# Player Rating and Evaluation Systems for Ultimate Frisbee


## Abstract

Since the advent of sports analytics, developing novel methods to rate or rank players has been a popular challenge due to its merits and the simplicity of its results. Recognizing the limitations of simplistic statistics in capturing the complexities of team sports, many researchers have developed unique methodologies for measuring impact or performance. Although most major sports are saturated with these models, the rapidly growing sport of Ultimate Frisbee has largely lacked such detailed analysis. This project aims to fill that gap by developing two innovative methodologies for rating players. Using data from the UFA (formerly the AUDL), the largest semi-professional league in the sport, this research employs models inspired by on/off ratings in basketball and mixed-effects linear models in baseball. The results reveal intriguing patterns, challenging preconceived notions of top players and prompting questions about potential oversights and the discovery of previously unknown insights. These ratings are publicly available to enhance the experience of any fan or player. This research contributes to the evolving field of sports analytics, offering a fresh perspective on player evaluation in Ultimate Frisbee and sparking broader considerations for player rating methodologies across various sports.


## 1 Introduction

Sports analytics have revolutionized nearly everything about sports. From Moneyball in baseball to the three point revolution in basketball, analytics has changed everything from how teams are made, how players train, and strategies themselves.

Ultimate Frisbee, a sport invented at a New Jersey high school in 1968, has grown substantially in its first 65 years of existence. The sport now boasts a semi-professional league, which recently rebranded to the Ultimate Frisbee Association (UFA), with 24 teams across the United States and Canada. The sport has an estimated 7 million players worldwide, and has governing bodies in 56 countries. The sport has broad cross-gender appeal and is often touted as primed for growth over the coming years and decades. Its limited equipment and ease of learning to play which is often touted by USA Ultimate (About ultimate) and its ability to generate highlight plays which are often featured on ESPN's SportsCenter (UFA) help it continue to grow. Although many aspects of Ultimate have advanced in recent years, there has been a minimal amount of analytics work created, in large part due to lack of accessible data. While the data

collection surrounding the sport has improved, statistical analysis has been slow to develop. Like any other sport, Ultimate Frisbee has the potential to benefit from increased focus on analytics.

Although there has been some creation of analytics of ultimate frisbee analytics, most have focused on box score formulas. As is the case within any team sport, box score metrics can only capture so much. Other similar sports with complex team dynamics have benefited from including non-box score metrics that can better assess a players ability, both within a team context and compared to expectations. This type of player assessment is needed especially for defense, as similar to other sports nearly all available box score data in Ultimate Frisbee are offensive in nature.

Although sports analytics can take many forms, its use in player ratings and ranking have been among the most prevalent (Williams 2024). Not only are they easy to understand, but they are relevant to teams, players, and fans alike. These same attributes make it an attractive option for an analytical approach to Ultimate Frisbee, as it has the potential to educate fans while also helping teams and players make smarter analytically driven decisions.

# 2 Description of Data

The key to any analytics project, no matter the field, is high quality data. The UFA (formerly the AUDL) has done an excellent job of providing good access to data for teams and fans alike for recent seasons. They have easily accessible player stats dating back to the league's first season in 2012 and also detailed play-by-play data beginning in the 2021 season. This data is available both through their website (stats) and through an API based Python package (yukikongju) that can provide the framework for many statistical techniques.

One piece of data that was not as easily available was player positions. Ultimate Frisbee can have murky positional roles, but these roles are essential to understanding the complex dynamics of player performance and value. Through the UFA website, positional information was available for approximately 28% of players. This motivated our two-stage approach to developing our player ranking models. The first stage involves using machine learning to predict player position for those players missing this information. Those predictions would then be used in our candidate models for evaluating player value, rating and ranking systems.

# 3 Methodology

The goal of this project was to create player rating techniques or systems that would accomplish two separate goals. One was to capture how a player performed relative to their expectations of the context they were given. The other was to capture a player’s overall effect on their team outside of quantifiable stats. To accomplish these goals I developed a two step approach that encompasses three different techniques across the fields of statistics and machine learning.

(1) Use a classification model to determine the position of players without positional data
(2a) Build a mixed-effects model to rank players across the league in terms of specific performance metrics
(2b) Develop an on-off plus minus model similar to those employed in basketball or hockey to measure player impact on a team level

## 3.1 Positional Classifier

A crucial component of many player value models is player positions. Out of the 2101 players with available stats only 585 (27.8%) have available positional info. The UFA divided these players into four different positions: handler, cutter, defender, and hybrid. This created a multiclass classification problem to assign positions to the other 72 percent of players without a known position. Although the goal of this project was to avoid an over reliance on box score metrics, for this particular problem it was the best solution available.

