Analyzing and Quantifying Quarterback Quality

Shailen Sheth, Rushil Gupta, Prabhjaap Singh, and Akhil Medarametla

I. Introduction

We have chosen to study the role of quarterbacks on NFL teams. Specifically, our goal was to quantitatively define the value of the most complex position in sports. Through data analysis, we delve into the multi-dimensional value of quarterbacks and pinpoint specific statistics that highlight their significant contributions to winning. Analyzing proficient quarterback play is essential for teams as it enables them to redefine their offensive game plan or potentially build a new game plan.

Effective Quarterback play can be measured by traditional statistics like QBR, Passer Rating, Completion Percentage etc. However, it is crucial to understand football is a team sport, where the cohesiveness of offensive and defensive players are the ones who can make or break a good quarterback. Wide receiver quality and their ability to create separation can elevate a quarterback's throwing stats to the next level. Similarly, the effectiveness of an offensive line in protecting a quarterback directly impacts the number of sacks and hits the player takes, which can negatively affect performance. Today, skilled offensive players are more important than ever, especially when the play breaks down and there is a need to improvise. Hence, our goal is to understand how we can determine a great quarterback? And is there a way to quantify such a complex concept?

II. Background

We used the 'NFLverse' dataset, a set of datasets of the National Football League using the professional league's data. The dataset has a numerous number of rows for each player who has ever played in the league and has counting stats. The rows we primarily focused on were those dedicated towards passing such as passing yards, passing attempts, touchdown passes and more. As well as rushing statistics such as rushing yards and carries to help us analyze the value of quarterbacks who are considered "dual-threat" or quarterbacks with more high rushing tendencies.

The time period we cover was from 2016 to 2022 as this 6-year span covered the period that a majority of NFL quarterbacks' career lengths would go to. For data cleaning, we removed all non-passing and rushing statistics for players designated as QBs. As well as, filtered out all stats prior to the 2016 season and quarterbacks who are below the desired passing attempt minimum. Specifically, removing quarterbacks with less than 16 games worth of passing attempts or 540 passing attempts from the dataset to ensure no outliers would remain.

III. Data Wrangling

A majority of the data transformation and feature engineering done for this report was creating more advanced variables for analysis. Specifically, we already had an idea in mind for how we wanted to explore. We wrangled data by separating it into different aspects of analysis. Hence, we start by developing efficiency metrics, which are metrics that determine how capable quarterbacks are at using their passing attempts. Our focus was on two computed metrics: Adjusted Net Yards per Attempt (ANY/A) and Passer Rating. ANY/A considers passing yards, touchdowns, interceptions, and sacks, providing a holistic measure of a quarterback's efficiency. On the other hand, Passer Rating involves a different calculation based on completion percentage, passing yards, touchdowns, and interceptions. We then identified the top quarterbacks and analyzed them further.

Adding more context to our data, we also manually gathered the data of top quarterbacks in terms of their win-loss record. This information was merged with the ANY/A data, creating a dataset that not only ranks quarterbacks by efficiency but also provides insight into their success on the field. Furthermore, we also wanted to look into the relationship between how often a quarterback completed a pass and how efficient they were. We used density plots to visualize the distribution and density of completion percentages concerning ANY/A and Passer Rating, creating a more in-depth perspective of quarterbacks.

The analysis extended beyond individual quarterback performance to explore team dynamics, specifically the role of O-linemen. We look specifically at pass protection events (sacks, QB hits, scrambles) for each team and correlate them with the team's average passer rating and completion percentage.

IV. Exploratory Analysis

This was our original exploratory analysis prior to finalizing how we wanted to collect and analyze data.

Efficiency Metrics (Y/A, AY/A, NY/A)

```
library("tidyverse")
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✔ dplyr     1.1.3     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.3     ✔ tibble    3.2.1
## ✔ lubridate 1.9.3     ✔ tidyr     1.3.0
## ✔ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── 
tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
```

```
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors
```

```
library("nflverse")
```

```
## ── Attaching packages ─────────────────────────────────────── nflverse
1.0.3 ──
## ✔ nflfastR 4.6.0     ✔ nflreadr 1.4.0
## ✔ nflseedR 1.2.0     ✔ nflplotR 1.2.0
## ✔ nfl4th   1.0.4
## ────────────────────────────────────────────────────────────────── Ready to
go! ──
```

```
library("ggrepel")
```

(Y/A)

```
qb_YA <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
filter(season_type == "REG" & position == "QB") %>% group_by(player_name) %>%
summarize(yards_per_attempt = sum(passing_yards)/sum(attempts),passing_yards
= sum(passing_yards), attempts = sum(attempts),currentTeam = recent_team)
```

```
## Warning: Returning more (or less) than 1 row per `summarise()` group was
deprecated in
## dplyr 1.1.0.
## ℹ Please use `reframe()` instead.
## ℹ When switching from `summarise()` to `reframe()`, remember that
`reframe()`
##   always returns an ungrouped data frame and adjust accordingly.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

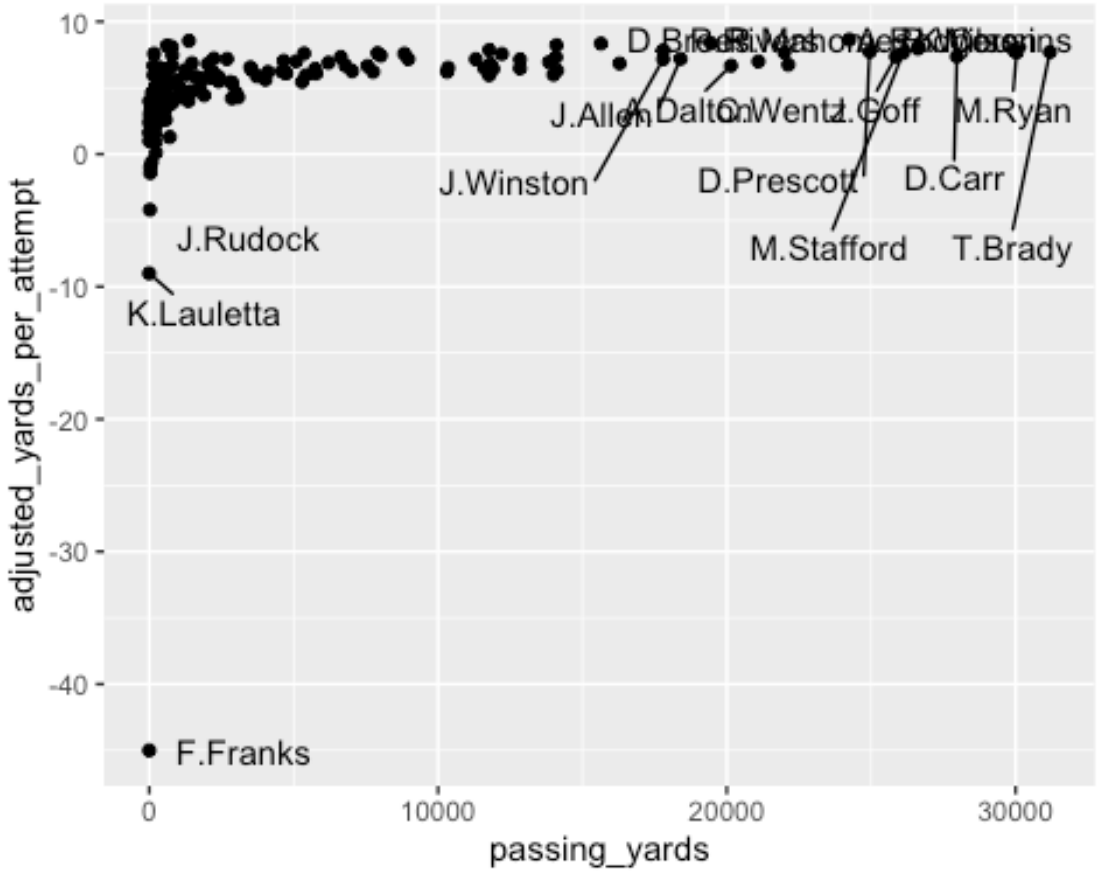
```
## `summarise()` has grouped output by 'player_name'. You can override using
the
## `.groups` argument.
```

(AY/A)

Weighted it by 20 so to ensure that QB touchdowns are still valued at a fair level

```
qb_AYA <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
  filter(season_type == "REG" & position == "QB") %>%
  group_by(player_name) %>%
  summarize(
    adjusted_yards_per_attempt = (sum(passing_yards) + 20 * sum(passing_tds)
- 45 * sum(interceptions)) / sum(attempts),
    passing_yards = sum(passing_yards)
  )
ggplot(data = qb_AYA, aes(x =passing_yards , y = adjusted_yards_per_attempt))
+ geom_point() + geom_text_repel(aes(label = player_name), box.padding = 0.5,
point.padding = 0.2, force = 0.5, max.overlaps = 35 )
```

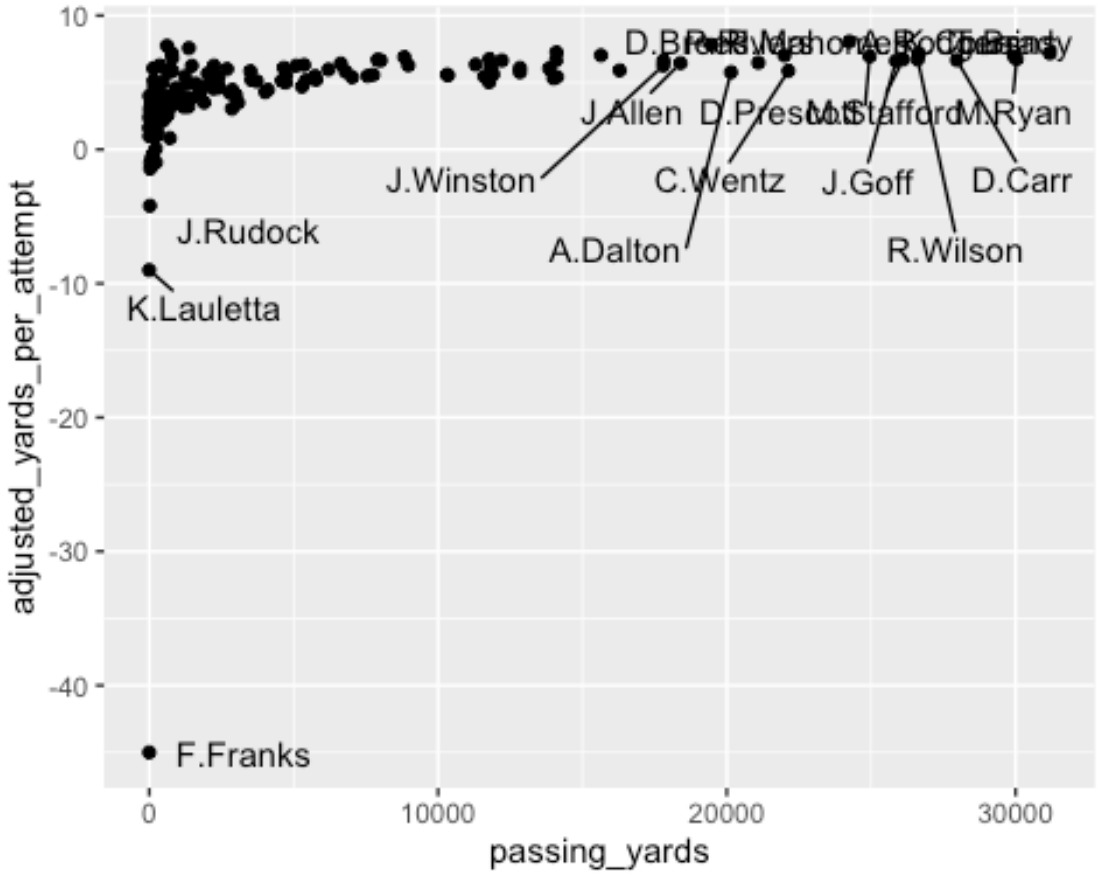
```
## Warning: Removed 5 rows containing missing values (`geom_point()`).

