Time to Throw in the National Football League

Vaughn Hajra

Professor Brad Hartlaub

December 14, 2023

Background and Project Overview

In 2018 the National Football League (NFL) and Amazon Web Services (AWS) introduced a new partnership: Next Gen Stats. These stats use machine learning paired with tracking cameras to capture new football statistics, such as defensive pressure rates, explosive plays, and more. In the 2023 season, Amazon also released a new broadcast, dubbed "Prime Vision". This broadcast tracked player routes and speed in real-time, as well as highlighting key players pre-play. Prime Vision has yet to catch on as a mainstream broadcast option, but to many, it signals a new wave of machine-learning-assisted stats.

Because this cutting-edge data is so new, much of it is yet to be explored. For example, a search of "NFL Time to Throw" on Google yields only one Pro Football Focus article (with the goal of selling their premium data subscription), a Reddit post consisting of a single scatterplot, and a plethora of fan blogs giving their best attempt at making sense of the "Time to Throw" statistic. Although these blogs and articles do provide useful exploratory analyses, there are little to no formal statistical tests being performed related to *time to throw* (TT) that are made easily accessible to the public.

This has led me to make time to throw in the NFL central to this research project. Specifically, I address the research questions of:

1. What is the relationship between time to throw and intended air yards for quarterbacks?
2. Does holding the ball longer than expected negatively affect quarterback ratings?
3. When using a clustering algorithm to group quarterbacks, what natural groupings emerge? How do these groups compare with each other?

I will answer these questions using both parametric and nonparametric statistical methods, and compare and contrast the results when using the two approaches. In addition to the methods we have discussed in class, I will also implement Density-Based Spatial Clustering of Applications with Noise (DBSCAN) as the clustering method of choice.

There are three main components of this paper, each related to one of the primary research questions. The first section employs tests for both correlation and comparison of groups. The second section implements various linear regression techniques, as well as testing for both location shift and dispersion differences between two distributions. Finally, the third section will use DBSCAN, and then compares multiple groups with general and ordered alternatives. Although there are multiple questions of interest for this project, time to throw and quality of quarterback play remain of central interest. Each question examines time to throw from a new perspective, with the goal of better understanding its significance in today's National Football League.

Data: Source, Wrangling, and Explanation

The data for this project was obtained from Amazon Web Service's (AWS) *Next Gen Stats*. The data ranges from 2018-2022, encompassing all of the time that AWS has tracked advanced statistics[1]. The specific data used in this project are the passing (quarterback) and receiving (wide receiver and tight end) advanced stats.

AWS provides these advanced statistics for quarterbacks who meet the following criteria: at least 15 pass attempts, multiplied by the number of weeks in a season, divided by two. This roughly translates to quarterbacks who play at least a quarter of the season. Similarly, to be included as a receiver in the data, players must have five targets (i.e. be the intended receiver for the pass), multiplied by the number of weeks, divided by two. Although the AWS method of selecting players may seem complicated, it helps account for the longer seasons that began in 2021. In this project, I use all of the players that meet AWS' criteria over the five-year span; this results in a sample size of 197 quarterbacks and 631 receivers.

Although the quarterbacks are the central focus of this study, including receiver data is important to control for the quality of players that the quarterbacks are passing to. If a quarterback is throwing to an elite receiver, poor throws may still be caught. On the other hand, if a receiver is especially bad at their job, even the best quarterbacks can struggle to win (just ask Patrick Mahomes). Because of this, I calculate a weighted average for the following variables: receiver separation, defensive back cushioning, catch percentage, and targeted air yardage. Target share (how many times a receiver is being passed to relative to their team) is calculated and used as the weights for these averages[2].

Finally, I use a many-to-one merge to add team receiving averages to each quarterback in the dataset. Although this merge is not perfect (multiple quarterbacks on the same team and year have the same receiving averages), I am confident that the weighted averages are adequate control variables for this project.

The final dataset consists of the 197 quarterbacks, their advanced statistics, and their team's weighted average receiving stats. The central variables of interest are Time to Throw (TT), and Quarterback Rating (QBR). When specifying models, variables will be explained but for summary statistics, see *Table A1* in the appendix of this paper.

---

[1] The incomplete 2023 season was not included as it is still ongoing at the time of writing.

[2] Total team target share is calculated by summing the number of total targets for each team in a given year. A receiver's percentage of targets relative to the team total is then used as the weight for that player when calculating their average.

Part 1: Time to Throw, Intended Air Yards, and Quarterback Rating

When considering time to throw, another advanced statistic is often also considered: intended air yards[3] (IAY). Pass catchers take time to get downfield, and one would expect that the farther downfield a quarterback is throwing, the longer it would take before the ball is thrown. Because of this, I first ran a correlation test between average intended air yards and average time to throw. When using the parametric approach (Pearson's Correlation Coefficient, *r*), the correlation estimate of 0.422 was obtained. This was accompanied by a 95% confidence interval of 0.300 to 0.531, and because zero is not included in the confidence interval we have sufficient evidence to show that the two variables are correlated with each other. When running the nonparametric approach (Kendall's tau) we are given the correlation estimate of 0.31 and a 95% confidence interval of 0.226 to 0.395.

Although both estimates are similar, Kendall's tau finds a weaker correlation between the two variables. This result may be because Kendall's tau is less sensitive to outliers and there are a few extreme points of long time to throw and high IAY. That being said, both estimates find a significantly positive relationship between average intended air yards and time to throw, which aligns with our intuitive expectations.