Using a player's career game stats as an input, several machine learning based classifiers were implemented to identify the optimal method for assigning positions to players. After initial analysis, random forest and logistic regression were identified as models with potential for being successful in this task. For model selection, a train test split of 80/20 was used with a stratified 5 k-fold GridSearchCV was used to test both models with various hyperparameters.

The initial versions of these models revealed that the logistic regression model outperformed the random forest model with a weighted F1 score of .59 compared to .55. However, the results showed a clear trend that the classifiers did a poor job of classifying “hybrid” players. This is due to two key factors: there were less players marked as hybrid (only about 80 compared to 125-175 in other positions) and that by definition a player who is a “hybrid” often seamlessly shifts their role between different positions. For the purposes of this project, the decision was made to limit predicted positions to cutter, handler, and defender. Both models were retested as three class classifiers with the random forest model slightly outperforming the logistic regression model with a weighted F1 score of .83 compared to .81.

These results helped choose the final positional classifier to be a random forest model with the following hyper parameters. A leave one out cross validated model was used to confirm results, which proved successful with a weighted F1 score of .80 (figure 1).

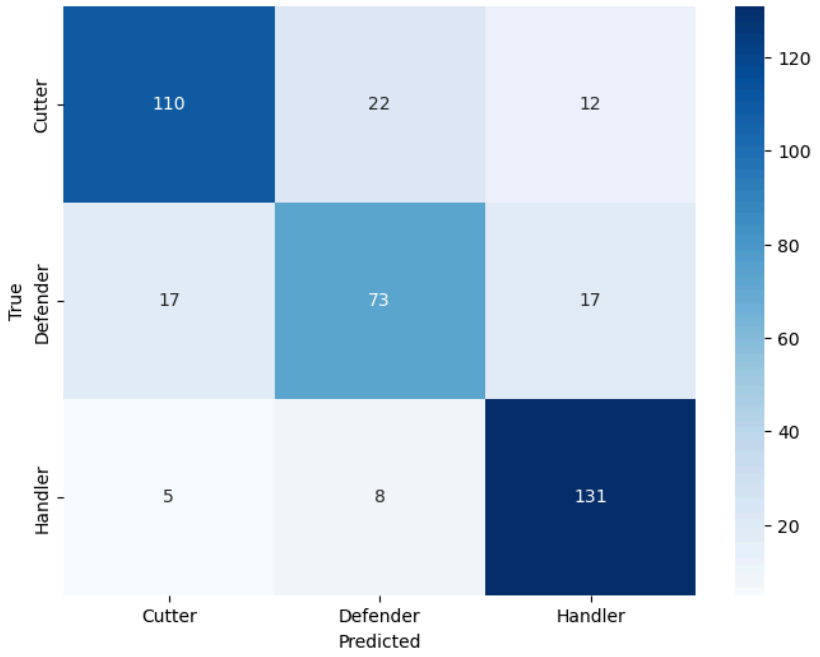
| max depth | min samples leaf | min samples split | n estimators |
|---|---|---|---|
| none | 2 | 5 | 100 |

Table 1: Hyperparamters for Random Forest Positional Classifier

Out of the 2101 total players, this model left the project with 541 cutters, 666 handlers, and 278 defenders. 29% of players were unable to be classified due to missing data, but most of those only played in the first years of the league. Using the top 8 most important features in the logistic regression model, the below graph takes a look at the defining features of different positions that the model found. Stats were averaged among positions and then normalized on a scale of 1-5 before being plotted (figure 2).

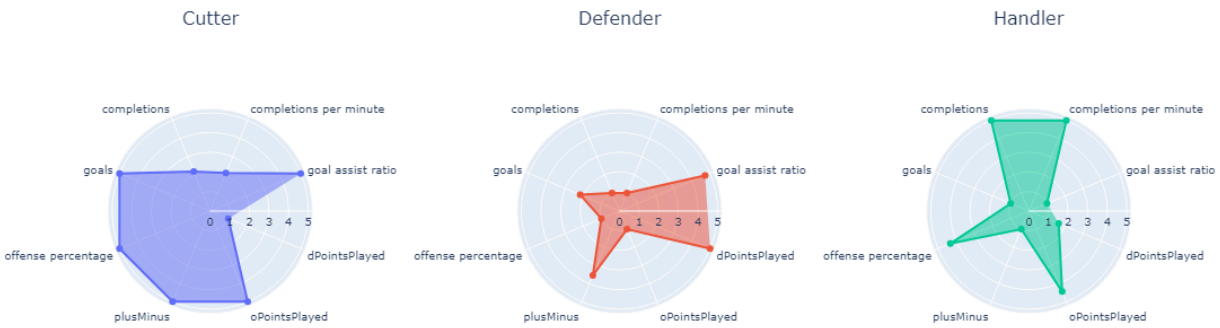
Figure 1: Leave-One-Out Cross Validation Confusion Matrix for Positional Classifier