## Warning: Removed 6 rows containing missing values (`geom_text_repel()`).

## Warning: ggrepel: 131 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



#(ANY/A)

```
qb_ANYA <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
  filter(season_type == "REG" & position == "QB") %>%
  group_by(player_name) %>%
  summarize(
    adjusted_yards_per_attempt = (sum(passing_yards) + 20 * sum(passing_tds)
- 45 * sum(interceptions) - sum(sack_yards)) / (sum(attempts) + sum(sacks)),
    passing_yards = sum(passing_yards)
  )


ggplot(data = qb_ANYA, aes(x =passing_yards , y =
adjusted_yards_per_attempt)) + geom_point() + geom_text_repel(aes(label =
player_name), box.padding = 0.5, point.padding = 0.2, force = 0.5,
max.overlaps = 35 )
```

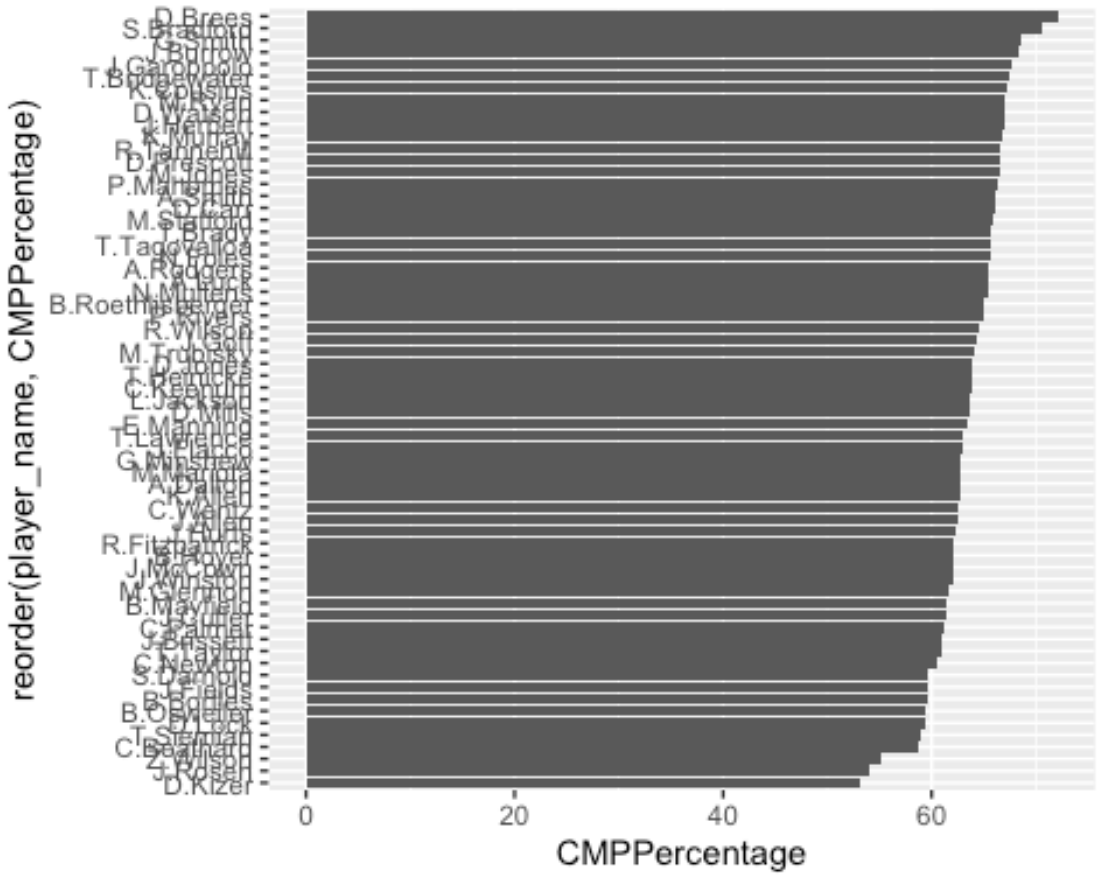
```
## Warning: Removed 5 rows containing missing values (`geom_point()`).

## Warning: Removed 6 rows containing missing values (`geom_text_repel()`).

## Warning: ggrepel: 131 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



Accuracy Metrics

Completion Percentage

```
qb_CMP <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
  filter(season_type == "REG" & position == "QB") %>%
  group_by(player_name) %>%
  summarize(
    CMPPercentage = (sum(completions) / sum(attempts)) * 100,
    average_yards_per_attempt = (sum(passing_yards) + 20 * sum(passing_tds) -
45 * sum(interceptions)) / sum(attempts), attempts = sum(attempts)
  ) %>% filter(attempts > 500)
ggplot(data = qb_CMP, aes(x = CMPPercentage,y= reorder(player_name,
CMPPercentage))) + geom_bar(stat = "identity")
```



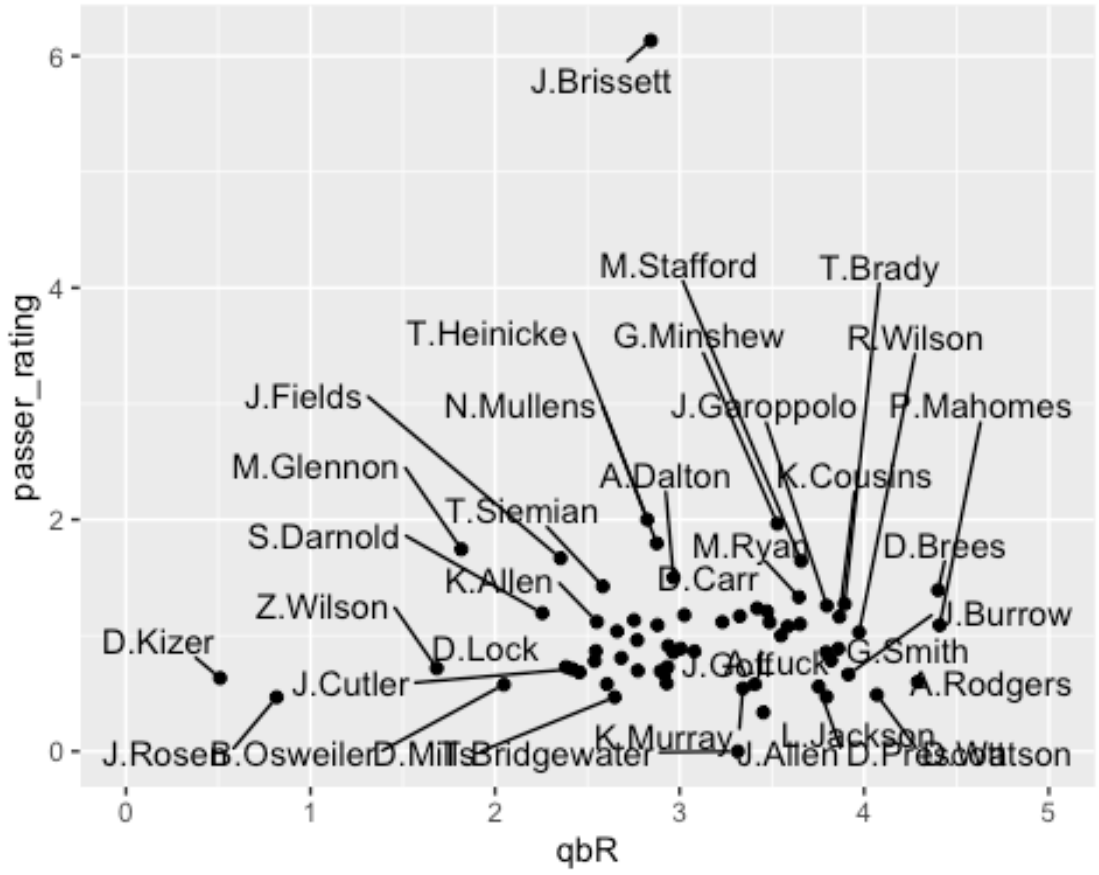
Basic Analysis: Using this graph, we can get a clear sense of the quarterbacks who are up there in terms of their completion percentage, such as Drew Brees, Sam Bradford, Geno Smith, and Burrow. These are relatively interesting to see, as there were other quarterbacks, who one would consider to be better like Patrick Mahomes and Tom Brady. Yet, they were relatively further down on the list. However, I do feel that the completion percentage with further analysis will be seen as being focus more on quarterback's who are focused more shorter yardage, whereas Mahomes and Wilson will be more aggressive in nature.

ADVANCED STATS

QB ratings

```
qb_QBR <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
  filter(season_type == "REG" & position == "QB") %>%
  group_by(player_name) %>%
  mutate(
    qb_CMP = (sum(completions) / sum(attempts)) * 100,
    qbYA = sum(passing_yards) / sum(attempts),
    tdPercent = (sum(passing_tds) / sum(completions)) * 100,
    intPercent = (sum(interceptions) / sum(completions)) * 100,
    qbR = ((qb_CMP - 30) / 20) + ((qbYA - 3) / 4) + (tdPercent / 5) -
(intPercent / 4),
```

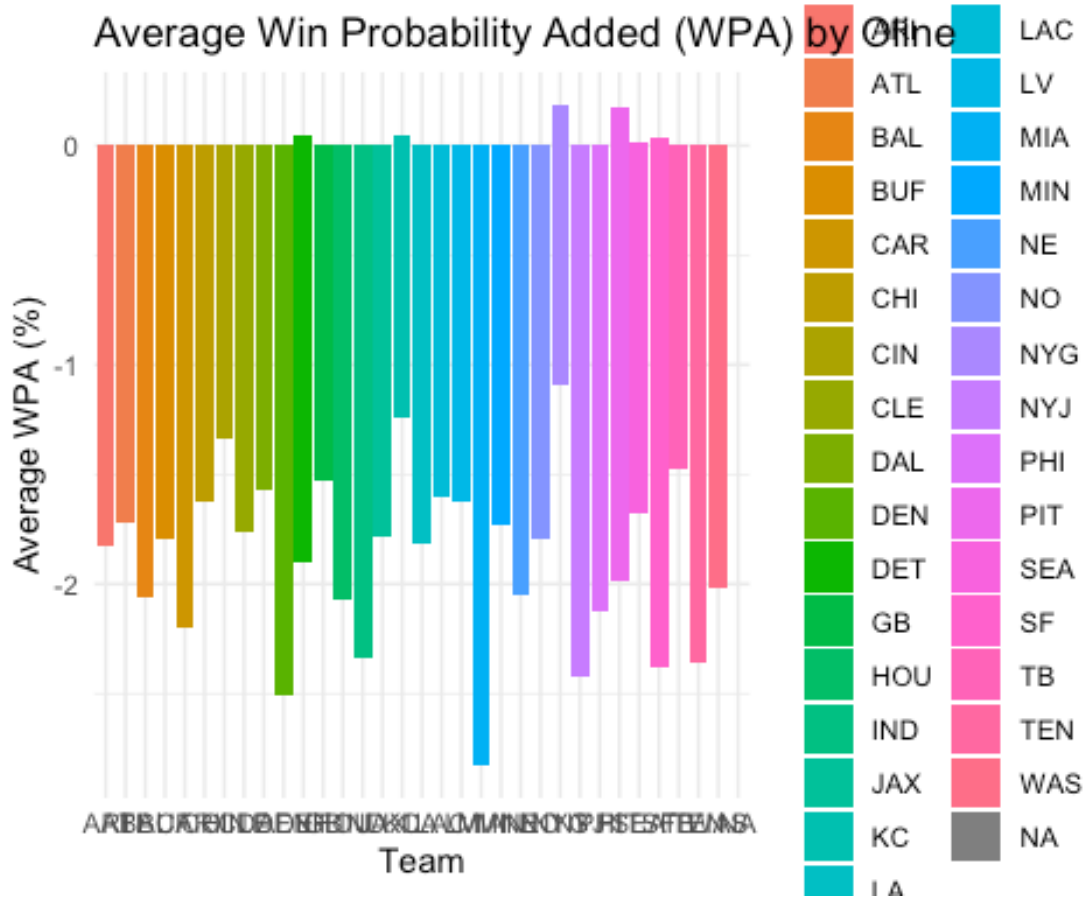
```
    passer_rating = pacr,
    attempts = sum(attempts)
  ) %>%
  ungroup() %>%
  distinct(player_name, .keep_all = TRUE) %>%
  filter(attempts > 500)

ggplot(data = qb_QBR, aes(x = qbR, y = passer_rating)) + geom_point() +
xlim(0,5) + geom_text_repel(aes(label = player_name), box.padding = 0.5,
point.padding = 0.2, force = 0.5, max.overlaps = 50)

## Warning: Removed 1 rows containing missing values (`geom_point()`).

## Warning: Removed 1 rows containing missing values (`geom_text_repel()`).

## Warning: ggrepel: 27 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



Analysis: Here we can see more clearly how good some quarterbacks are. Drew Brees continues to be a high quality player even in the twilight of his career. Furthermore compared to completion percentage, we see mahomes and Aaron Rodgers joining as high quality players. While Rosen and Kizer continue to be Quarterbacks who are not considered valuable to their teams as well as consistently not demonstrating any value in terms of Quarterback analytics.

O-line Quality

```
 pbp <- nflreadr::load_pbp(2022) %>% filter(season_type == "REG") %>%
mutate(qbh_inc = ifelse(qb_hit == 1 & incomplete_pass == 1, 1,0), qb_int =
ifelse(qb_hit == 1 & interception == 1, 1,0))

nfl_plays <- pbp %>% group_by(week, posteam) %>% summarize( total_plays =
sum(play_type == "run" | play_type == "pass", na.rm = TRUE) )

## `summarise()` has grouped output by 'week'. You can override using the
## `.groups` argument.

# Calculate Total Line Value

nfl_line_value <- pbp %>% filter( sack == 1 | tackled_for_loss == 1 |
yards_gained <= 2 | qb_scramble == 1 | qb_hit == 1 | qbh_inc == 1 | qb_int ==
1 ) %>% left_join(nfl_plays, by = c("posteam" = "posteam", "week" = "week"))
%>% group_by(week, posteam) %>% summarize( nfl_sum_wpa = sum(wpa, na.rm =
TRUE), nfl_total_plays = sum(unique(total_plays)), nfl_avg_wpa = (nfl_sum_wpa
/ nfl_total_plays * 100))

## `summarise()` has grouped output by 'week'. You can override using the
## `.groups` argument.

ggplot(nfl_line_value, aes(x = posteam, y = nfl_avg_wpa, fill = posteam)) +
geom_bar(stat = "identity", position = "dodge") + labs(title = "Average Win
Probability Added (WPA) by Oline", x = "Team", y = "Average WPA (%)") +
theme_minimal()

## Warning: Removed 18 rows containing missing values (`geom_bar()`).
```



Analysis: Here we did not get as far as desired in terms of demonstrating Oline quality in regards to talent. However I do believe that we can get more out of this if we were to figure out a more adept QUarterbacks tat to put into comparison. However, we can tell that Minnesota is getting the most value out of their Oline compared to many others. However, looking at actual in-game footage, they have been weaker than can be seen. However, we need to look more into what makes an Oline work, especially if we go through it play-by-play rather than just in a cumulative manner.