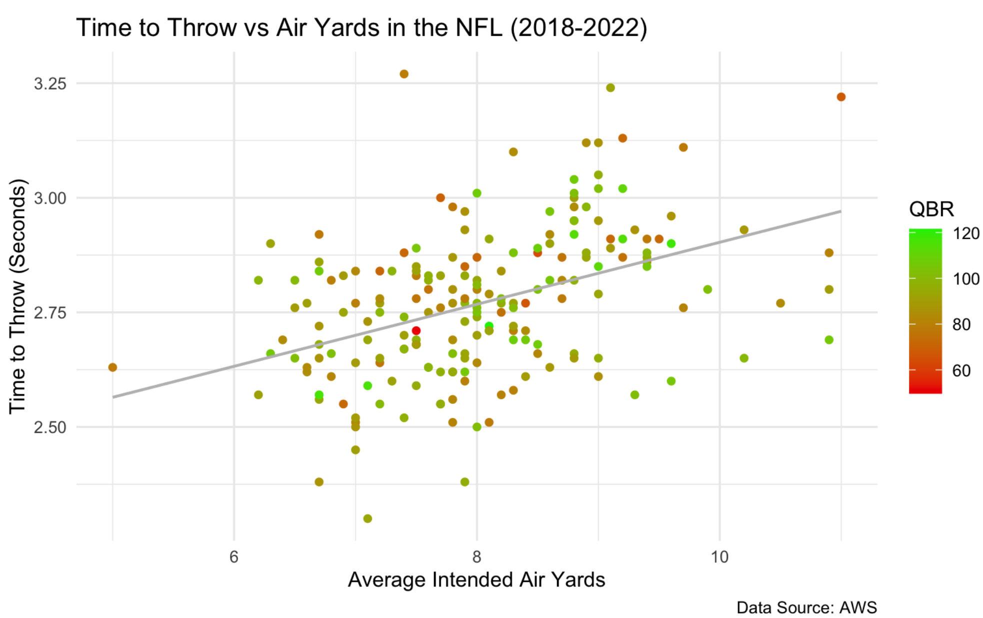

Figure 1. Time to Throw (TT) vs Intended Air Yards (IAY) in the NFL (2018-2022). This scatterplot shows the relationship between TT and IAY for all 197 quarterbacks in the sample. Additionally, the points are colored by quarterback rating (QBR). A line of best fit is also on the graph, highlighting the positive relationship between TT and IAY.

[3] In the case of this analysis, I am using average intended air yards for a quarterback in a given year.

When a third variable, quarterback rating (QBR) is factored in, a new theoretical framework must be developed. I am specifically interested in the combination of IAY and TT that gives the best quarterback rating. In a blog post by T. Troy Russell, he claims that the biggest issues arise for quarterbacks when they're taking a long time to throw short or a short time to throw long. I agree with this reasoning because it makes sense that discrepancies in time to throw and intended air yards often are due to "broken" plays that don't work out as intended.

Russell, in his blog article, claims that there should be two groups of quarterbacks: those that have discrepancies with how long they're taking to throw and those that don't. I would argue that there should be four groups, split up by throwing short/long and being quick/slow to throw. This would give us short/quick, short/slow, long/quick, long/slow. In R, I divided these players into four groups, and the number of quarterbacks in each group is as follows:

TT	Quick		Slow	
IAY	Short	Long	Short	Long
Count	67	37	32	61

Table 1. Counts for quarterback Time to Throw and Intended Air Yards groups. Groups were divided using the median of each variable (which was very similar to the mean, resulting in roughly the same number of observations in each group).

When examining the counts for each group, an interesting observation arises: the data are not evenly distributed between each of the four groups. Upon further inspection, a possible explanation for this is made. As I previously mentioned, one would expect consistent quarterbacks to generally be better, and these quarterbacks are likely the ones with the starting jobs. Because AWS requires a minimum threshold of passes thrown to be included in the dataset, we may be missing a slew of backup quarterbacks who might make up more of the conflicting groups. If it is possible to get more data on these quarterbacks (potentially by paying for Pro Football Focus' premium subscription or finding a way to get the entire AWS database) using **all** passers would be an improvement.

Alternatively, there is also a chance that these inconsistent quarterbacks aren't playing in games because they are inconsistent. They are not as good at the job, and because of their inconsistencies, they don't get to see the field nearly as often. Regardless of this, there are enough observations in each group where I feel comfortable evaluating them. To test if there are any differences in the groups, I will use the Analysis of Variance (ANOVA, parametric) and Kruskal-Wallis (nonparametric) procedures. The models for each are as follows:

$$\underline{\text{ANOVA model:}}\ QBR_{ij} = \mu + \alpha_j + \varepsilon_{ij},\ where\ \varepsilon \sim N(0, \sigma^2)$$

μ refers to the grand mean, α refers to the group effect, and ε is the error term.

$$\underline{\text{Kruskal-Wallis model:}}\ QBR_{ij} = \theta + \alpha_j + \varepsilon_{ij},$$

where ε has zero median, and forms a continuous distribution.

θ refers to the overall median, α refers to the group effect, and ε is the error term.

