# High-resolution shot capture reveals systematic biases and an improved method for shooter evaluation

Rachel Marty
University of California, San Diego
ramarty@ucsd.edu

Evaluating shooting ability is a critical component of player comparison and player development. However, players are often evaluated on a limited number of shots, exposing assessment to high variation and inaccurate, anecdotal conclusions. The aim of this paper is to explore the potential of high-resolution shot data to improve shooter evaluation. Using over 22 million shots captured in high-resolution by Noahlytics, we reveal previously hidden systematic biases in entry left-right and entry depth from all positions on the court. Then, we focus on the high-resolution shot data from 509 NBA, college and high school players to train a machine-learning algorithm that predicts shooting ability from 25-shot sessions. The algorithm outperforms conventional methods and better ranks players by skill-level. We conclude by encouraging coaches and players to re-evaluate their largely anecdotal assessment methods and implement more effective, data-driven methods to enhance shooter development and shooter ranking.

## 1. Introduction

### 1.1. Motivation

Six years ago, Kirk Goldsberry revolutionized shot analysis with new shot chart visualizations that introduced a spatial component to shooting percentage[1]. He advocated for analysis to go beyond summary percentages and to consider shooter ability from different positions on the court. His research has changed the way the NBA evaluates shooters. Now, with newly-developed technology, we can build on these findings to get greater insight into factors that influence shooting percentage. Previously we could only expose low percentage shooting from different areas of the court, while now we can explore why shooters miss from those positions. Moreover, rather than assessing shooting ability on shooting percentage alone, where low sample sizes can introduce bias, we now can capture and analyze valuable information about how players achieve their shooting percentages.

Knowing this "why" and "how" has significant ramifications in the NBA for shooter development and shooter ranking. First, high shooting percentages make the game more exciting for fans, and understanding why players miss from specific areas on the court reveals actionable changes for shooter development. Second, ranking players based on their shooting ability is a critical component in drafting and trading NBA players. Currently, players are assessed by their shooting percentage on a minimal number of shots[2]. This approach is prone to high sampling error and inaccurate results, leading to suboptimal team rosters.

In this paper, we explore the value of high-resolution shot data to augment shooter development and shooter ranking. This paper is built upon 22 million shots captured in high-resolution by Noahlytics. High-resolution shot data allows the exploration of not only shot position on the court, but also how the ball approaches and interacts with the hoop. We start by redesigning shot charts through the lens of spatial shot patterns at the rim to expose systematic miss patterns in the population. Through unsupervised learning, we explore the potential of high-resolution rim pattern data capture to augment shooter evaluation. Then, we train a spatial rim pattern-based supervised algorithm to rank shooter ability that outperforms conventional assessment approaches.

We foresee that this research will have long-term impact in the NBA by materially increasing shooting percentages and reforming shooter ranking methods. As more players adopt high-resolution data capture training methods, we predict that systematic shooting biases will be resolved. As more coaches adopt high-resolution shooting analysis approaches, players will be evaluated for draft or trade based upon quantitative assessment of shooting skill rather than inaccurate, small sample size evaluation.

## 1.2. Data collection

This study uses over 22 million shots captured by the Noahlytics system. The Noahlytics system utilizes a sensor that hangs about 13 feet above the basket, capturing all shots taken. The system accurately and automatically determines the shot location, whether the shot was made or missed, and how the shot was made or missed using high-resolution shot attributes. The data in this paper comes from 5,649 individuals. Shooters are both male and female from all levels (NBA, WNBA, NCAA, high-school, etc.).

## 1.3. High-resolution shot attributes

For each shot taken, the system collects data about how and where the basketball shot enters the plane of the hoop. We analyze three attributes of shot entry: Left-Right, Depth and Angle. Since shots are taken from all positions of the court, these shot entry attributes are measured from the perspective of the shooter; the point on the hoop closest to the shooter is always defined as the front of the hoop. These attributes have previously been described in great detail[3].

In brief, Left-Right is the left to right deviation of the shot at the hoop. A shot which lands exactly on the leftmost part of the hoop from the perspective of the shooter has a Left-Right value of -9”, a straight shot has a Left-Right value of 0” and a shot which lands on the rightmost part of the hoop has a Left-Right value of +9”. Depth is the entry depth of the shot into the hoop. A shot which lands directly on the front of the hoop from the perspective of the shooter has a Depth value of 0” and a shot which lands directly on the back of the hoop has a Depth value of 18”. Angle is the entry angle of the shot into the hoop. A relatively flat shot can have an Angle value of 36° and a relatively high arcing shot can have an Angle value of 55°.