Touchdown to Interception Ratio

```
qb_T2I <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
filter(season_type == "REG" & position == "QB") %>% group_by(player_name) %>%
summarize(T2I = sum(passing_tds)/sum(interceptions))
```

Sack Metrics

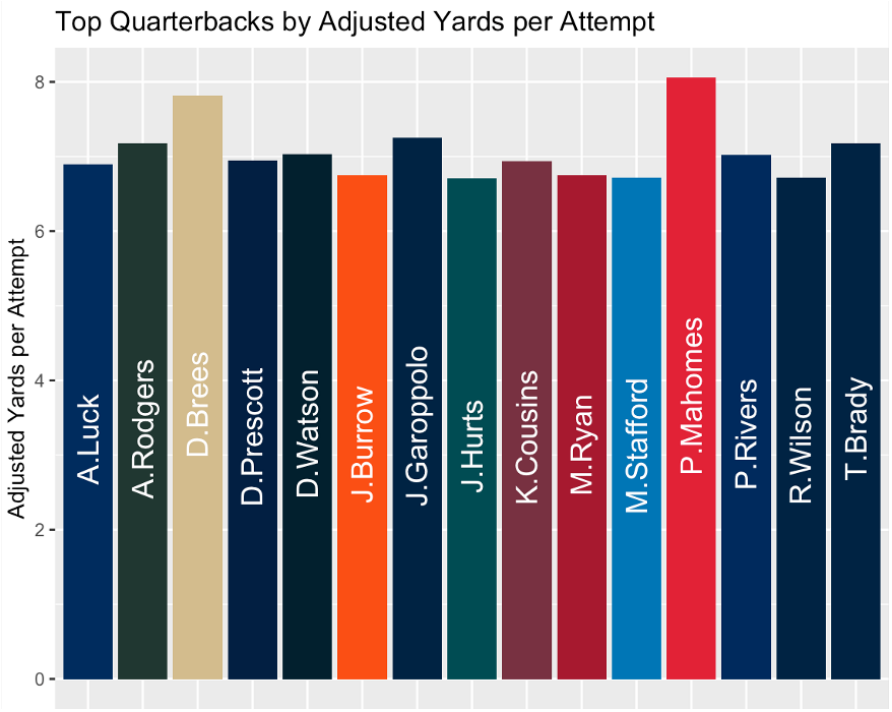
```
qb_SackRate <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
filter(season_type == "REG" & position == "QB") %>% group_by(player_name) %>%
summarize(sum(sacks)/sum(attempts))
```

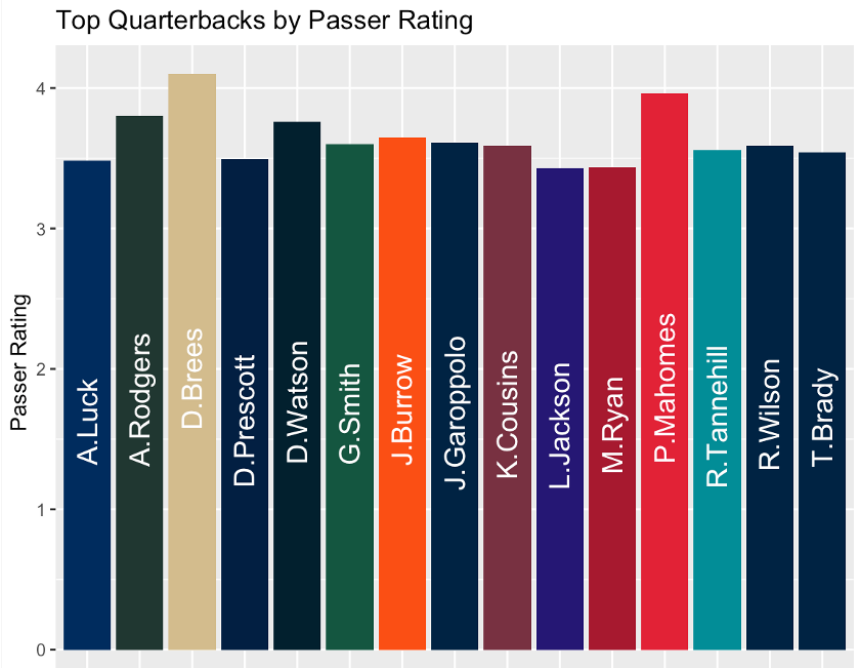
Air Yards

```r
qb_AirYards <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
filter(season_type == "REG" & position == "QB") %>% group_by(player_name) %>%
summarize(totalAirYards = sum(passing_air_yards))

qb_YAC <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
filter(season_type == "REG" & position == "QB") %>% group_by(player_name) %>%
summarize(totalYAC = sum(passing_yards_after_catch))
```

V. Final Analysis

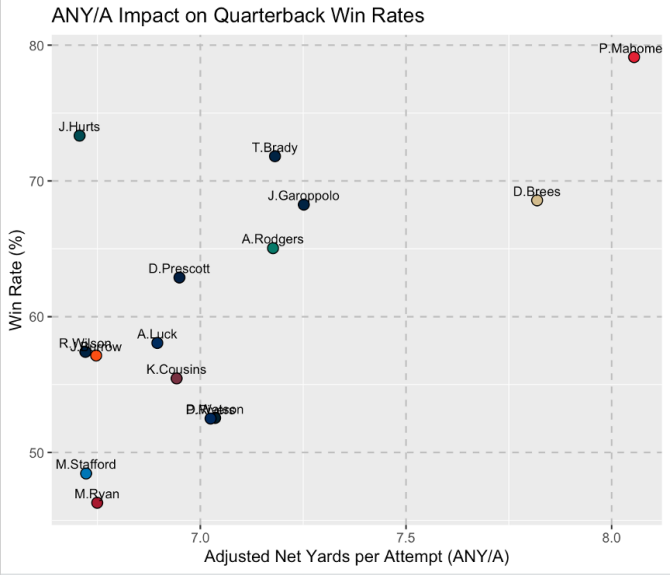
How Passer Rating Defines Quarterback Quality and Team Success



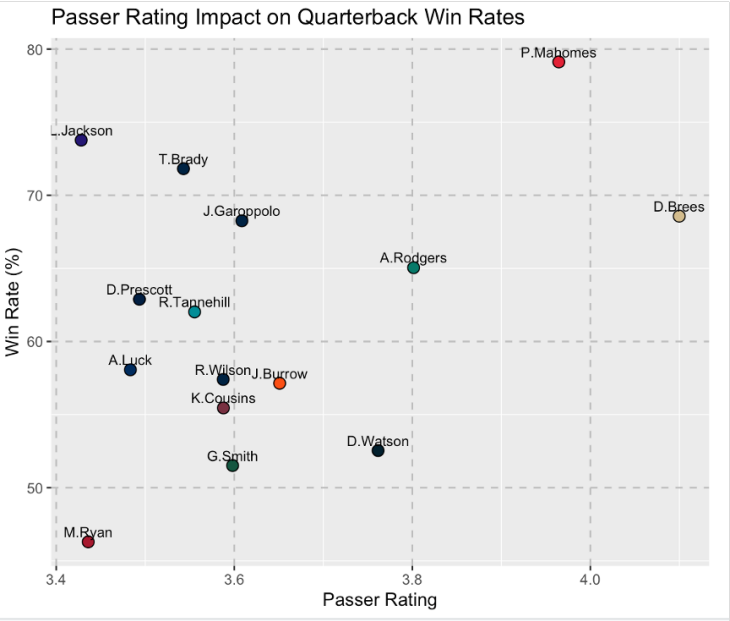


When looking at the analysis for quarterback passer rating, we must first determine the metrics for Adjusted Net Yards per Attempt (ANY/A) and Passer Rating. Since these values were given by the "NFLverse" dataset, we were able to create a bar plot that compared the top quarterbacks for their ANY/A and passer rating. In this histogram, it was clearly seen that Drew Brees and Patrick Mahomes had the highest statistics compared to the other quarterbacks. While this visualization does give insight for specific qualities that could affect the WPA, we wanted to take a closer look at how these metrics affect the overall team win rate.

In order to do this, we made two scatter plots that showed the team win rate with either the ANY/A or passer rating metrics. For the ANY/A scatter plot, we were able to see that most of the data was plotted in the lowest two quadrants, indicating that lower ANY/A corresponded to a lower team win rate. However, we do encounter some outliers in this chart, where we see that Mahomes and Brees, who had the highest ANY/A, had an increased win rate. This indicates that there is a potential positively linear relationship between ANY/A and win rate. However, Jalen Hurts is an additional outlier in this situation where we see that he has one of the lowest ANY/A values out of all the quarterbacks, but a relatively high win rate.

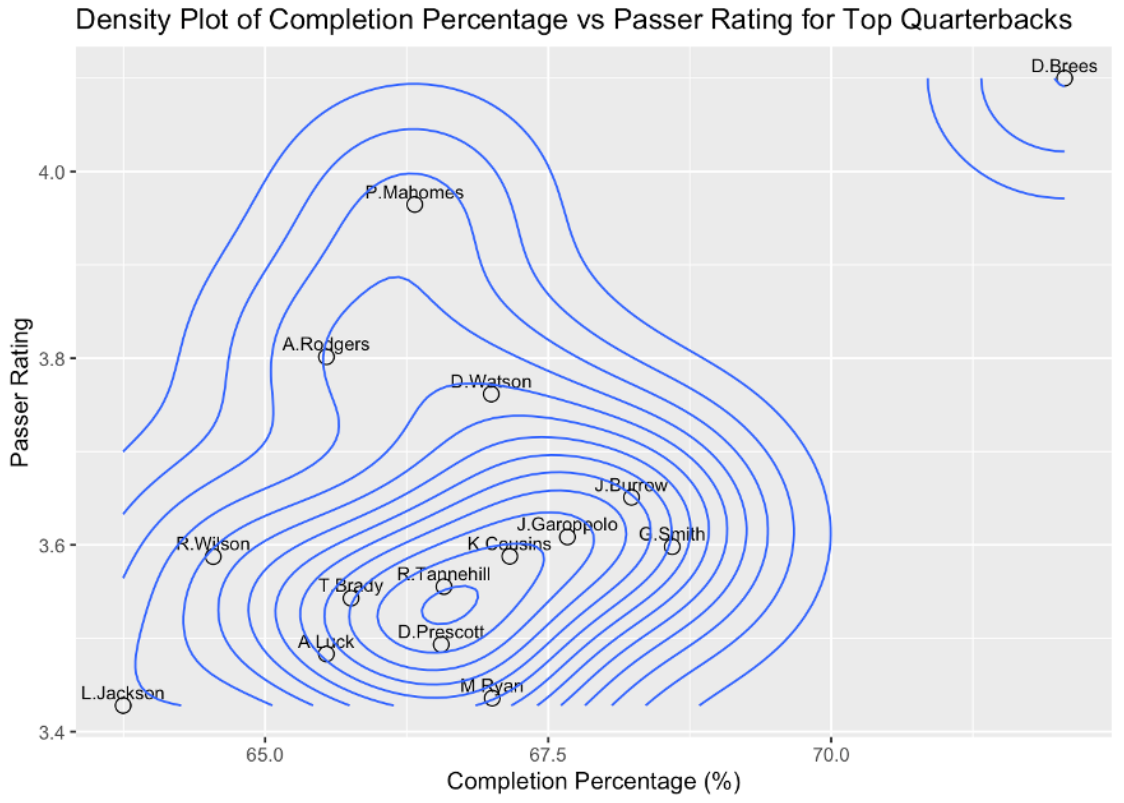


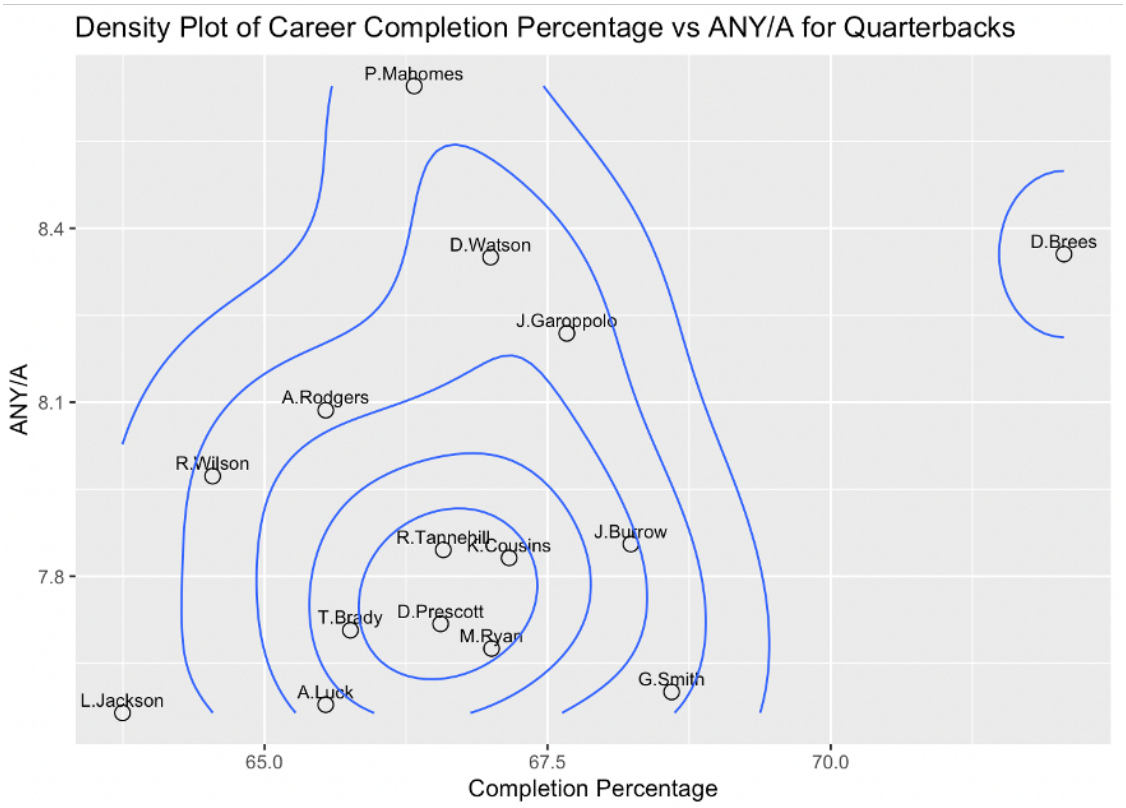
When looking at the passer rating scatter plot, we see a similar trend, where most of the data is plotted in the last two quadrants. While the data is more spread out in this scatter plot, it is clear that there is a positive linear correlation, where a lower passer rating indicates a lower win rate, while higher passer rating generally indicates a higher team win rate. Again, we see that Mahomes and Brees have the highest passer ratings as well as the highest team win rate. Additionally, we also see an outlier, Lamar Jackson, who has the second lowest quarterback passer rating but the second highest team win rate. Overall, it is clear to see from these two scatter plots that ANY/A and passer rating do provide a useful metric to understand their WPA. However, there are still some outliers in this situation that need additional metrics and player background to be understood.