Both of these models have j = 4 groups. Note that a key difference between the two models is that ANOVA uses the grand mean while Kruskal-Wallis uses the grand median. Additionally, ANOVA relies on a normal distribution of error, while Kruskal-Wallis requires a continuous distribution. For ANOVA, the null hypothesis for this model is that there is no difference in the mean quarterback rating for each group ($\alpha_j = 0$). The alternative hypothesis is that not all the treatment means are the same (at least one α_j != 0). For Kruskal-Wallis, the null hypothesis for this model is that there is no difference in the median quarterback rating for each group ($\alpha_j = 0$). The alternative hypothesis is that not all the treatment medians are the same (at least one α_j != 0).

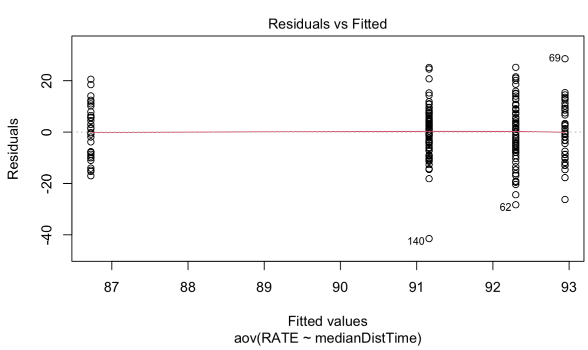

Figure 2: *Residuals vs Fitted for the ANOVA model*

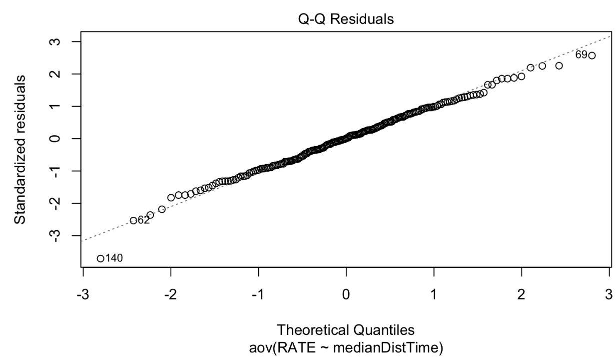

Figure 3: *Normal Q-Q plot for the ANOVA model*

When running ANOVA, the assumption of a normally distributed error term is met, with a zero mean. That being said, it is also appropriate to run the Kruskal-Wallis test because the error distribution for that model also satisfies the zero median and continuous distribution conditions. When running the tests themselves, we are given very similar results. For the ANOVA, we have df = (3, 193), F = 2.17, and p = 0.0929. For Kruskal-Wallis we have chi-squared = 6.5, df = 3, and p = 0.089. In both of these cases, we fail to reject the null hypotheses. In the case of ANOVA, we do not have sufficient evidence that there are significant differences in the means of the different groups' quarterback ratings. In the case of the Kruskal-Wallis test, we do not have sufficient evidence to conclude that the groups' quarterback ratings have significantly different medians.

Although I was planning to do multiple comparisons with ordered alternatives for these groups, it is not an appropriate procedure to run because we failed to reject the null hypothesis for both tests at the five percent level. This was a surprising result to me, and my feeling was that the result was due to the grouping being subpar rather than the differences being nonexistent. To summarize the current section, I have found a positive correlation between intended air yards and time to throw, but no significant differences in four groups of quarterbacks (created by dividing observations at the median of the two variables). This led me to tackle a different grouping method and provided the basis for my subsequent research question.

Part 2: Time to Throw, and Quarterback Rating using the Regression Residual

After some background research related to my perceived grouping problem, I decided to take inspiration from the financial world. Events studies are used in finance to track stock portfolios and their performance after distinct events. Essentially, analysts use the best regression analysis possible to predict stock returns and classify portfolios as "overperforming" or "underperforming" based on the abnormal returns (what actually happened vs. what was expected to happen).

This framework, paired with my results in the previous section, led me to wonder if I can better classify passers as "quick" or "slow" to throw the ball by using multiple variables. From the correlation analysis of the previous section, we know time to throw and intended air yards are positively correlated, and, through more exploratory analysis, I found other advanced stats that are also correlated with time to throw. Specifically, I aimed to create groups using a similar method to an events study, and then compare the quarterback rating of the two newly created groups.

The goal then became the following: find the best linear regression model that I could to predict time to throw, **without using any variables directly related to quarterback rating**. Then, use the residual of the regression model to classify quarterbacks either as quicker or slower than expected. Finally, I will compare these two groups of quarterbacks using both parametric and nonparametric methods.

Before jumping straight into multiple linear regression, I wanted to use simple linear regression as a starting point for subsequent regression analysis. My first model predicted time to throw (TT) using intended air yards (IAY), and the parametric models are as follows:

Parametric Approach: $TT = \beta_0 + \beta_1 IAY + \varepsilon$

Nonparametric Approach: $TT = \alpha + \beta_1 IAY + \varepsilon$

Where β_0 and α are the intercepts, and β_1 is the coefficient for intended air yards.

When fitting the model, we find that both the intercept and IAY coefficient are significant and very similar using both the parametric and nonparametric approaches. Each result is the same value to the nearest hundredth, and therefore for this application, the models are essentially interchangeable. The coefficient of interest, IAY, is significantly different from zero with t = 6.5 and p = roughly zero. This result makes intuitive sense, and the positive coefficient also aligns with what I found in the previous section (correlation test). To see residual plots for this analysis, please see the appendix.