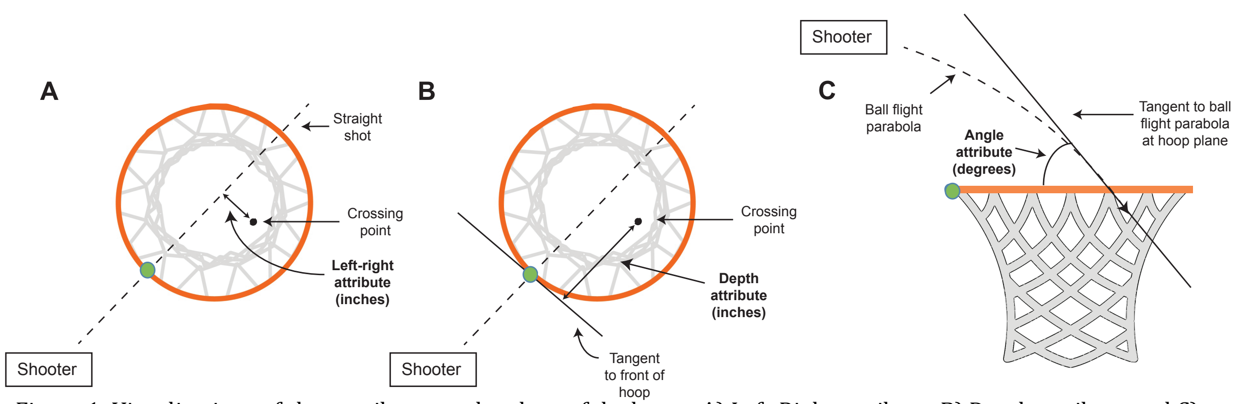



Figure 1: Visualizations of shot attributes at the plane of the hoop - A) Left-Right attribute, B) Depth attribute and C) Angle attribute.

# 2. Systematic Biases

In this section, we utilized the high-resolution shot capture data for over 22 million shots to construct court floor maps which reveal systematic population shooting biases [4,5]. Each square in the heat maps represents a six-inch by six-inch square on the basketball court. The squares are colored according to the mean value for all shots taken from that position of the specified shot attribute, first for the Left-Right attribute and then for the Depth attribute.

## 2.1 Reduced shooting percentage due to Left-Right biases in corner 3-point shots

When taken as a group, all shots have an average Left-Right value of 0 inches (straight), and excellent shooters have an average Left-Right value of 0 inches with minimal Left-Right variation from shot to shot. Despite these summary statistics, we hypothesized that there may be systematic Left-Right biases from specific positions on the court that are invisible to the naked eye. To assess this hypothesis, we divided the 22 million high-resolution shots into six-inch by six-inch squares on the court based on their shot location. For each of these squares, we calculated the mean Left-Right value across the entire population (**Figure 2A**). As expected, court positions that are dominated by bank shots vary significantly from 0 inches (straight) as shown by the dark red and blue areas emanating at 45 degrees from the left and right sides of the hoop. Ignoring these two bank shot dominated areas, the right side of the court has a left bias (red color) and the left side of the court has a right bias (blue color). Thus, on the right side of the court players shoot to the left of center, and on the left side of the court, players shoot to the right of center.