## 3.2 Mixed Effects Model




The first player evaluation model that was developed was a linear mixed-effects model. These models are an extension of linear regression models, but are able to consider both fixed and random effects. Fixed effects are factors that are assumed to affect all entities in the same or

Figure 2: Radial Graphs of Normalized Positional characteristics




similar way. The random effects attempt to account for the "other" things that affect it. For example random effects could look at the difference between student 1 and student 2 in the same class through looking at all of their test scores. These models have been used in a large variety of contexts from medicine to marketing. In the sports analytics world they have primarily been used in baseball. For example Gerber and Craig (2021) looked at forecasting results of baseball players transitioning between playing in Japan to playing in the MLB. It has also been used to evaluate players in other sports such as in the NBA (Brahme 2022). The strength of these models in regards to player rating models is that they allow us to look at how a player performed relative to expectation by taking into account the fixed effects that could have helped or hurt a player.

The design of the mixed effects model for this project aimed to assess how players performed relative to the expectation that the fixed effects have. In order to create the model within Python, the statsmodels library was used. The fixed effects ($\beta_i$) or factors that should impact all players in the same way that were used for the model were team, position, year, offensive points played, and defensive points played. These should account for differences in team quality, positional and yearly tendencies, and the amount that a player plays. The random effects ($\gamma$) were the individual players. As it looked at each season of a player's career as a separate data point, it had several data points for each player to assess how they performed; in other words, longitudinal data. We considered four distinct target features ($y$) in order to get an overall picture of the player. The four different target variables that were used were goals, assists, blocks, and offensive efficiency (oeff). Offensive efficiency is defined as the ratio of team scores while on the field to total possessions.

Equation for mixed effects model

$$y = \beta_0 + \Sigma\beta_i x_i + \gamma$$

The output of the model contains slope values for the fixed effects and intercept values for the random effects. In this scenario the random intercepts represent how a player performed relative to expectations; a higher intercept value for a player indicates stronger performance. The intercepts from the four different models were averaged to create a composite rating for every player that accounts for both offensive (goals, assists, OEFF) and defensive (blocks) performance.

## 3.3 On-Off Plus Minus Model

This model builds off of models that are frequently used in other sports, particularly basketball. It is often alternately referred to as net plus-minus or Roland Rating (Net Plus-Minus 2017). It's a technique that was pioneered by Roland Beech, an NBA analytics guru who is currently the VP of Basketball Operation for the Dallas Mavericks. It utilizes play by play data to understand a player's impact on a game. The traditional model that is used in basketball looks how a team performs with a player on the court compared to when they are not on the court.

Net-On-Off Model Equation - Basketball

$$NetOnOff\ Rating\ =\ Team\ Net\ Rating\ While\ On\ Court\ -\ Net\ Rating\ While\ Off\ court$$

The model outlined in this paper adapts that existing framework for an Ultimate Frisbee context. One major difference in Ultimate Frisbee is that players generally stay on for points without substitutions. A point begins with one team on offense and the other on defense and ends when a team scores or the quarter ends. In the UFA the team that starts on offense scores about 68 percent of the time. This requires that adjustment is made to the model in order to properly account for the balance of offense and defense. In order to accomplish this, separate defensive and offensive models were created. A Total rating was then made by taking a weighted average of the two ratings.

On-Off Plus Minus Equations for Ultimate Frisbee

$$Oon/off\ Rating\ =\ 100\ *\ [(playerOscores\ /\ playerOpoints)\ -\ \{(teamOScores\ -\ playerOscores)/(teamOpoints\ -\ playerOpoints)\}]$$
$$Don/off\ Rating\ =\ 100\ *\ [(playerDscores\ /\ playerDpoints)\ -\ \{(teamDScores\ -\ playerDscores)/(teamDpoints\ -\ playerDpoints)\}]$$
$$Total\ on/off\ Rating\ =\ (Oon/off\ Rating\ *\ playerOpoints\ +\ Don/off\ Rating\ *\ playerDpoints)/playerTotalPoints$$

# 4 Results

This results section will discuss an overview and some specific takeaways from the models. The full results of the model have been made publicly accessible via a Streamlit web app which can be accessed at https://ultimateanalyticsapp.streamlit.app/.