Finally, in the next analysis, we utilized density plots to explore the relationships between career completion percentages and ANY/A as well as passer rating. In both of these plots

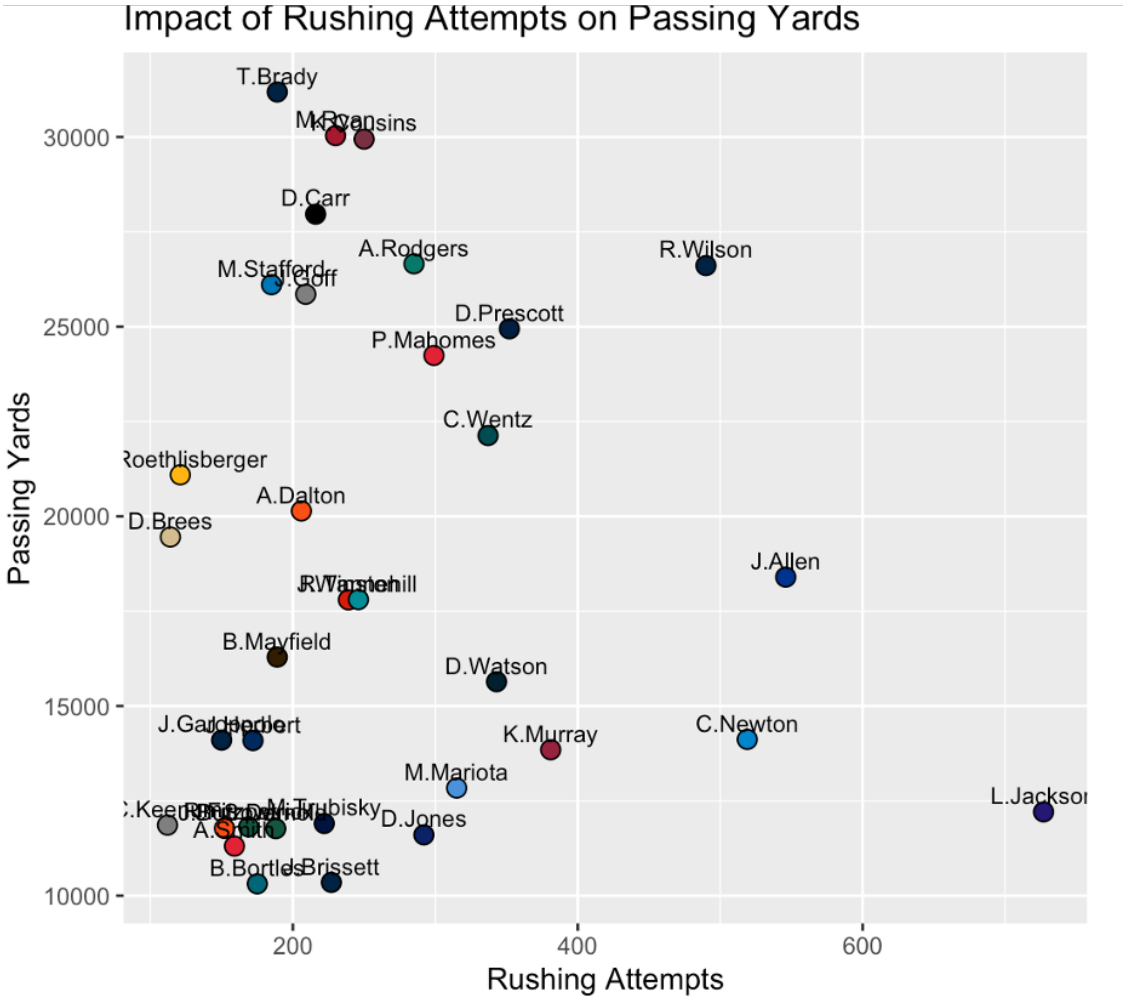
we can see that most of the data is centered in the middle of the plot, showing that players with similar completion percentages and passer ratings or ANY/A contribute similar WPA to the teams they are on. Notably, Drew Brees stands out as an outlier in these graphs, where we see that he has the highest completion percentage compared to the other quarterbacks. Overall, this contour plot was able to show the similarity in WPA for quarterbacks with certain completion percentages, ANY/A, and passer rating.



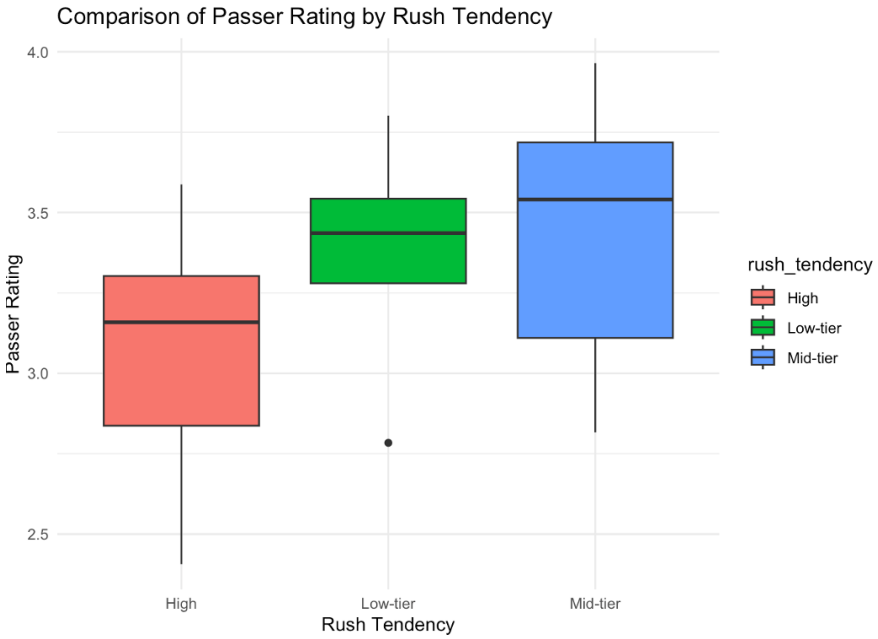


Impact of Rushing on Quarterback Quality

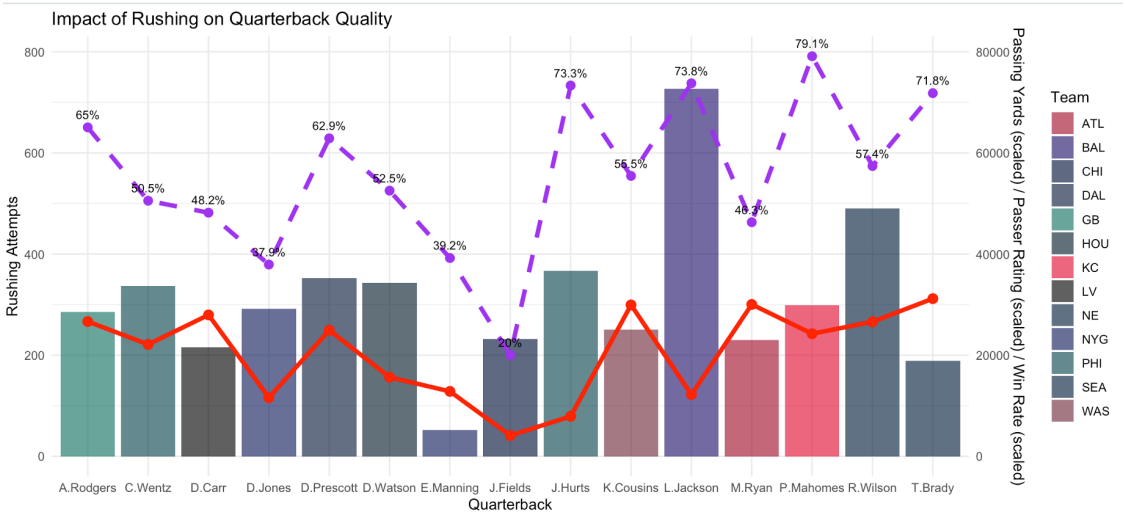
To begin determining how the rushing metrics of a quarterback's game affects their overall quality as a quarterback, we first made a chart that plotted the impact of their rushing attempts on passing yards for all statistics between after the 2016 season and filtered out all the data before 2016 seasons, in addition to filtering out data for quarterbacks who are below the desired passing minimum attempt to ensure there is a decent sample size to perform this analysis. The number of times a quarterback has attempted to rush is plotted on the x axis, while the number of passing yards a quarterback has thrown for is plotted on the y axis. After plotting this graph, there were several things we found out. The first thing we concluded were which quarterback, or quarterbacks, threw for more yards and who had more passing attempts. The quarterback with the most rushing attempts is Lamar Jackson of the Baltimore Ravens. On the other end, the quarterback with the most passing yards was Tom Brady, a retired NFL quarterback who was a member of the New England Patriots for the majority of his career and a member of the Tampa Bay Buccaneers for the last couple years of his career. These two quarterbacks sit on the far end of the spectrum, having the most rushing attempts and passing yards, respectively.



Upon analysis, we decided that we had to define the different run tendencies for quarterbacks into three different categories, which were high rushers, moderate rushers, and non-rushers. Each of these categories has a certain set of minimum numbers to reach to qualify for that category. High rushers had to average more than thirty rushing yards a game, three rushing touchdowns in a season, and average at least five rushes a game. Moderate rushers had to average ten to thirty yards a game, attempt between two and five rushes a game, and have one to two rushing touchdowns in a season. A non-rusher would have to have less than ten yards a game, less than two attempts per game, and no rushing touchdowns in a season. We use these different sets of standards and baselines to categorize quarterbacks, ultimately giving us insight to each quarterback's physical build and how it contributed to their style of play, efficiency, and their success in the NFL as a quarterback.



We constructed a box-and-whisker plot to analyze and compare the passer ratings by rush tendencies and categorized the quarterbacks into the categories we created. This box-and-whisker plot has the rush tendency listed on the x-axis, categorized by the different ties we created, and the passer rating is listed on the y-axis. There are few quarterbacks we can place into each of the categories to group similar quarterbacks. In the high-rusher-tier, there are quarterbacks like Lamar Jackson, Josh Allen, Cam Newton, Jalen Hurts, Russell Wilson, and Justin Fields. Quarterbacks in the high-rusher-tiers tend to be on the smaller end of NFL quarterbacks, contributing to their high-mobility. In the moderate-rusher-tier, there are quarterbacks like Daniel Jones, Deshaun Watson, Patrick Mahomes, Dak Prescott, Carson Wentz, and Kirk Cousins. Quarterbacks in the mid-rusher-tier tend to be bigger than those in the high-rusher-tier, but smaller than those in the non-rusher-tier, contributing to their ability to run down the field. In the non-rusher-tier, there are quarterbacks like Eli Manning, Matt Ryan, Derek Carr, Aaron Rodgers, and Tom Brady. Quarterbacks in the non-rusher-tier tend to be bigger and taller than the others, making them less mobile. The main point of this box-and-whisker plot was to analyze the spread of quarterbacks in each tier. From this, there are a few things we can conclude about each tier. The first is that non-rusher-tier quarterbacks have the smallest spread in terms of passer rating, between 3.25 and 3.5, and they are between the moderate-rusher-tier and high-rusher-tier in terms of rush tendency. Next, moderate-rusher-tier quarterbacks seem to have the largest spread in passer rating, from around 3.1 to 3.7, and they seem to have the highest rushing tendency. Finally, the high-rusher-tier quarterbacks have a decent, albeit lower, spread in terms of passer rating, from 2.6 to 3.3, and are lowest in terms of rushing tendency compared to the other tiers.



Based on our analysis, we can determine that rushing does have an impact on the overall quality of a quarterback. The last step in our analysis of the impact that rushing has on a NFL quarterback's overall quality is to a bar graph that measures three things about the quarterback: their passing yards, passer rating, and win rate. We plotted 15 quarterbacks across the multiple tiers to give an accurate and diverse representation of quarterbacks. The bar graph has three different visual elements to represent the three statistics that we measured for the quarterbacks. The first visual element is the quarterback's team, which is prepared as a bar and differentiated by color, and that bar measures the number of rushing attempts from the quarterback for their team. The second visual element is a dotted purple line that measures the quarterback's completion percentage over the time frame we established. The third visual element is a solid red line that measures the number of rushing yards the quarterback has. This graph is the culmination of our analysis of the impact that rushing has on the quarterback's overall quality. This graph really describes to us the different-tiered quarterbacks, their effect on the team, their own statistics, and helps us further understand the variation in rushing and its overall effect on the quarterback's quality.

Offensive Line Elevates Quarterback Quality

Offensive Line play is a key determinant of Quarterback play and quality.

The role of the offensive line is considered an essential part of improving quarterback play. Prior to going over how, we first need to understand what an offensive line is.