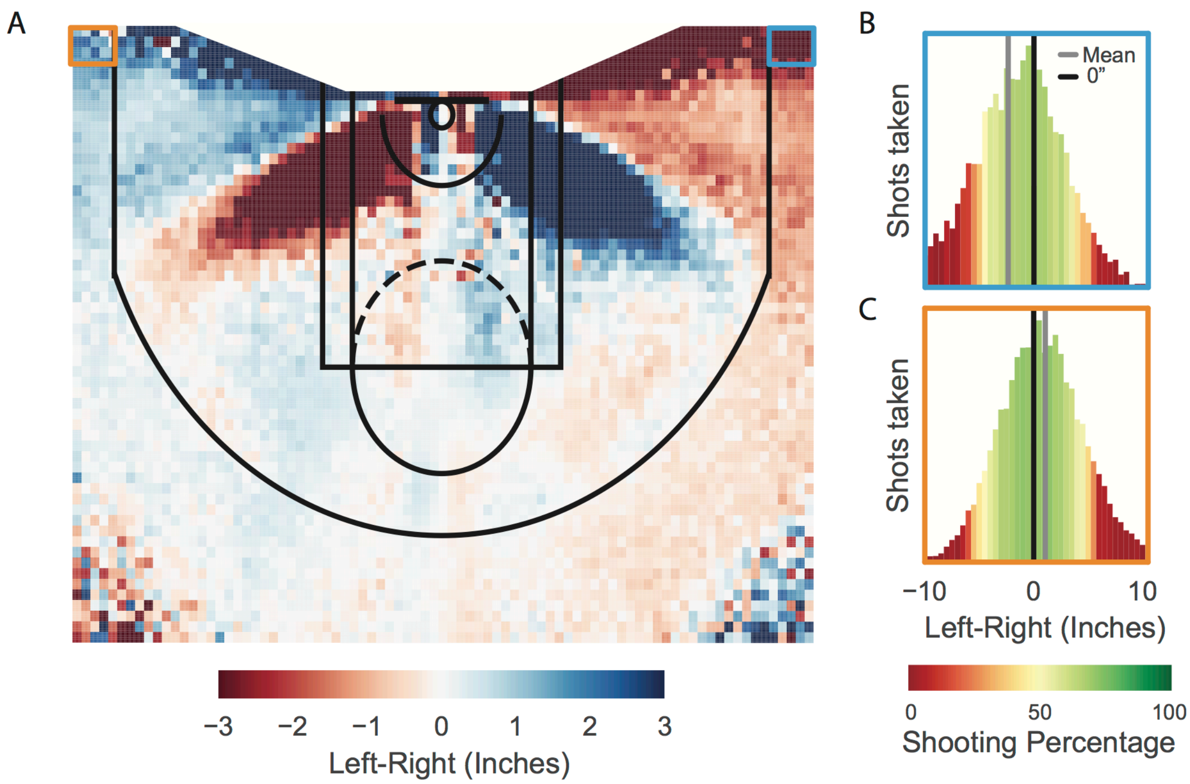



Figure 2: Systematic Left-Right Biases. (A) A heat map of the mean Left-Right values across all 22 million shots for each six-inch square of the court. (B-C) Distribution of the number of shots taken across different Left-Right values colored by percentage made for (B) right corner (blue box) and (C) left corner (orange box). Center of the hoop is denoted with a black line and the Left-Right mean is denoted with a grey line.

This systematic bias is particularly prominent for 3-point shots, with the largest bias occurring for baseline 3-point shots. The baseline 3-point shot is strategically important in the NBA because it is the shortest shot resulting in 3 points of value. On average, players shooting from the right corner shoot 2.34 inches left of the hoop center (**Figure 2B**). On average, players shooting from the left corner shoot 1.05 inches right of the hoop center (**Figure 2C**). An average NBA player shoots with Left-Right standard deviation of about 4 inches from the corner 3-point distance. Using this standard deviation, we simulated Left-Right distributions with different Left-Right averages. Thus, we could estimate shooting percentage loss at each suboptimal Left-Right value for an average NBA player. Given these distributions, an NBA player with average Left-Right distribution is sacrificing up to 4% of shooting percentage from the right corner and up to 2% of shooting percentage from the left corner. Notably, the bias in the right corner is much more extreme. While the Noahlytics system does not currently collect data on the handedness of the shooters, we hypothesize this difference is due to a much higher percentage of right-handed shooters.