After running simple linear regression, I moved to incorporate more variables to better predict time to throw. I chose variables that I thought may impact time to throw but are not directly used in calculating quarterback rating[4]. These variables are intended air yards (IAY), number of attempts(ATT), year (YEAR, a factor to control for potential changes in the NFL over time), completed air yards (CAY, how far the quarterback is actually completing the ball), and two controls for receiver quality. The receiver controls are average cushioning that the defensive backs give receivers (CUSHavg) and average receiver separation (SEPavg). I also investigated potential quadratic and other relationships between these variables but found that the linear relationships are the best to use for this regression.

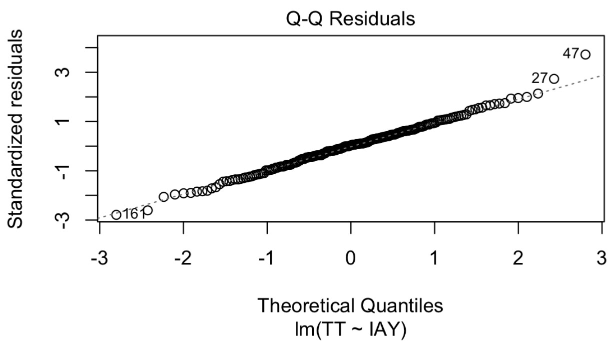

Figure 4*: Normal Q-Q plot for the parametric multiple linear regression model*

The full parametric model, using all of these variables is as follows:

$$TT = \beta_0 + \beta_1 IAY + \beta_2 CUSHavg + \beta_3 SEPavg + \beta_4 ATT + \beta_5 CAY + \beta_{6-9}(YEAR) + \varepsilon$$

Where the variables are as specified above, and β_{6-9} represents four indicator variables (one for each year after 2019, with 2018 being the base year). After running this initial regression, and also running the nonparametric equivalent of it, I decided to move forward with the nonparametric multiple linear regression models. I decided to use the nonparametric approach because the results were extremely similar, but because there were a number of outliers in the data the nonparametric approach would be less sensitive to those outliers.

---

4 Quarterback rating is calculated using yards, completions, touchdowns, and interceptions (all per attempt).

Although I thought defensive back cushioning (how much space receivers have pre-play) and receiver separation may have an impact on time to throw on an individual play-by-play basis, I was unsure if they would significantly contribute to the model when looking at yearly data. This thought arose because much of this comes from the defense's skills rather than the receivers themselves, and when playing many games throughout the season it may lose some of its importance. To test for this, I ran a drop in dispersion test.

The hypotheses for this drop in dispersion test are as follows:

Ho: All of the receiver variables in the model are zero in the presence of the quarterback stats
Ha: At least one of the receiver variables in the model is non-zero in the presence of the quarterback stats

After running this test, we are given an F value of 12.4 and a p-value of roughly zero. This means that in the presence of the receiver statistics, at least one of the other coefficients in the model has a nonzero coefficient. Upon further inspection, I found that separation **is** a significant contributing factor in the model (p = roughly zero), while cushioning is **not** a significant contributing factor (p = 0.92) when looking at the individual t-tests for variables. In addition to this finding, I also found that the year dummy variables and completed air yards were not worth including in the model, using a mix of (manual) backwards elimination and my own reasoning. After removing the year variables, and the cushioning variable, the final model for the nonparametric multiple regression is as follows:

$$TT = \alpha + \beta_1 IAY + \beta_2 SEPavg + \beta_3 ATT + \varepsilon$$

Where alpha is the intercept. The fitted model is:

$$\widehat{TT} = 21.66 + 0.074 * IAY + 0.20 * SEPavg - 0.00015 * ATT$$

This test has all variables significant at the 5 percent level, with IAY and SEPavg also significant at the one percent level. The robust R-squared for this model is 0.249 and when considering how little information the model is being provided (only three predictor variables and no defensive statistics), I am generally pleased with these results.

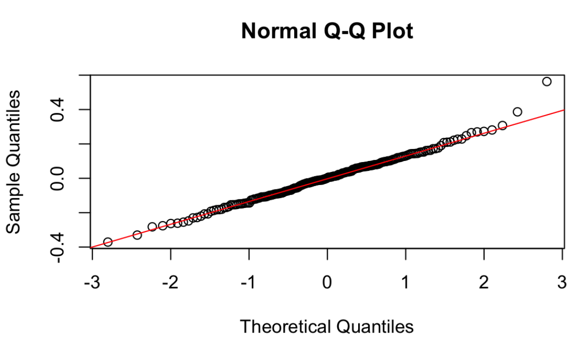

Figure 5*: Normal Q-Q plot for the final nonparametric multiple linear regression model*

Now that the "best" regression model for my needs has been obtained, I calculate the residuals using the fitted model and classify passers as *quick* if they have a negative residual (throwing faster than the model would predict) and *slow* if they have a positive residual[5] (throwing slower than the model would predict). This classification was as evenly split as possible, with 99 of the 197 quarterbacks being classified as *quick* and 98 being classified as *slow*.

Based on my background knowledge of football, I would expect players with a "quick release" to be better than the *slow* group. I also was interested in seeing if there was a difference in dispersion between the two groups. After these groups were created, the next step to further investigate this was to take a look at the two distributions.