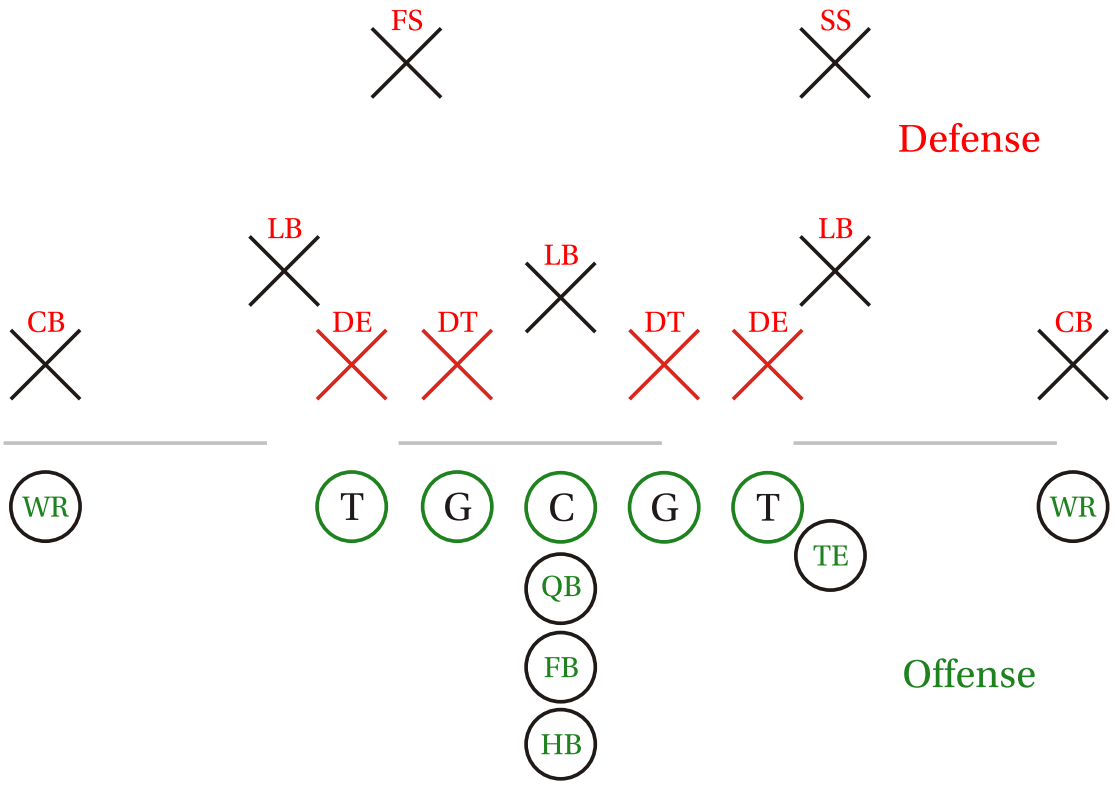
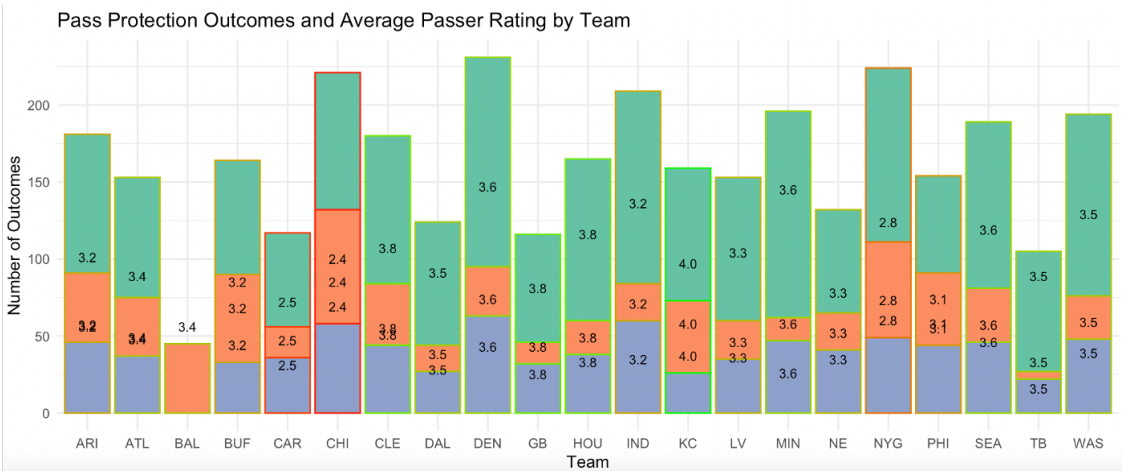


Figure I. Offensive Line Structure

Looking at the figure for reference, the offensive line consists of a center, who snaps the ball, two guards who flank the center, and two tackles on either side of the guards. We chose to measure offensive line quality by creating a bar graph that collects. To measure offensive line quality, we decided to use the data from nflverse and utilize it to make a bar graph showing key metrics. Performance of an offensive line is measured by the absence of sacks, QB scrambles, weak yards gained(QB runs of 2 yards or less), etc. As such the statistics we used from the data on our bar chart include Quarterback Hits, Quarterback scrambles and Quarterbacks sacks and categorized these as “outcomes”. We utilized these stats and plotted them on a graph which displayed the Number of these outcomes over Time. These results were clearly evident. We made 32 graphs for the 32 teams for the outcome data from 2016-2022 and plotted a summarized number of team/QB passer ratings in the middle of each graph. From here there was a clear trend that teams with more “negative” outcomes led to a lower team passer rating. Teams like ChicagoBears , Indianapolis Colts, and New York Giants all had team passer ratings below 3.0 while having the highest outcome rates. One outlier could be seen as Denver while also had one of the highest outcome rates as well as one of the highest passer ratings at 3.6. This can be attributed to historic data on the quarterbacks who play in Denver and their offensive game plan which is built around QB scrambles and extending the play.



Conclusion

Our exploration has exposed a crucial issue with our desire to quantitatively define good quarterback play: there is no single statistic that could comprehensively cover all aspects of quarterback success. Instead, the multifaceted nature of the playstyle demanded means that we need to look past basic metrics or even advanced metrics to define them. Our analysis of Adjusted Net Yards per Attempt (ANY/A), passer score, speeding dispositions, and the effect of the offensive line has highlighted the many elements influencing quarterback performance.

Efficiency metrics like passer score and ANY/A provide insights into how efficient a quarterback can be. However, they also don't attempt to go over the impact of a quarterback's ability to rush or even the quality of the team's offensive linemen. While certain traits and correlations are obvious in determining success, the presence of outliers, together with Jalen Hurts or Lamar Jackson, shows how individualized the nature of quarterback success is. A quarterback's effectiveness isn't always determining how successful they are and how valuable to the team they are, instead, we need to look at how all of these different aspects intersect with one another.

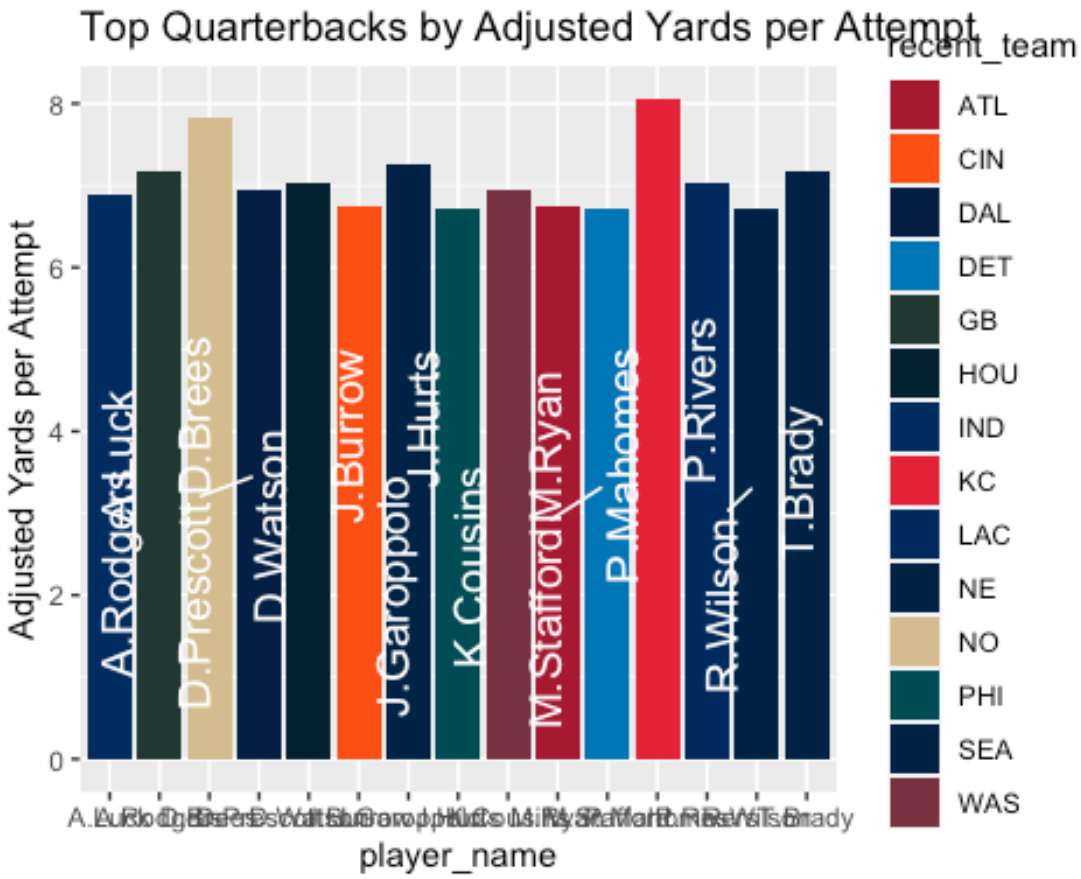
Our analysis leads us to the conclusion that quantitatively determining good quarterback play requires a balanced consideration of numerous metrics and factors. Efficiency, rushing ability, and the guidance of the team all contribute considerably to a quarterback's effect, however, it is clear that they are not all the define a good quarterback. We instead need to acknowledge the various approaches of quarterbacks or how what may have made one quarterback successful on the team will not always correlate or work on another.

Code Appendix

Efficiency Metrics

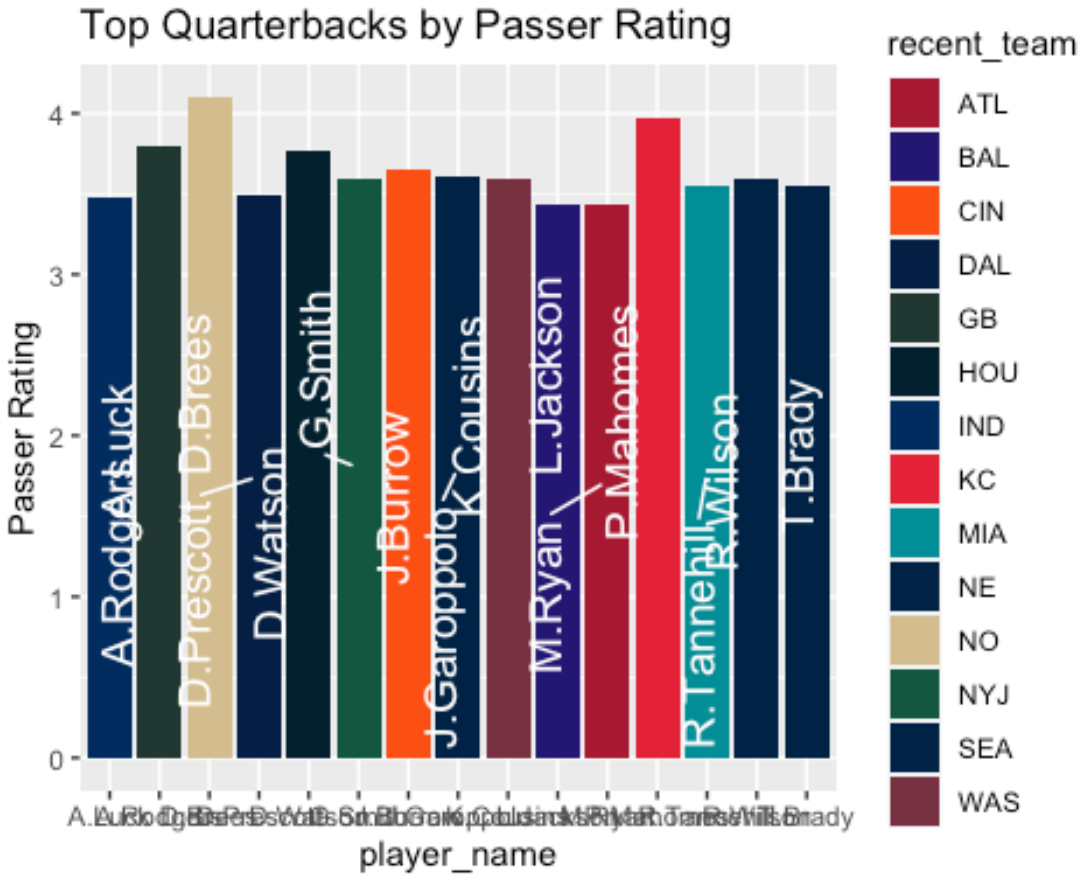
```
#(ANY/A)
qb_ANYA <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
  filter(season_type == "REG" & position == "QB") %>%
  group_by(player_name) %>% filter(sum(attempts) > 554) %>%
  summarize(
```

```r
    adjusted_yards_per_attempt = (sum(passing_yards) + 20 *
(sum(passing_tds)) - 45 * (sum(interceptions)) - sum(sack_yards)) /
((sum(attempts)) + sum(sacks)),
    passing_yards = sum(passing_yards),
    recent_team = first(recent_team)  # Use an appropriate aggregation
function for recent_team
  )

#take top 10
top_qbs <- qb_ANYA %>%
  arrange(desc(adjusted_yards_per_attempt)) %>%
  slice(1:15)



# Modify ggplot code to include team colors
ggplot(data = top_qbs, aes(x = player_name, y = adjusted_yards_per_attempt,
fill = recent_team)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = team_colors) +  # Assign team colors
  # Updated geom_text_repel with adjustments
  geom_text_repel(aes(label = player_name),
                  box.padding = 0.5,
                  point.padding = 0.2,
                  force = 0.5,
                  max.overlaps = 35,
                  vjust = -0.5,  # Adjust vertical position inside the bar
                  angle = 90,   # Flip the text vertically
                  color = "white",
                  position = position_stack(vjust = 0.5),
                  size = 5)  + # Adjust the size of the text
                  labs(title = "Top Quarterbacks by Adjusted Yards per
Attempt", y = "Adjusted Yards per Attempt")
```