## 2.2 Reduced shooting percentage due to longer shots landing shorter in the hoop

Last year, we reported that 3-point shots are most likely to score when shot at the center of the Guaranteed Make Zone (GMZ), the region of the hoop where a shot is guaranteed to score[3]. We also found a systematic population bias for 3-pointers to be shot 2 inches short of this center point. In last year's study, we grouped all 3-point shots together and did not consider how shot distance

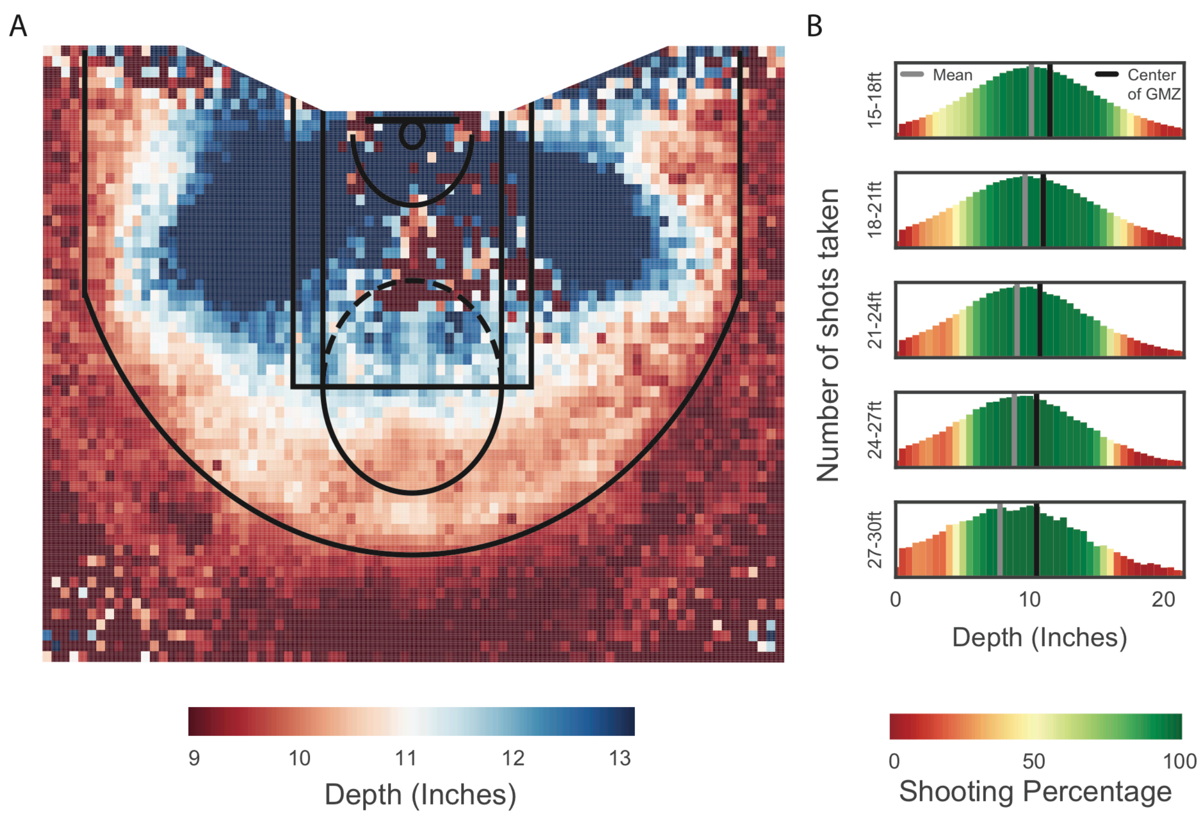



Figure 3: Systematic Depth Biases. (A) A heat map of the mean Depth values across all 22 million shots for each six-inch square of the court. (B) Distributions of the number of shots taken across different Depth values colored by percentage made. Charts visualize the shots taken 15-18 feet from the basket on the top and 27-30 feet from the basket on the bottom. The mean Depth is denoted with a grey line and the center of the GMZ is denoted with a black line.

impacts shot Depth. This year we grouped the 22 million high-resolution shots into six-inch by six-inch squares on the court (**Figure 3A**). Again, bank shots stand out in this plot. They are the blue regions emanating at 45 degrees from the left and right sides of the hoop. Bank shots are shot on a trajectory that would proceed past the rim if the backboard were removed, so bank shots appear to be on a trajectory that goes well past the hoop. For this analysis, we ignore the bank shot regions and focus on non-bank shots. Non-bank shots demonstrate a clear trend; shorter shots go deeper in the hoop and longer shots go shorter in the hoop. This can be seen most clearly in the bottom half of the figure where bank shots are rare. Mid-range shots average about 11 inch Depth (white) while long 2-point shots average about 10 inch Depth (light red) and 3-point shots average about 9 inch Depth (dark red). This trend is consistent in every direction and holds for median-based statistics.