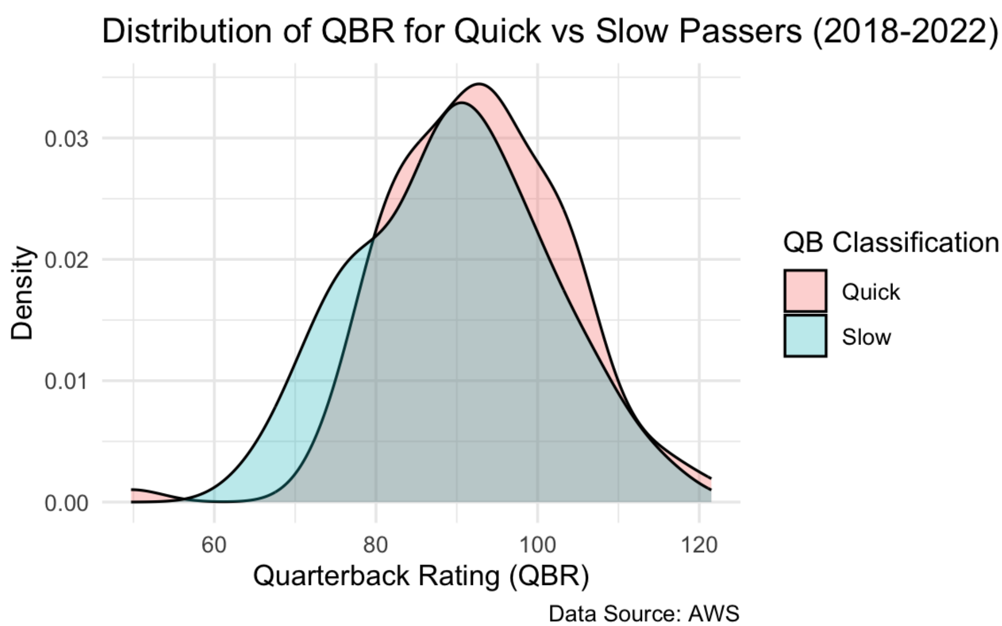

Figure 6. Distribution of Quarterback Rating (QBR) for quick and slow passers in the NFL (2018-2022). This scatterplot shows the distributions of the two groups of quarterbacks in the NFL, classified as *quick* or *slow* based on the residual multiple linear regression model. *Quick* passers have a mean QBR of 92.65, while *slow* passers have a mean QBR of 89.

To test the for a location shift, I used the parametric Welch two-sample t-test (means), and the nonparametric Wilcoxon rank sum test (medians), with the following hypothesis:

Ho: There is no difference in mean/median QBR between quick and slow passers
Ha: The difference in mean/median QBR between quick and slow passers is greater than zero

For the t-test, I found t = 1.90, df = 194, and p = 0.030. The Wilcoxon rank sum test gave W = 5,614.5 and a p-value of 0.028. Both of these tests gave the same result, at the five percent

[5] Note that this classification of "quick" and "slow" by using the residual of the linear model is different from the first section that used the median for classification.

significance level: we have sufficient evidence that the true location shift is greater than zero for the two groups. This is a different result than found when comparing the IAY and TT groups in the previous section, and I feel much more confident in this grouping method as it accounts for more information (still unrelated to QBR) when creating the groups.

Next, I was interested in testing for dispersion between the two groups. For the parametric approach, I used the F-test to compare two variances, and for the nonparametric approach, I employed the Miller Jackknife procedure (this was chosen because we know the medians are not equal but we still want to test for differences in dispersion). The null hypothesis for the F-test is that the true ratio of variances is equal to one, and the two-sided alternative hypothesis is that the true ratio of variances is **not** equal to one. When running the test in R, we are given $F = 1.12$, numerator $df = 97$, denominator $df = 98$, and $p = 0.57$. For the Miller Jackknife procedure, we are given the Jackknife estimator equal to 1.1 and $p = 0.66$. Based on these results, the nonparametric and parametric approaches both agree that we fail to reject the null hypothesis and that there is not sufficient evidence to reject the null hypothesis that the two distributions have differences in dispersion.

To recap the findings of this section, I first used simple and then multiple linear regression to best predict time to throw for quarterbacks in the NFL, and then used the final multiple linear regression to predict the fitted values. Then, I used the residual of the actual minus predicted values to group quarterbacks as either slow or quick passers. Analyzing these two groups, I found a significant difference in quarterback rating and no significant difference in variance.

Part 2: Clusters and Time to Throw

After being much more satisfied with the conclusions in the second part, I wanted to address the final research question:

When using a clustering algorithm to group quarterbacks, what natural groupings emerge? How do these groups compare with each other when considering time to throw?

For this analysis, I am interested in the **opposite** framework of the previous sections. What happens if I give a clustering algorithm many factors directly related to quarterback rating, and then compare the time to throw for those groups? Is there a specific time to throw range that the most successful quarterbacks have in common?

Before doing the actual clustering itself, I had to decide on a clustering algorithm to use. This led me to explore if there were nonparametric clustering algorithms, and why they were categorized as nonparametric. I found that nonparametric clustering algorithms are ones that "operate by considering the number of clusters unknown" while parametric clustering requires a pre-specified number of clusters. I identified a well-regarded algorithm, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), which was introduced in 1996 and won an award in 2014 for standing the test of time.