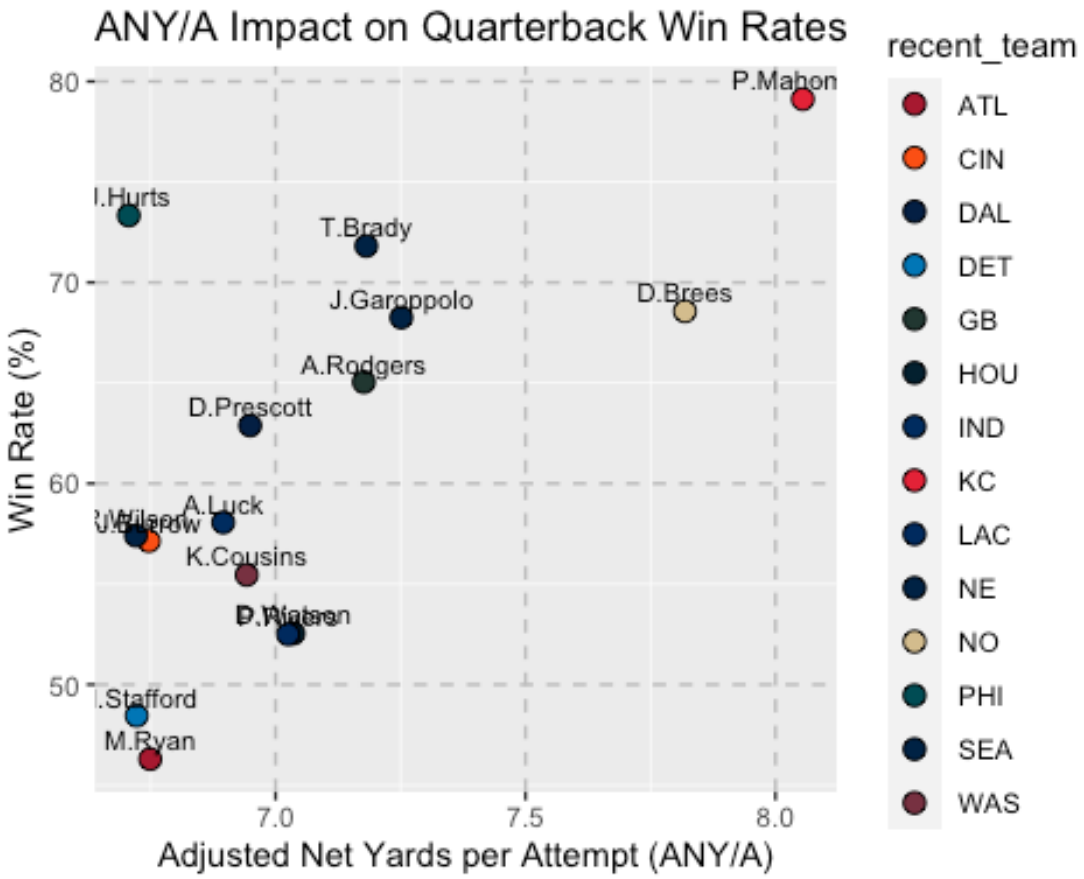

```
qb_QBR <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
  filter(season_type == "REG" & position == "QB") %>%
  group_by(player_name) %>%filter(sum(attempts) > 554) %>%
  summarize(
    completions = sum(completions),
    attempts = sum(attempts),
    passing_yards = sum(passing_yards),
    passing_tds = sum(passing_tds),
    interceptions = sum(interceptions),

    # Calculate individual components
    C = (completions / attempts) * 100,
    Y = passing_yards / attempts,
    TD = (passing_tds / attempts) * 100,
    INT = (interceptions / attempts) * 100,

    # Calculate Passer Rating
    passer_rating = ((C - 30) / 20) + ((Y - 3) / 4) + (TD / 5) - (INT /
4),recent_team = first(recent_team)
  )

#take top 10
top_qbs_QBR <- qb_QBR %>%
```

```
  arrange(desc(passer_rating)) %>%
  slice(1:15)

ggplot(data = top_qbs_QBR, aes(x = player_name, y = passer_rating, fill =
recent_team)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = team_colors) +  # Assign team colors
  # Updated geom_text_repel with adjustments
  geom_text_repel(aes(label = player_name),
                  box.padding = 0.5,
                  point.padding = 0.2,
                  force = 0.5,
                  max.overlaps = 35,
                  vjust = -0.5,  # Adjust vertical position inside the bar
                  angle = 90,   # Flip the text vertically
                  color = "white",
                  position = position_stack(vjust = 0.5),
                  size = 5) +   # Adjust the size of the text
  labs(title = "Top Quarterbacks by Passer Rating", y = "Passer Rating")
```

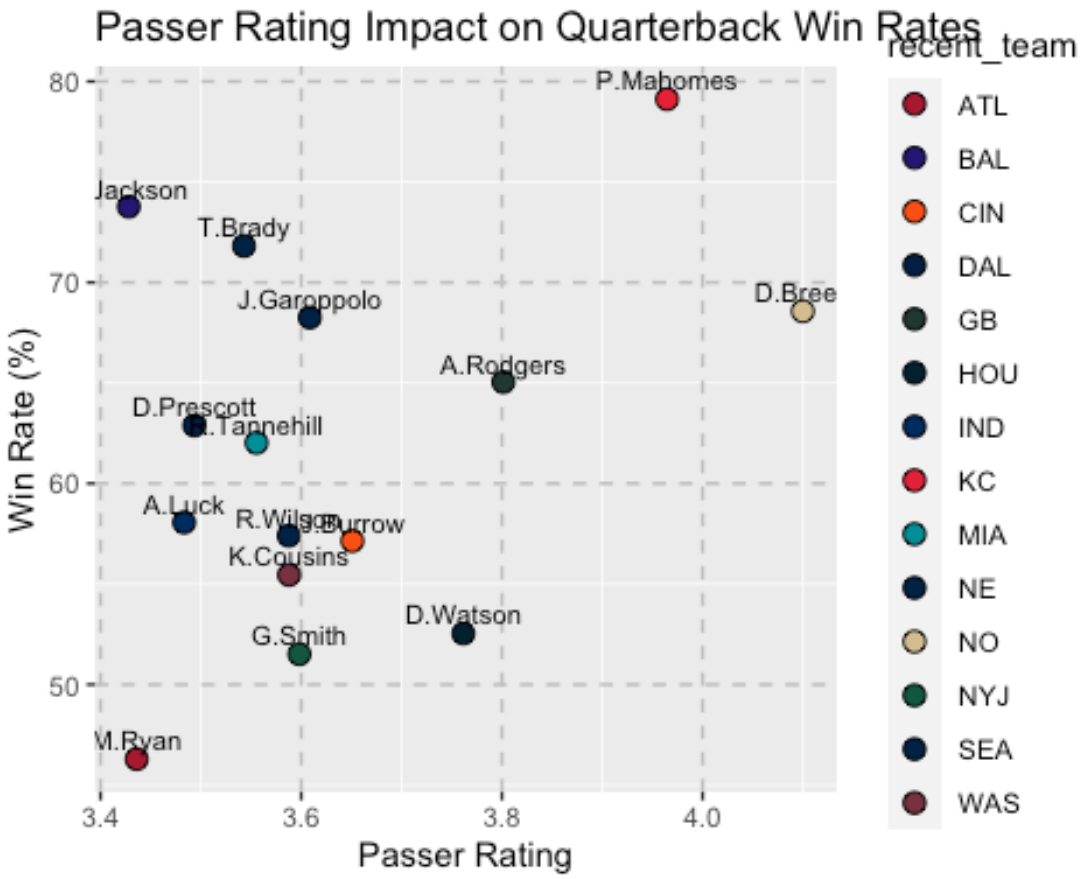



```
generic <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
  filter(season_type == "REG" & position == "QB")
```

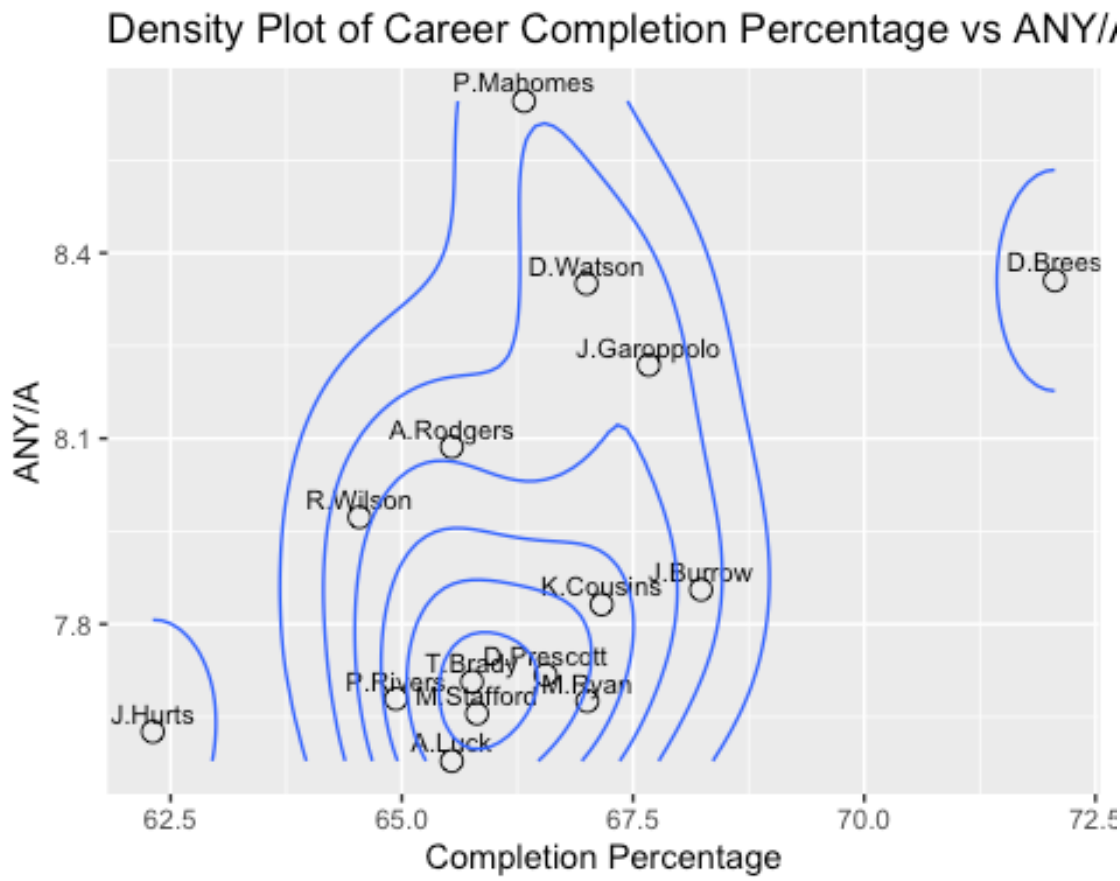
```r
##ANY/A with records
## get win numbers per player
qb_records <- data.frame(
  player = c("P.Mahomes", "D.Brees", "J.Garoppolo", "T.Brady", "A.Rodgers",
"D.Watson", "P.Rivers", "D.Prescott", "K.Cousins", "A.Luck", "M.Ryan",
"J.Burrow", "R.Wilson", "L.Jackson", "G.Smith", "M.Stafford", "J.Hurts",
"R.Tannehill"),
  wins = c(72, 48, 43, 79, 67, 31, 42, 61, 61, 18, 50, 24, 62, 45, 17, 47,
33, 49),
  losses = c(19, 22, 20, 31, 36, 28, 38, 36, 49, 13, 58, 18, 46, 16, 16, 50,
12, 30)
)

top_qbs_with_records <- inner_join(top_qbs, qb_records, by = c("player_name"
= "player"))



ggplot(top_qbs_with_records, aes(x = adjusted_yards_per_attempt, y = (wins /
(wins + losses))*100, label = player_name,fill = recent_team)) +
  geom_point(size = 3, shape = 21, color = "black") +
  geom_text(aes(label = player_name), vjust = -0.5, hjust = 0.5, size = 3) +
# Add labels above the points
  scale_fill_manual(values = team_colors) +  # Assign team colors
  labs(title = "ANY/A Impact on Quarterback Win Rates",
       x = "Adjusted Net Yards per Attempt (ANY/A)",
       y = "Win Rate (%)") +
  theme(
    panel.grid.major = element_line(color = "gray", linetype = "dashed"),
  )
```