To assess the relationship between shot distance, Depth and make percentage, we plotted the distribution of shots taken at several different shot distance ranges, coloring each bar with the make percentage in the population (**Figure 3B**). To enable cleaner analysis, we only considered non-bank shots that went straight (-2 < Left-Right > 2). The dark green color in each chart represents the region of the hoop with the highest percentage of shots scored, also known as the GMZ. Notably, the mean Depth value decreases as shots get longer, from 10.1 inches at 15-18 feet to 7.8 inches at 27-30 feet (denoted with the grey line). On the other hand, the center of the GMZ region remains more constant, only shifting from 11.5 inches at 15-18 feet to 10.5 inches at 27-30 feet (denoted with a black line). Thus, players consistently shoot shorter, on average, than the center of the GMZ. We used an NBA player with standard Depth variation to estimate the percentage loss due to these biases. Depth distributions are consistently skewed short (non-normal), so we estimated the percentage loss by varying the median Depth of the empirical Depth distribution of the player. Simulations showed about a 1.5% decrease in shooting percentage for the typical NBA 3-point shot of 24-27 feet as compared to the optimal Depth. Since NBA teams are increasingly shooting the 27-30 foot 3-point shot to open up the lane interior, it is worth noting that the long 3-point shot has an even more dramatic 2.9% decrease in shooting percentage. These shooting percentage decreases can be recovered by simply training to shoot at the center of the GMZ.

These case studies of shot attributes across the population demonstrate the value of high-resolution data capture to gaining important insights that were previously unnoticed. They also demonstrate the importance of ensuring players are shooting shots in the ideal ranges of these shot attributes in order to maximize shooting percentage.

# 3. Ranking players by shooting ability

Ranking shooters by skill-level is essential to comparing players and assessing player improvement. In this section, we focus on shots of length 18-22 feet in order to expound the limitations of conventional shooting ability assessment. We recommend a superior method based on the high-resolution shot data described in Section 2.

## 3.1 Limitations of shooting percentage

Typically, a quantitative assessment of a player's shooting ability is based on a shooting percentage from a limited number of shots. Unfortunately, low sample sizes can lead to wide distributions of observed shooting percentages across multiple sessions. When shooting ability is derived from a single shooting session, it is often an inaccurate representation of the shooter's true ability.

In order to visualize and quantify the variation of shooting percentages at different sample sizes, we looked at three actual players – referred to here as Player A, Player B and Player C. We tracked these players over 10 months. Players A, B and C took 27,636, 25,031 and 10,524 18-22 foot shots, respectively, over this time period. The players averaged 50%, 58% and 75% made shots, respectively. Given the large number of shots, we can confidently discern that Player C is the best shooter, followed by Player B and eventually by Player A. However, some days Player C might produce a lower shooting percentage than Player A by chance. To visualize the variation and likelihoods of possible shooting percentages for each player, we broke all of each player's shots into shooting sessions of different sizes based on time stamps from the shots. Due to a strong correlation between player movement and shooting percentage, we only considered sessions with low player movement for all analyses in Section 3. (See appendix for details.) On a day when a player took 600 shots, we would extract 24 25-shot sessions, 6 100-shot sessions and 1 500-shot session. The likelihood of a player to shoot a particular shooting percentage is denoted by the height of the distribution (**Figure 4A**). When only 25-shot sessions are considered, the overlap between shooters is very high, meaning it is very challenging to confidently rank players based on shooting percentage from 25-shot sessions. As sessions of higher shot counts are considered, the distributions start to separate, but there is still high overlap. Sessions of over 1,000 shots are necessary to rank shooters reliably.