## 4.1 Mixed-Effects Model

The results of these models have the ability to uncover insights on many types of players each in their own way. The mixed effect model whose target variables were primarily statistical results revealed players that outperformed their expectations in four statistical categories: goals, assists, blocks, and offensive efficiency. As it looked how players performed throughout different seasons in their career it produced career long ratings for players. The data needed for this model was available since the inaugural season of the AUDL in 2012 creating the ability to generate ratings for all players for the mixed model. However, for the purpose of this paper, the results shown for the mixed model will be based on a version of the model run only for the 2021-2024

seasons. This allows direct comparison to the on/off plus minus model which required play by play data which only became available beginning in the 2021 season.

The leaderboard of the mixed model results is full of both players often regarded as among the best of the sport and those who are rarely discussed. Similar to other leagues, the UFA has season end awards and teams recognizing the best players in the sport. By looking at the released awards and teams from 2021-2023, one can analyze how well the model aligns with traditional thinking in the sport. Seven of the top eight players in composite rating were recognized in some capacity by the UFA from 2021-2023. This includes the 2021 and 2022 MVP (Ben Jagt and Ryan Osgar). This gives somewhat of a confirmation that the model is doing a reasonable job at identifying quality players.

Table 2: Leaders for Mixed Model (min 3 years and 30 games playd)

| name | position | games | years | oeff rating | goal rating | assist rating | block rating | composite rating |
|---|---|---|---|---|---|---|---|---|
| Sean McDougall | Cutter | 51 | 4 | 0.71 | 3.13 | 1.64 | 3.04 | 2.13 |
| Ryan Osgar | Handler | 42 | 3 | 0.64 | 2.04 | 6.71 | -0.98 | 2.10 |
| Pawel Janas | Handler | 53 | 4 | 2.70 | -1.21 | 5.25 | -0.98 | 1.44 |
| Rocco Linehan | Handler | 42 | 4 | 1.34 | 1.21 | 1.09 | 2.12 | 1.44 |
| Ben Jagt | Cutter | 55 | 4 | 0.01 | 3.37 | 1.02 | 1.14 | 1.39 |
| Jordan Kerr | Handler | 52 | 4 | -0.77 | 1.33 | 3.91 | 1.02 | 1.37 |
| Alex Atkins | Handler | 30 | 3 | -0.40 | 1.78 | 0.96 | 2.69 | 1.26 |
| James Pollard | Cutter | 47 | 4 | -0.68 | 1.64 | 1.41 | 2.65 | 1.25 |
| Jace Bruner | Defender | 35 | 4 | 1.39 | 1.20 | -0.51 | 2.89 | 1.24 |
| Ben Lewis | Defender | 47 | 4 | 0.38 | 3.79 | -0.04 | 0.84 | 1.24 |

Another interesting outlook is to look at players that the model thinks of highly that haven't received notoriety for their play generally. These players may be undervalued. Three players within the top ten have not received any formal honors from the UFA. These players are Rocco Linehan, Jace Bruner, and Ben Lewis. Out of all the players on the leaderboard Rocco Linehan has by far the most balanced ratings. The difference between his lowest and highest component ratings (oeff, goal, assist, and block) is only 1.03. The second lowest difference is more than double at 2.42. This incredible balance that he has in his performance could lead to his lack of recognition. For Jace Bruner and Ben Lewis, they are both the lowest two players on the leaderboard and the only two defenders who made the top ten. This leads to a couple interesting possibilities. One, the attributes that make a great defender likely aren't captured by this model, because otherwise the highest rated defenders would likely have won defensive player of the year award or at least have made an all defensive team. Second, defenders may simply have a lack of recognition in Ultimate similar to what often happens in other sports.

## 4.2 On-Off Model

On the contrary to the mixed effects model the on/off plus-minus model is an impact metric which looks at a players impact on their team success. A rating can be read at how

percentage increases or decreases that a player gives a team when on the plate. For example Ryan Osgar had an on off rating of 25 in 2022 meaning that his team scored 25% higher rate when he was on the field versus when he wasn't.