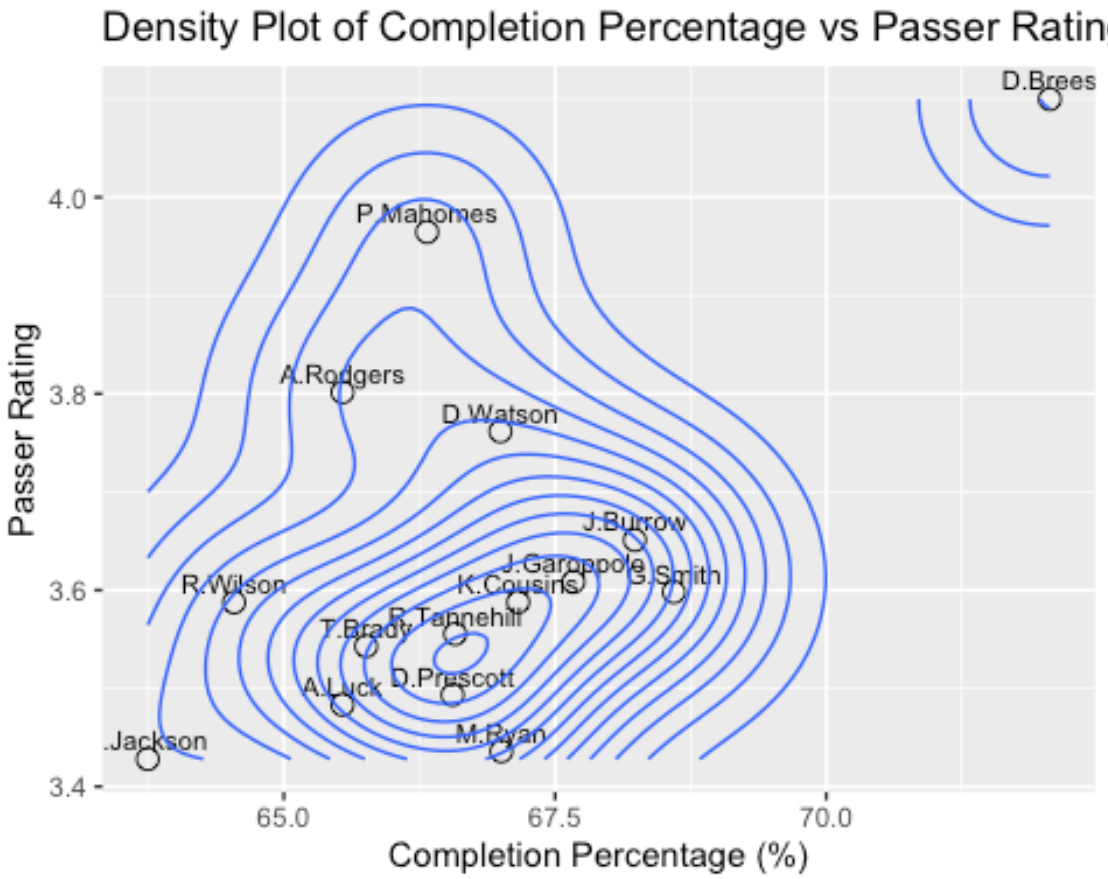

```r
# do it for passer rating
top_qbsQBR_with_records <- inner_join(top_qbs_QBR, qb_records, by =
c("player_name" = "player"))

ggplot(top_qbsQBR_with_records, aes(x = passer_rating, y = (wins / (wins +
losses))*100, label = player_name,fill = recent_team)) +
  geom_point(size = 3, shape = 21, color = "black") +
  geom_text(aes(label = player_name), vjust = -0.5, hjust = 0.5, size = 3) +
# Add labels above the points
  scale_fill_manual(values = team_colors) +  # Assign team colors
  labs(title = "Passer Rating Impact on Quarterback Win Rates",
       x = "Passer Rating",
       y = "Win Rate (%)") +
  theme(
    panel.grid.major = element_line(color = "gray", linetype = "dashed"),
  )
```




```r
#Completion Percentage v ANY/A
qb_CMP <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
  filter(season_type == "REG" & position == "QB") %>%
  group_by(player_name) %>% summarize(CMPPercentage =
((sum(completions)/sum(attempts)))*100, average_yards_per_attempt =
(sum(passing_yards) + 20*(sum(passing_tds)) -
45*(sum(interceptions)))/sum(attempts))

combined_data_ANYA <- left_join(top_qbs_with_records, qb_CMP, by =
"player_name") #any/a

ggplot(combined_data_ANYA, aes(x = CMPPercentage, y =
average_yards_per_attempt)) +
  geom_point(size = 3, shape = 21, color = "black") +
  geom_text(aes(label = player_name), vjust = -0.5, hjust = 0.5, size = 3) +
# Add labels above the points
  scale_fill_manual(values = team_colors) +  # Assign team colors
  geom_density_2d() +
  labs(title = "Density Plot of Career Completion Percentage vs ANY/A for
Quarterbacks",
       x = "Completion Percentage",
       y = "ANY/A")
```




```r
combined_data_PASR <- left_join(top_qbsQBR_with_records, qb_CMP, by =
"player_name")#passer rating

#Completion percentage v Passer rating
ggplot(combined_data_PASR, aes(x = CMPPercentage, y = passer_rating)) +
  geom_point(size = 3, shape = 21, color = "black") +
  geom_text(aes(label = player_name), vjust = -0.5, hjust = 0.5, size = 3) +
# Add labels above the points
  scale_fill_manual(values = team_colors) +  # Assign team colors
  geom_density_2d() +
  labs(title = "Density Plot of Completion Percentage vs Passer Rating for
Top Quarterbacks",
       x = "Completion Percentage (%)",
       y = "Passer Rating")
```

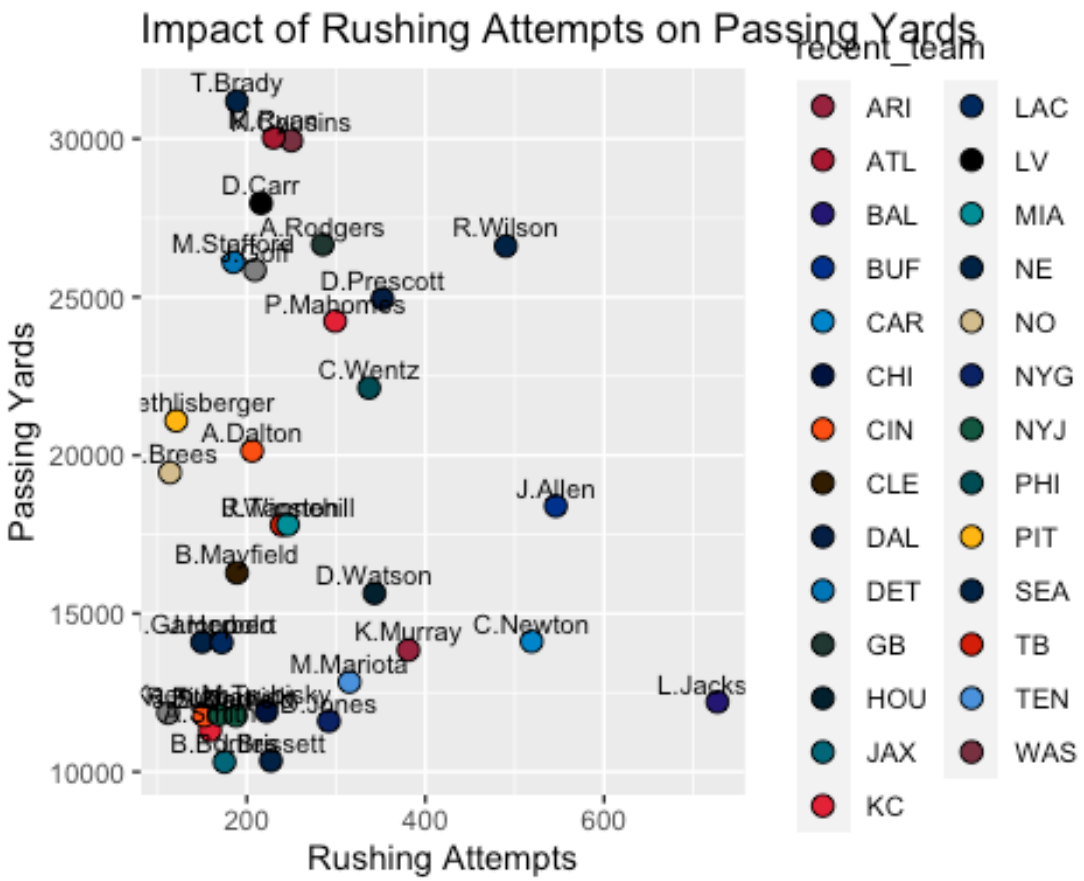


Rushing Analysis

```
qbRushers <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
  filter(season_type == "REG" & position == "QB") %>%
  group_by(player_name) %>%
  filter(sum(attempts) > 554) %>% summarize(passing_yards =
sum(passing_yards),rushing_attempts = sum(carries),
recent_team = recent_team, rushing_yards = sum(rushing_yards), rushing_tds =
sum(rushing_tds))
```

```
## Warning: Returning more (or less) than 1 row per `summarise()` group was
deprecated in
## dplyr 1.1.0.
## ℹ Please use `reframe()` instead.
## ℹ When switching from `summarise()` to `reframe()`, remember that
`reframe()`
##   always returns an ungrouped data frame and adjust accordingly.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

## `summarise()` has grouped output by 'player_name'. You can override using
the
## `.groups` argument.
```

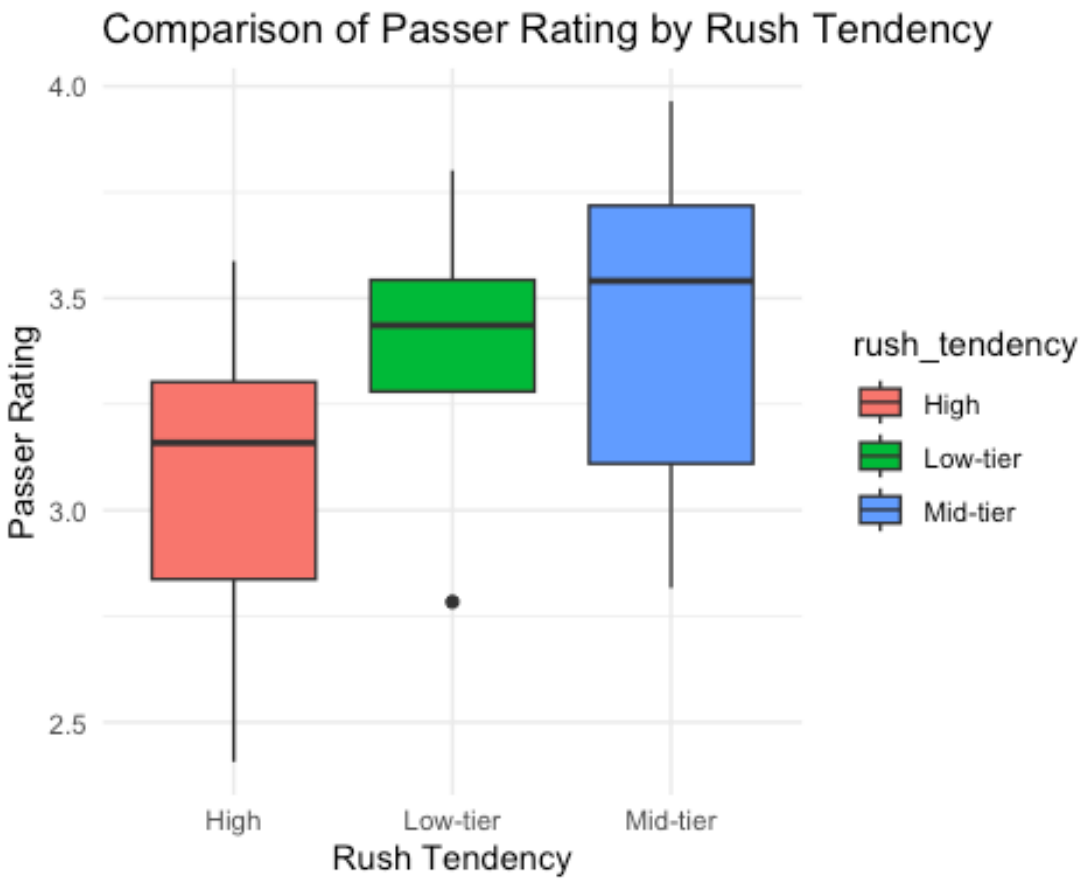
```r
qbRushers <- qbRushers %>% distinct(player_name, .keep_all = TRUE) #keep
nonduplicates
#Scatter Plot
qbRushers <- qbRushers %>% filter(rushing_attempts >= 100 & passing_yards >=
10000)
ggplot(qbRushers, aes(x = rushing_attempts, y = passing_yards,label =
player_name,fill = recent_team)) +
  geom_point(size = 3, shape = 21, color = "black") +
  geom_text(aes(label = player_name), vjust = -0.5, hjust = 0.5, size = 3) +
# Add labels above the points
  scale_fill_manual(values = team_colors) +  # Assign team colors
  labs(title = "Impact of Rushing Attempts on Passing Yards",
       x = "Rushing Attempts",
       y = "Passing Yards")
```




```r
qbRushers <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
  filter(season_type == "REG" & position == "QB") %>%
  group_by(player_name) %>%
  filter(sum(attempts) > 554)

qb_names <- c("L.Jackson", "J.Allen", "C.Newton", "K.Murray", "J.Hurts",
"R.Wilson", "J.Fields",
              "D.Jones", "D.Watson", "P.Mahomes", "D.Prescott", "C.Wentz",
```

```
"K.Cousins", "E.Manning",
               "M.Ryan", "D.Carr", "A.Rodgers", "T.Brady")


QBR_Rushing <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
  filter(season_type == "REG" & position == "QB") %>%
  group_by(player_name) %>%
  filter(sum(attempts) > 554 & player_name %in% qb_names) %>%
  summarize(
    completions = sum(completions),
    attempts = sum(attempts),
    passing_yards = sum(passing_yards),
    passing_tds = sum(passing_tds),
    interceptions = sum(interceptions),

    # Calculate individual components
    C = (completions / attempts) * 100,
    Y = passing_yards / attempts,
    TD = (passing_tds / attempts) * 100,
    INT = (interceptions / attempts) * 100,

    # Calculate Passer Rating
    passer_rating = ((C - 30) / 20) + ((Y - 3) / 4) + (TD / 5) - (INT / 4),
# Assign rush-tendency labels
    rush_tendency = case_when(
      player_name %in% c("L.Jackson", "J.Allen", "C.Newton", "K.Murray",
"J.Hurts", "R.Wilson", "J.Fields") ~ "High",
      player_name %in% c("D.Jones", "D.Watson", "P.Mahomes", "D.Prescott",
"C.Wentz", "K.Cousins") ~ "Mid-tier",
      player_name %in% c("E.Manning", "M.Ryan", "D.Carr", "A.Rodgers",
"T.Brady") ~ "Low-tier",
      TRUE ~ "Unknown"
    ),
    recent_team = first(recent_team)
  )

## Warning: Returning more (or less) than 1 row per `summarise()` group was
deprecated in
## dplyr 1.1.0.
## ℹ Please use `reframe()` instead.
## ℹ When switching from `summarise()` to `reframe()`, remember that
`reframe()`
##   always returns an ungrouped data frame and adjust accordingly.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

## `summarise()` has grouped output by 'player_name'. You can override using
the
## `.groups` argument.
```