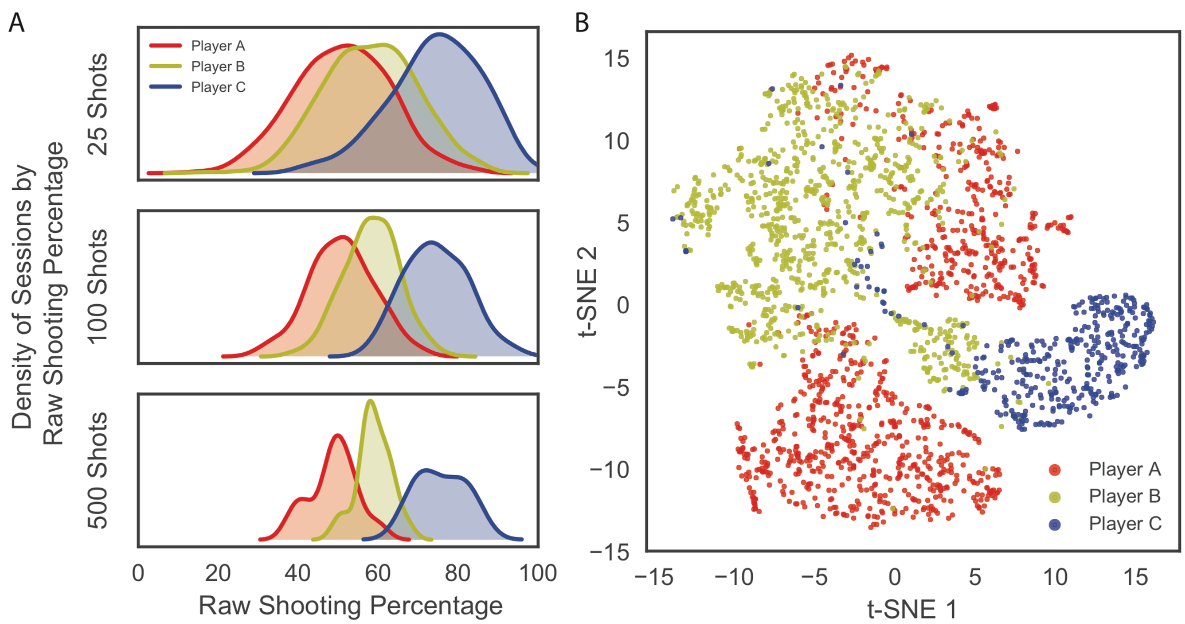



Figure 4: Limitations of shooter ability assessment by raw shooting percentage. (A) Distributions of raw shooting percentage densities of three real players. Sessions of different sizes are visualized: 25, 100, 500 from top to bottom. (B) Nine shooter dimensions from 25-shot sessions visualized using t-SNE dimension reduction technique. Sessions are colored according to player.

In addition to measuring raw shooting percentage for a session, high-resolution spatial rim patterns are also captured for each session. This data allowed us to identify systematic shot characteristics based on location, as shown in Section 2. Here, we employ those same tools to understand player ranking based on a small sample size. The **spatial rim pattern** is defined as the following nine features: the mean for each shot attribute (Left-Right, Depth and Angle), the standard deviation for each shot attribute and spearman correlation between each pair of shot attributes. These nine features are calculated for each session. While raw shooting percentage for a session only gives one dimension of measurement to evaluate a shooter's skill, these rim pattern statistics give an additional nine dimensions of measurement for each session. T-distributed stochastic neighbor embedding (t-SNE) is an unsupervised machine learning technique used to visualize high dimensional data in two dimensions [6,7]. The algorithm calculates a probability distribution over pairs of data with the goal of visualizing two points that are close in high dimensional space as close in two-dimensional space. We applied this technique to the 25-shot sessions of the three players. Notably, this unsupervised approach clusters the sessions by players without having any prior knowledge about the player (**Figure 4B**). This analysis suggests that the nine additional high-resolution spatial rim pattern dimensions contain valuable information for player classification that may be predictive of shooting skill as well.