There are two big reasons why DBSCAN is the best option for this project. The first (and primary) reason is that it allows for an unspecified number of groups, and also allows for noise: not all observations have to be classified in a group. The secondary reason for using it is that there exists a well-regarded R package that provides all the necessary tools for implementation.

The algorithm functions by starting at a single point, moving to the nearest neighbor until there are no more neighbors within the specified distance, and then starting a new group. Additionally, it will only label a group as a cluster if it meets the minimum number of points specified. The clustering algorithm itself therefore takes two inputs: epsilon (to be used as the nearest distance) and the minimum number of points. To find an appropriate value for epsilon, K-nearest neighbors are used.

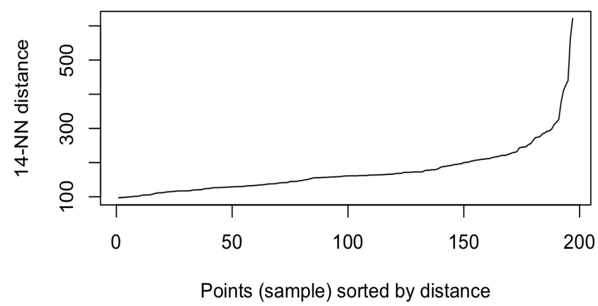

Figure 7. K-Nearest Neighbor Plot. This plot was used to decide on the appropriate epsilon value for the algorithm, using k-nearest neighbors. The square root of the sample size (roughly 14) was used as a starting point for the k value, as recommended as a starting point by existing literature[6]. The point at which the slope (roughly 150) changes was used as the value for epsilon for the clustering algorithm.

After determining the appropriate epsilon and minimum point values, the clustering algorithm was then implemented. This led to four distinct groups being classified, plus one group being not classified within a cluster. These groups, upon further inspection, had four different qualities of quarterbacks. Using my background knowledge of football, cluster two seemed to have the “best” passers, followed by cluster one, and then cluster four, and finally cluster three (which seemed to be mostly backup-tier quarterbacks). One shortcoming of this grouping method is that it does not take into account running skills (Lamar Jackson stands out as not being appropriately grouped), but because I am interested in specifically passing, the limitations in considering quarterback running ability are not a major concern for subsequent analysis.

---

[6] Literature was found using articles mentioned in a stack exchange answer originally linked here

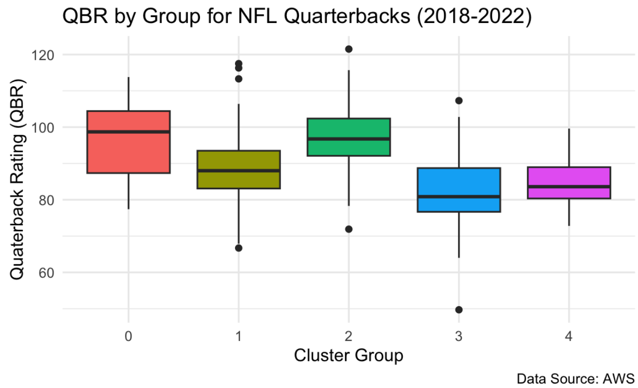

Figure 8. Quarterback Rating by Cluster. This plot shows the quarterback rating distribution for each cluster (1-4) in the sample of quarterbacks, with group zero not being clustered. Sample Sizes are 65 for cluster 1, 80 for cluster 2, 32 for cluster 3, and 10 observations in cluster 4.

First, to ensure that the groups did in fact have different quarterback ratings, I again employed the Kruskal-Wallis test. The null hypothesis for this test is that there is no difference in the median quarterback rating by group, and the alternative hypothesis is that at least one of the clusters has a different median QBR. When running this analysis[7], I found chi-squared = 52, df = 3, and p = roughly zero. This means that we have sufficient evidence of differences in at least one of the groups' QBR, at the one percent level. We reject the null hypothesis that there are no significant differences in the median QBR for the different clusters. I also ran ANOVA for this procedure, with the same hypotheses except using the median rather than the mean. Although ANOVA came to the same conclusion (df = (3, 183), F = 20.75, p = roughly zero), I am extremely cautious to proceed with parametric procedures. As seen in Figure 8, there are 8 points classified as outliers in the data, which means nearly one in twenty observations is an outlier. Because of this, I decided to only use nonparametric procedures going forward with this analysis.

In addition to the Kruskal-Wallis analysis, I was also interested in testing the ordered alternative, following my previous reasoning using my background knowledge. For this, I used the Jonkheere-Terpstra test for ordered alternatives, with the order from smallest to least median QBR being clusters 3, 4, 1, 2. This gave the same null hypothesis as the Kruskal-Wallis (no difference in median QBR between clusters) but a new alternative hypothesis:

$$H_2\colon \tau_3 \leq \tau_4 \leq \tau_1 \leq \tau_2 \text{ with at least one strict inequality}$$

When running the appropriate Jonkheere-Terpstra procedure (with ties, so p-values are based on a conditional null distribution), we find J = 8,759 and p = roughly zero[8]. This again aligns with our intuitive expectations and shows that at least two of the groups **do** have differing levels of quarterback rating.