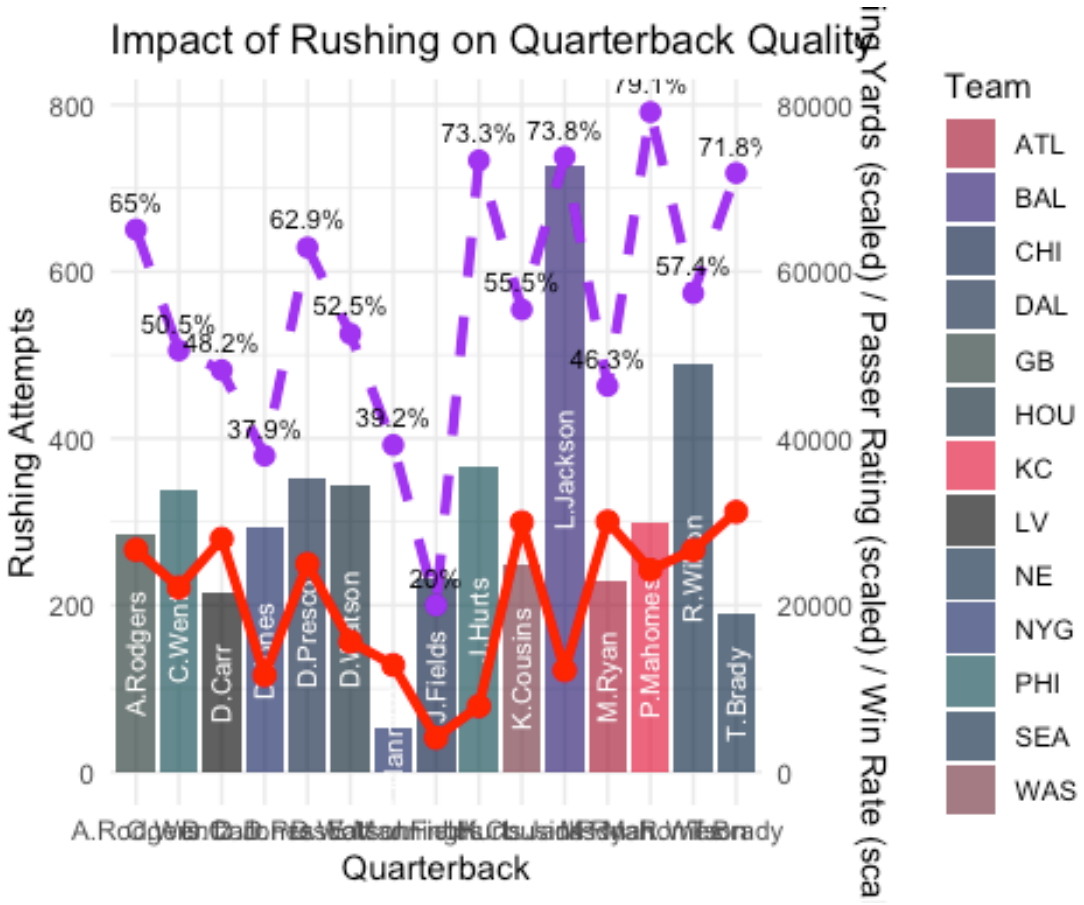
```r
QBR_Rushing <- QBR_Rushing %>% distinct(player_name, .keep_all = TRUE) #keep
nonduplicates


ggplot(QBR_Rushing, aes(x = rush_tendency, y = passer_rating, fill =
rush_tendency)) +
  geom_boxplot() +
  labs(title = "Comparison of Passer Rating by Rush Tendency",
       x = "Rush Tendency",
       y = "Passer Rating") +
  theme_minimal()
```




```r
#Dual Axis

qb_records <- rbind(qb_records,
                    data.frame(player = "J.Fields", wins = 5, losses = 20),
                    data.frame(player = "D.Jones", wins = 22, losses = 36),
                    data.frame(player = "C.Wentz", wins = 46, losses = 45),
                    data.frame(player = "E.Manning", wins = 20, losses = 31),
                    data.frame(player = "D.Carr", wins = 53, losses = 57)
)
```

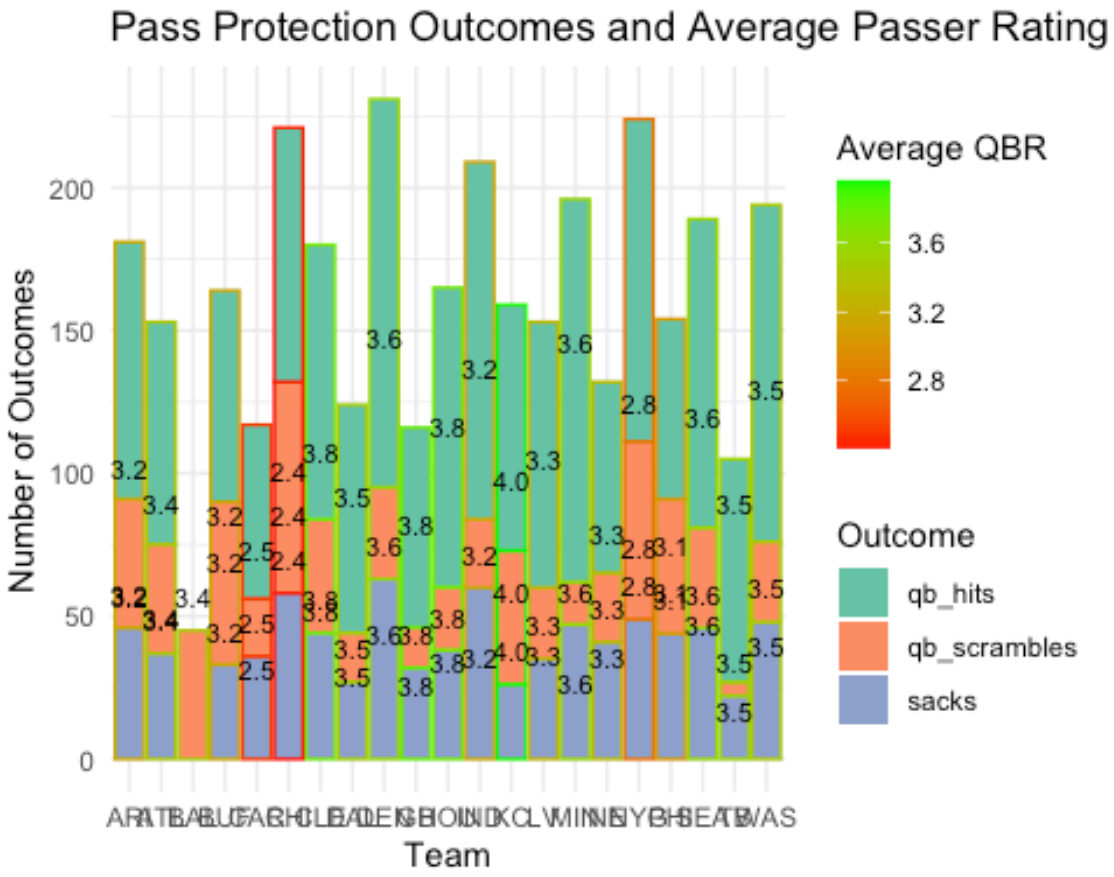
```r
qb_names <- c("L.Jackson", "J.Allen", "C.Newton", "K.Murray", "J.Hurts",
"R.Wilson", "J.Fields",
              "D.Jones", "D.Watson", "P.Mahomes", "D.Prescott", "C.Wentz",
"K.Cousins", "E.Manning",
              "M.Ryan", "D.Carr", "A.Rodgers", "T.Brady")
QBR_Rushing <- nflreadr::load_player_stats(seasons = 2016:2022) %>%
  filter(season_type == "REG" & position == "QB") %>%
  group_by(player_name) %>%
  filter(sum(attempts) > 554 & player_name %in% qb_names) %>%
summarize(rushing_attempts = sum(carries),    completions = sum(completions),
attempts = sum(attempts), passing_yards = sum(passing_yards),
passing_tds = sum(passing_tds),
interceptions = sum(interceptions),

# Calculate individual components
C = (completions / attempts) * 100,
Y = passing_yards / attempts,
TD = (passing_tds / attempts) * 100,
INT = (interceptions / attempts) * 100,
# Calculate Passer Rating
passer_rating = ((C - 30) / 20) + ((Y - 3) / 4) + (TD / 5) - (INT /
4),recent_team = recent_team)

## `summarise()` has grouped output by 'player_name'. You can override using
the
## `.groups` argument.

combinedQBRWins <- inner_join(QBR_Rushing, qb_records, by = c("player_name" =
"player"))
combinedQBRWins <- combinedQBRWins %>% distinct(player_name, .keep_all =
TRUE) #keep nonduplicates


ggplot(combinedQBRWins, aes(x = player_name)) +

  # First axis (left) for rushing attempts
  geom_bar(aes(y = rushing_attempts, fill = recent_team), stat = "identity",
alpha = 0.7) +
  scale_fill_manual(values = team_colors) +

  # Add player names to the bars on the left axis
  geom_text(aes(y = rushing_attempts, label = player_name),
            position = position_stack(vjust = 0.5),
            color = "white",
            size = 3,
            angle = 90,
            hjust = 0.5) +

  # Second axis (right) for passing yards
  geom_line(aes(y = passing_yards * 0.01), color = "red", size = 1.5, group =
```

```r
1) +
  geom_point(aes(y = passing_yards * 0.01), color = "red", size = 3, shape =
19) +

  geom_point(aes(y = (wins / (wins + losses)) * 100 * 10), color = "purple",
size = 3, shape = 16) +
  geom_line(aes(y = (wins / (wins + losses)) * 100 * 10), color = "purple",
size = 1.5, linetype = "dashed", group = 3) +

  # Add win-loss percentages above the points
  geom_text(aes(y = (wins / (wins + losses)) * 100 * 10, label =
paste0(round((wins / (wins + losses)) * 100, 1), "%")),
            color = "black",
            vjust = -1,
            hjust = 0.5,
            size = 3) +

  # Add labels and titles
  labs(
    title = "Impact of Rushing on Quarterback Quality",
    y = "Rushing Attempts",
    x = "Quarterback",
    fill = "Team",
    color = "Passing Yards (scaled) / Passer Rating (scaled) / Win Rate
(scaled)"
  ) +

  # Set the y-axis labels for the second axis
  scale_y_continuous(sec.axis = sec_axis(~.*100, name = "Passing Yards
(scaled) / Passer Rating (scaled) / Win Rate (scaled)")) +

  # Customize appearance
  theme_minimal()+  scale_linetype_manual(name = "Metrics", values =
c("solid", "dashed", "solid"), labels = c("Passing Yards", "Passer Rating",
"Win Rate"))
```




```
#O-line Quality
pbp <- nflreadr::load_pbp(2022) %>%
  filter(season_type == "REG") %>%
  mutate(qbh_inc = ifelse(qb_hit == 1 & incomplete_pass == 1, 1,0),
         qb_int = ifelse(qb_hit == 1 & interception == 1, 1,0))

nfl_plays <- pbp %>%
  group_by(week, posteam) %>%
  summarize(
    total_plays = sum(play_type == "run" | play_type == "pass", na.rm = TRUE)
  )

## `summarise()` has grouped output by 'week'. You can override using the
## `.groups` argument.

# Calculate Total Line Value
nfl_line_value <- pbp %>%
  filter(
    sack == 1 | qb_hit == 1 | qb_scramble == 1
  ) %>%
  group_by(posteam) %>%
  summarize(
    sacks = sum(sack),
```

```r
    qb_hits = sum(qb_hit),
    qb_scrambles = sum(qb_scramble)
  )
pass_protection_long <- nfl_line_value %>%
  pivot_longer(cols = c(sacks, qb_hits, qb_scrambles), names_to = "Outcome",
values_to = "Count")

team_qbr_completion <- QBR_Rushing %>%
  group_by(recent_team) %>%
  summarize(
    avg_qbr = mean(passer_rating, na.rm = TRUE),
    avg_completion_pct = mean(C, na.rm = TRUE)
  )

pass_protection_qbr_pct_team <- merge(pass_protection_long,
team_qbr_completion, by.x = "posteam", by.y = "recent_team")

ggplot(pass_protection_qbr_pct_team, aes(x = posteam, y = Count, fill =
Outcome, color = avg_qbr)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = c("#66c2a5", "#fc8d62", "#8da0cb"), name =
"Outcome") +
  scale_color_gradient(low = "red", high = "green", name = "Average QBR") +
  geom_text(aes(label = sprintf("%.1f", avg_qbr), y = Count + 5), color =
"black", vjust = -0.5, size = 3) +  # Display QBR values above the bars
  labs(
    title = "Pass Protection Outcomes and Average Passer Rating by Team",
    x = "Team",
    y = "Number of Outcomes"
  ) +
  theme_minimal()
```


Pass Protection Outcomes and Average Passer Rating
Number of Outcomes
0
50
100
150
200
Team
Average QBR
3.6
3.2
2.8
Outcome
qb_hits
qb_scrambles
sacks