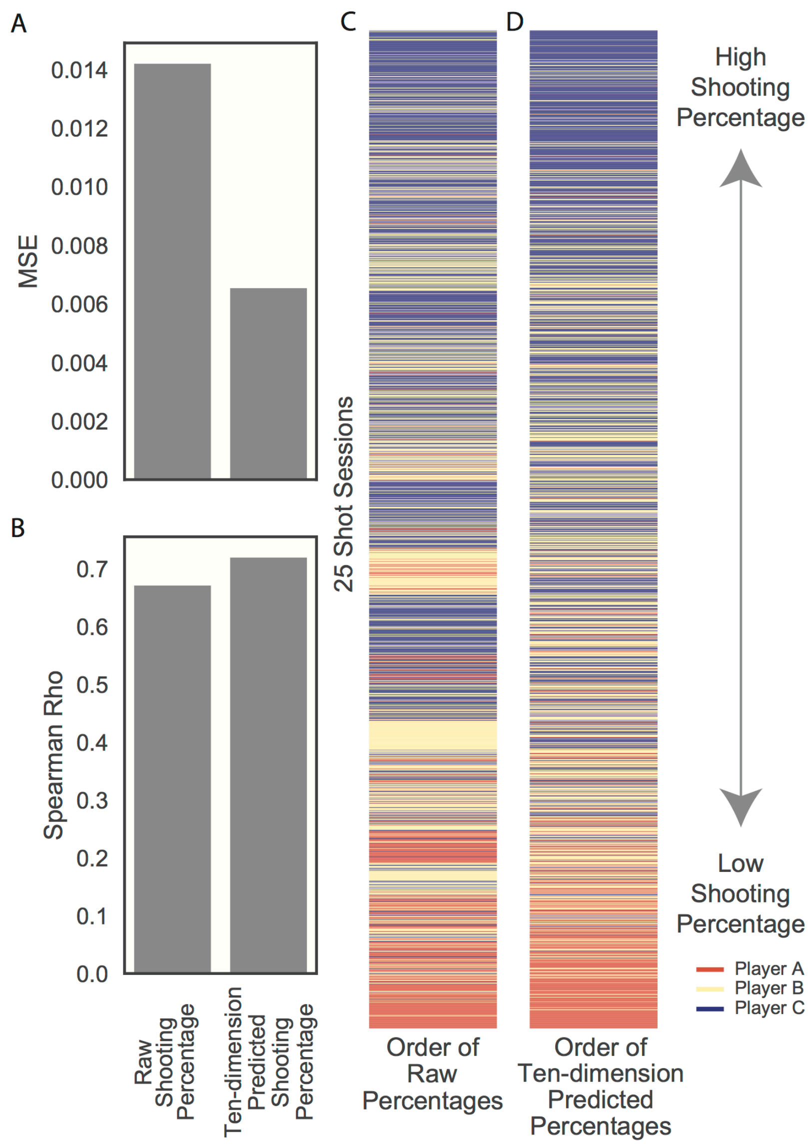



Figure 5: Results of prediction algorithm. (A) Mean Squared Error of the test set for the two models. (B) Spearman rho between the test set for the two models. (C-D) 25-shot sessions colored by player and ordered by (C) raw shooting percentage and (D) ten-dimension predicted shooter percentage.

### 3.2 Prediction model

We created a supervised model to use raw shooting percentage augmented with the nine high-resolution spatial rim pattern features. We hypothesized that we could train this ten-dimension model to assess shooter ability more accurately over a small number of shots than by using the single dimension of raw shooting percentage alone. We chose to use Gradient Boosting Regression to allow for the optimization of arbitrary differentiable loss functions. We built a data set by extracting

25-shot sessions from players with more than 1,000 stationary shots in the 18-22 foot range. The labels were the overall shooting percentage for all of a player's stationary shots in the distance range. The data was split into a training set (with 2/3 of the players) and test set (with 1/3 of the players). It was also down-sampled to achieve an even distribution across all observed shooting percentages. All results are reflective of the performance of the model on the test set.

### 3.3 Ten-dimension prediction model improves player ranking

The quality of the high-resolution model, which includes raw shooting percentage data augmented with nine additional spatial rim pattern dimensions, was tested in comparison to raw shooting percentage data alone. Both metrics were compared to the gold standard of shooting percentage evaluation – the cumulative percentage from each of the player's thousands of shots. First, we calculated a mean squared error. The model has half the mean squared error of raw shooting percentage, suggesting very strong performance (**Figure 5A**). We also performed a Spearman rank test on the results, revealing an increased correlation between overall shooting ability and the ten-dimension prediction model shooting percentage as opposed to raw shooting percentage alone (**Figure 5B**). There is currently a bias in the estimator due to the limited range of training percentages that inflates the predicted shooting percentages of poor players and deflates the predicted shooting percentages of great players. (See appendix for details.) This bias inflates the loss in mean squared error, but it does not impact the Spearman rank test. Even with this bias, shooters can still be ranked on ability from small shot sessions better than ever before using the ten-dimension prediction model. There are methods to correct this bias that will allow even better separation of player shooting ability from small shot sessions, but these methods will not be developed further in this paper. The difference between ranking players based on raw shooting percentage from small shot sessions compared to the ten-dimension model is visualized for Players A, B and C (**Figure 5C-D**). In the ten-dimension model, Player A is more frequently predicted to be a poor shooter and Player C is more frequently predicted to be a better shooter despite the variability of their performance on any given day.