Similarly to the mixed model, the majority of the top ten players have been recognized since 2021 by the UFA in their year-end awards. However, for the most part there was a different cross-section of the best players in the sport. The only two players that appeared in the top ten in both models were Pawel Janas and Ryan Osgar. This isn't surprising as they have consistently been recognized as some of the league's best players over the last several years as two of only three players that were on the league's first team every year from 2021-2023. There are other patterns that appear that may reveal some of the model's strengths and weaknesses.

Table 3: Leader from On-Off Plus-Minus Model (min 3 years and 30 games played)

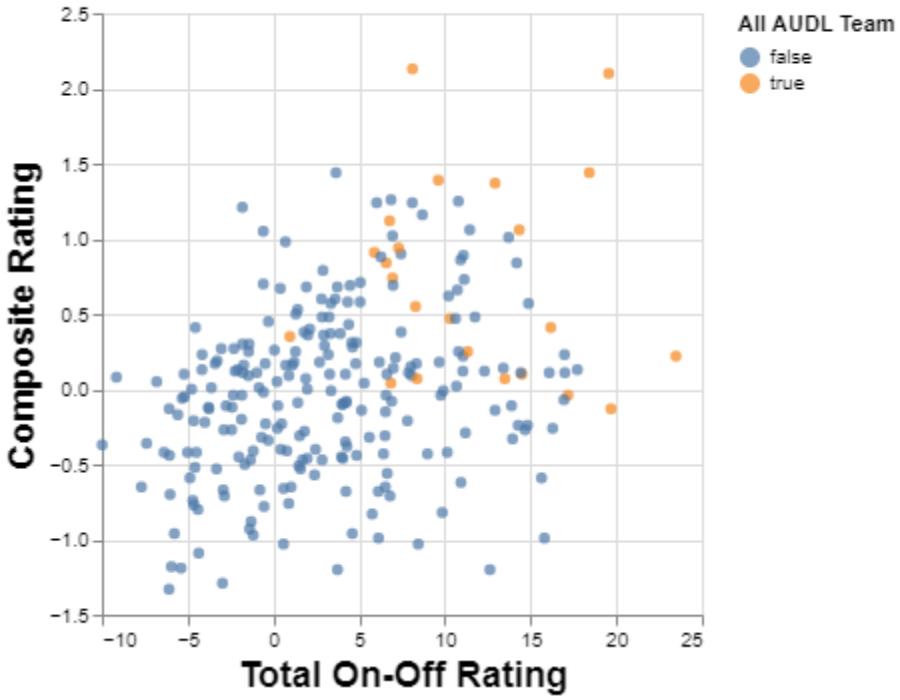
| name | position | games | years | offensive on off rating | defensive on off rating | total on off rating |
|---|---|---|---|---|---|---|
| Jack Williams | Handler | 55 | 4 | 25.82 | -0.06 | 23.48 |
| Austin Taylor | Cutter | 50 | 4 | 20.13 | 8.18 | 19.69 |
| Ryan Osgar | Handler | 42 | 3 | 19.73 | -1.62 | 19.56 |
| Pawel Janas | Handler | 53 | 4 | 18.55 | 1.85 | 18.43 |
| Bobby Ley | Handler | 43 | 4 | 17.87 | 4.92 | 17.73 |
| Jacob Miller | Handler | 52 | 4 | 19.72 | 7.96 | 17.18 |
| Elliott Chartock | Handler | 49 | 4 | 17.31 | 12.82 | 17.00 |
| Sean Connole | Handler | 40 | 3 | 17.52 | 8.18 | 16.98 |
| Luc Comire | Handler | 40 | 4 | 17.89 | 6.93 | 16.95 |
| John Lithio | Cutter | 52 | 4 | 17.61 | -11.74 | 16.29 |

Likely the most noteworthy trend in the results are the positional trends. All but two of the top ten players in on-off rating were handlers. Not only did no defenders make the cut for the top ten, but the highest rated defender ranked 62nd overall. There are likely several reasons for this model's preference for handlers. This bias was made possible by it being a position-blind model. The mixed model took into account a players position so players ratings were adjusted to be compared against players in their same position. The biased results of a position-blind model likely signal some other biases either in modeling for the project or within the sport itself. As most players' positions were determined by the stats-based positional classifier, that could be part of the cause of the bias. As stats are more offensively focused, players with low stats were generally given a label of defender, which could create a bias of players that are generally worse being labeled as defenders. Another hypothesis is that handlers are often the most skilled players on the field in Ultimate, so the value of them being on the field could just generally be higher than nearly all defenders. More research would need to be done to discover the reason for the bias.