---

[7] I only included data that were clustered in this analysis, which excludes group zero on the above boxplot
[8] Found using 10,000 Monte Carlo iterations

Now that it has been established that quarterback rating varies by cluster[9] I am interested in seeing if time to throw also varies. The goal of this is to investigate the idea that we know these groups have varying levels of success at passing, so will there be differences in how long it takes them to throw the ball? To address this question, we will again use the ANOVA (mean) and Kruskal-Wallis (median) tests. The hypotheses are as follows:

Ho: There is no difference in mean/median time to throw between clusters
Ha: There is at least one mean/median difference in time to throw between clusters

When running the ANOVA test, we are given df = (3, 183), F = 0.71, and p = 0.547. This means that we fail to reject the null hypothesis that there is no difference in the mean time to throw between clusters. Similarly, when we run Kruskal-Wallis we are given chi-squared = 1.47, df = 3, and p = 0.68. We again fail to reject the null hypothesis, this time we do not have sufficient evidence that the median time to throw is different between clusters. When examining the data in a box plot (Figure 9), our results align with the trends we see in the data.

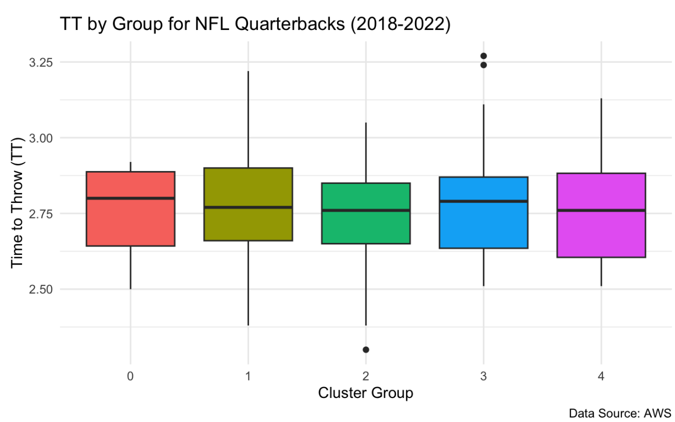

Figure 9. Time to Throw by Cluster. This plot shows the time to throw distribution for each cluster (1-4) in the sample of quarterbacks, with group zero not being clustered. Sample Sizes are 65 for cluster 1, 80 for cluster 2, 32 for cluster 3, and 10 observations in cluster 4.

To summarize this section, I first divided quarterbacks using DBSCAN. Then, I verified there were at least two groups distinct from each other with respect to quarterback rating. Finally, I found that although the quarterback rating was significantly different between groups, there was actually no difference in time to throw between any of the clusters. This non-significant result underscores a key finding: quarterbacks can find success at different times to throw, and there is no “right” timing that quarterbacks should aim for. I would argue that this means as long as a quarterback is throwing with appropriate timing within their offensive system, they have the potential to find success.

[9] Note that although multiple comparisons were considered for this analysis, I did not decide to include a multiple comparisons procedure here because the quarterback ratings are not central to the research question. I am more interested in potential differences in time to throw

Conclusions and Future Directions

Although time to throw is a relatively new and perplexing variable, I feel confident that this analysis can serve as a sound starting point for future research. In future studies, a major improvement would be to track which quarterbacks/receiver pairs were playing together on a game-by-game basis. This would be very time-consuming to do with well over 1,000 games to track, but could lead to a more robust dataset. Additionally, tracking offensive coordinators (play callers for the offense) could lead to more interesting findings to look at changes within the same offensive system. For example, San Francisco quarterback Brock Purdy has improved from 29th in time to throw to 6th, within the same offensive system. This has been a major improvement for the young quarterback, and finding ways to analyze these changes on a broader scale would be a very interesting future project.

When looking at all NFL quarterbacks from a yearly perspective though, it is much harder to find a “good” time to throw. In the first section, when comparing intended air yards and time to throw, I found no significant differences in quarterback ratings when grouping by time to throw and intended air yards. This finding is limited because it doesn’t take into account differences in offensive systems, but it still provides useful information going forward.

Subsequently, when considering more factors, I found that throwing the football faster than expected is related to a higher quarterback rating. Although this difference in QBR is not dramatic, it is significant and the little differences make up the biggest changes in professional football. Over the course of a season having a marginally better quarterback can greatly improve a team. Finally, I found that when clustering quarterbacks by quality of play, there is no difference in time to throw between low-quality and high-quality passers. This means that there is no “right” time to throw the ball for all quarterbacks in the NFL.

This begs the question, then, How important is time to throw when evaluating prospective additions to a football team? I would argue, **very important**. Because of its nuanced relationship with quality of play, it is not often used as an indicator of success by front offices. I would recommend isolating individual players within one offensive system, and inspecting how their time to throw changes over time. If they get faster throwing the ball, without increasing their interception rate, I would highlight that passer as a potential breakout player. Another addition to using this to evaluate players as a front office is that because it is not yet widely used, it could give teams an edge in both draft and free agent selection.

Works Cited:

The Chiefs' Wide Receiver Problem Is Now a Full-Blown Crisis, www.si.com/nfl/2023/12/11/chiefs-wide-receivers-emergency-patrick-mahomes-toney. Accessed 13 Dec. 2023.

"DBSCAN: Density-Based Clustering for Discovering Clusters in Large Datasets with Noise - Unsupervised Machine Learning." *STHDA*, www.sthda.com/english/wiki/wiki.php?id_contents=7940. Accessed 13 Dec. 2023.

News, SIGKDD. "Sigkdd News." *SIGKDD News : 2014 SIGKDD Test of Time Award*, www.kdd.org/News/view/2014-sigkdd-test-of-time-award. Accessed 13 Dec. 2023.