The below graph looks at the relationship between the two developed models. It plots the ratings of the 250 players who played at least 3 seasons and 30 games from 2021-2024. It shows that there is a slight positive relationship between the two models, although it is also clear that the

two models have major disagreements on many players. Every player who made an all-ufa team from 2021-2023 had a positive on-off rating, meaning their teams were better while they played than when they didn't. Only two of these players had a negative composite rating over the same span meaning that they underperformed how the mixed model would expect them to perform.

Figure 3: Total On-Off Rating vs. Composite Rating (Mixed Model)




# 5 Future Work

Although the results of this project reveal much about player performance and evaluation in Ultimate Frisbee, it also raises several new questions that could be answered via future research. One particular area of future interest would be positional classifications in Ultimate. Researchers could look into either better ways to determine the positions of players or look at clustering as a technique to give players roles/positions that may not be considered any of the traditional positions. Another option could be to treat Ultimate as positionless sport and continue to look at value irrespective of position.

One of the primary drawbacks of on-off ratings is they fail to capture the interactions between players that are on the field. For example if I (a not very good player) were to only play with amazing players, my on-off would likely be very high. On the other hand, if the best player on a team often plays with lineups made up of bench players in an effort to lift them up, their on-off rating will be negatively impacted. There are many scenarios that can cause on-off ratings to become misleading. The primary way to overcome this is to also create x-player combination on-off ratings. These ratings look at the effect a combination of players have on the game while

they are playing. This can be done in groups of players from two all the way to a full lineup of seven players. This type of analysis gets rid of many of the biases that a single player on-off rating contains.

The potential of these other endeavors could create significant change within the world of Ultimate. They could help coaches make lineup decisions, general managers make contract decisions, and aid fans in their understanding and enjoyment of the game.

# 6 Conclusion

In this paper, we have shown that it is possible to use existing analytics frameworks to create player rating systems for Ultimate Frisbee. By adjusting frameworks originally developed for baseball and basketball, we created evaluation metrics that passed the "eye test" by identifying many of the highly awarded athletes in the sport. However, the metrics differed in key areas both from each other and traditional wisdom, thus creating new perspectives to analyze players. These insights have the potential to help coaches and general managers in their day-to-day. These models both show that Ultimate has the potential to benefit from analytics and that analytics developed for team sports can often be adapted for unintended uses.

# References


*About Ultimate*. (n.d.). Retrieved September 3, 2024, from https://archive.usaultimate.org/about/

Brahme, A. (2022, July 21). *Understanding Scoring Propensity: A Mixed Model Approach to Evaluating NBA Players*. Medium

Gerber, Eric & Craig, Bruce. (2021). A mixed effects multinomial logistic-normal model for forecasting baseball performance. Journal of Quantitative Analysis in Sports. -1. 10.1515/jqas-2020-0007.

Goldberg, M. (n.d.). *Evaluating Lineups and Complementary Play Styles in the NBA*. Retrieved September 5, 2024, from https://dash.harvard.edu/handle/1/38811515

*History of Ultimate*. (n.d.). WFDF. Retrieved September 3, 2024, from https://wfdf.sport/history/history-of-ultimate/

LEBRON Introduction. (n.d.). *Basketball Index*. Retrieved September 3, 2024, from https://www.bball-index.com/lebron-introduction/

. *Lewis, M. (*2004). Moneyball. WW Norton.

*Net Plus-Minus (Roland Rating) Explained*. (2017, May 8). https://www.nbastuffer.com/analytics101/net-plus-minus/

*Williams, Benjamin & Schliep, Erin & Fosdick, Bailey & Elmore, Ryan. (2024). Expected Points Above Average: A Novel NBA Player Metric Based on Bayesian Hierarchical Modeling.*

*Stats | WatchUFA*. (n.d.). Retrieved September 3, 2024, from https://watchufa.com/stats/player-stats?year=2012

UFA Ultimate Frisbee Association. (2016, September 2). *Top SportsCenter Moments* [Video recording]. https://www.youtube.com/watch?v=y2a4wZc2MjA

Wurtztack, P. (n.d.). *Better Box Score Metrics: The AUDL's Offseason Data Upgrade—Ultiworld*. Retrieved September 3, 2024, from https://ultiworld.com/2022/04/06/better-box-score-metrics-the-audls-offseason-data-upgrade/

yukikongju. (n.d.). *Audl · PyPI*. Retrieved September 3, 2024, from https://pypi.org/project/audl/