"Ngs: NFL next Gen Stats." Next Gen Stats, nextgenstats.nfl.com/stats/passing#yards. Accessed 13 Dec. 2023.

"Ngs: NFL next Gen Stats." Next Gen Stats, nextgenstats.nfl.com/stats/receiving#yards. Accessed 13 Dec. 2023.

Non-Parametric Clustering Using Deep Neural Networks | IEEE Journals ..., ieeexplore.ieee.org/document/9171232. Accessed 13 Dec. 2023.

Russell, T. Troy. "Fine, Let's Talk about Time to Throw." Stampede Blue, Stampede Blue, 16 Oct. 2021, www.stampedeblue.com/2021/10/16/22727432/fine-lets-talk-about-time-to-throw.

Wagoner, Nick. "Forget Regression, It Seems like 49ers' Brock Purdy Is Actually Getting Better." ESPN.Com, ESPN, 3 Oct. 2023, www.espn.co.uk/blog/san-francisco-49ers/post/_/id/42998/forget-regression-it-seems-like-49ers-brock-purdy-is-actually-getting-better.

Appendix

Variable	Abbreviation	Min	Q1	Median	Q3	Max	Mean	SD
Quarterback Rating	QBR	49.70	83.90	91.20	99.10	121.50	91.13	11.37
Time to Throw	TT	2.30	2.65	2.77	2.87	3.27	2.77	0.16
Attempts	ATT	131.00	297.00	441.00	544.00	733.00	419.54	155.39
Touchdowns	TD	2.00	11.00	19.00	26.00	50.00	19.56	10.42
Interceptions	INT	2.00	6.00	9.00	12.00	30.00	9.19	4.18
Receiver Separation	SEPavg	2.22	2.75	2.93	3.07	3.67	2.91	0.25
Receiver Cushioning	CUSHavg	4.59	5.68	5.95	6.18	6.98	5.93	0.40

Table A1. Summary statistics for key variables. Data is obtained from Amazon Web Services' Next Gen NFL Statistics, and there are 197 observations for each variable.

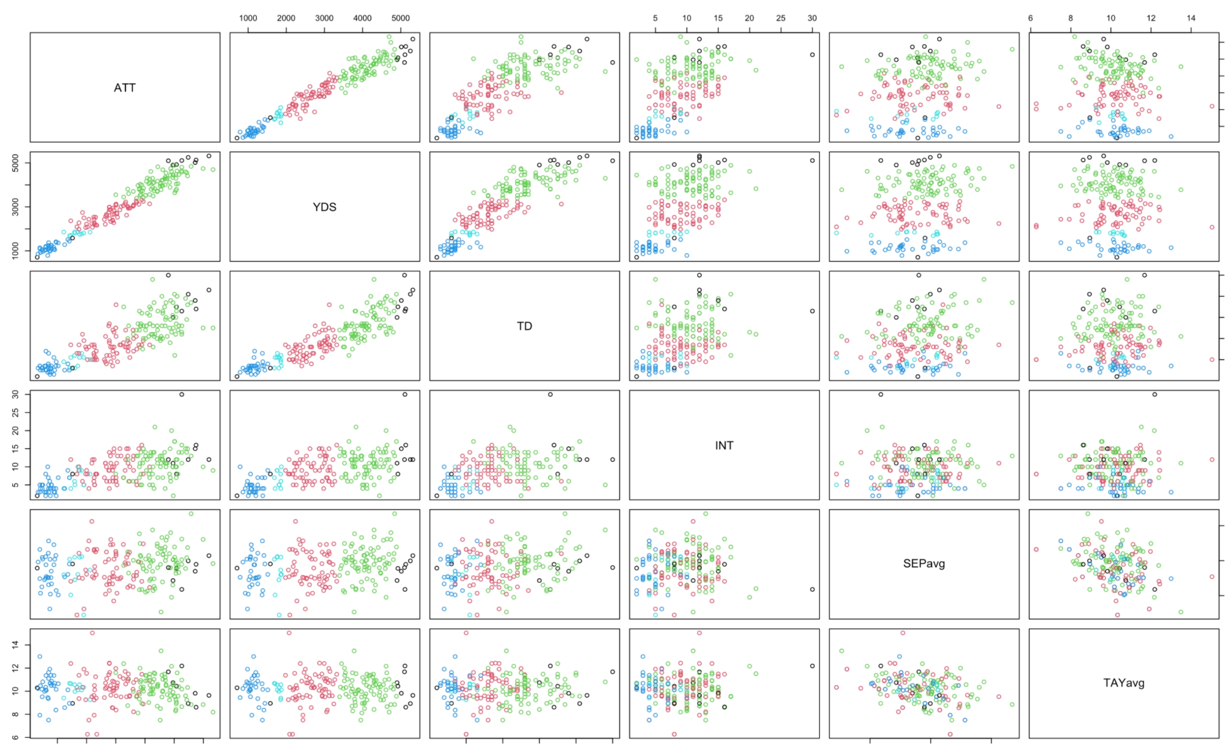

Figure A1. Cluster Plot. Data that was clustered from DBSCAN in section 3 of the above paper, using the above variables. A scatter plot of all the variables' relationships is shown, with colors representing four distinct clusters. Black points are treated as noise and not grouped into a cluster.