## 4. Discussion

Conventional shot charts inform a player about their shooting percentage but lack any information about why the player attains that specific shooting percentage. We describe spatial rim pattern shot charts to give players information that is both informative and actionable. This information will provide players with the resources they need to develop and improve as shooters. Although we looked at trends across the entire population, each individual player has their own personal spatial biases. In the future, we foresee the development of spatial rim pattern court maps for individual players. These court maps will allow coaches to describe weaknesses to players and give them actionable advice for shooting percentage improvement. Furthermore, comparison with a database of millions of shots will give players improvement incentives by accurately measuring their potential shooting percentage increase with each spatial improvement. These spatial rim pattern shot charts will change the way coaches and players approach shooter development.

Conventional shooter evaluation approaches rely on raw shooting percentages derived from a minimal number of shots, resulting in inaccurate rankings. We demonstrated that the approach using high-resolution shot data to construct a ten-dimension prediction model better ranks shooters according to their ability. As data is collected over several years, we foresee the model learning to predict potential future shooter ability in addition to current shooter ability. This capability will have huge impacts on player drafting, player trades and player development in the NBA. The ability to capture high-resolution spatial shot rim patterns has opened these ideas to reality. The only barriers standing in the way of additional progress are the accumulation of more high-resolution rim pattern data and the application of machine learning techniques.

## 5. Conclusion

We argue that current approaches to shooter assessment are inadequate because they are prone to low sample size variation and neglect shooter spatial rim patterns. In this paper, we explored 22 million shots captured with high-resolution Noahlytics technology. First, we established the power of high-resolution shot capture technology to expose systematic shooting biases that impact shooting percentage. Second, we developed and recommended a shooter evaluation approach that integrates spatial rim patterns with raw shooting percentage into a ten-dimension prediction model. In the end, we concluded that high-resolution shot capture has the potential to improve shooter development and shooter ranking in the NBA.

## 6. Acknowledgements

John Carter, CEO of Noah Basketball, and Logan Buchanan, Developer at Forty AU, for providing the Noahlytics high-resolution shot capture data used in the analysis.

## 7. Resources

Code for this project is hosted at https://github.com/Rachelmarty20/Noah.

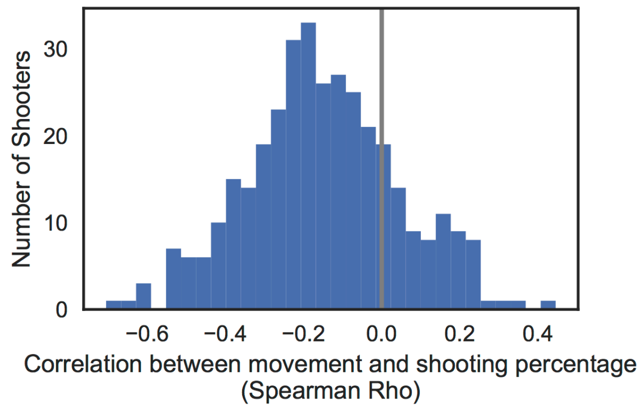
## References

[1] Goldsberry, Kirk. "Courtvision: New visual and spatial analytics for the nba." *2012 MIT Sloan Sports Analytics Conference*. 2012.
[2] NBA Combine. https://stats.nba.com/draft/combine-spot-up/
[3] Marty, Rachel, and Simon Lucey. "A data-driven method for understanding and increasing 3-point shooting percentage." *2017 MIT Sloan Sports Analytics Conference*. 2017.
[4] Tjortjoglou, Savvas. How to Create NBA Shot Charts in Python. http://savvastjortjoglou.com/nba-shot-sharts.html
[5] Seaborn: statistical data visualization. https://seaborn.pydata.org/
[6] Maaten, Laurens van der, and Geoffrey Hinton. "Visualizing data using t-SNE." *Journal of Machine Learning Research*9.Nov (2008): 2579-2605.
[7] Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.

# Appendix

## Section 1

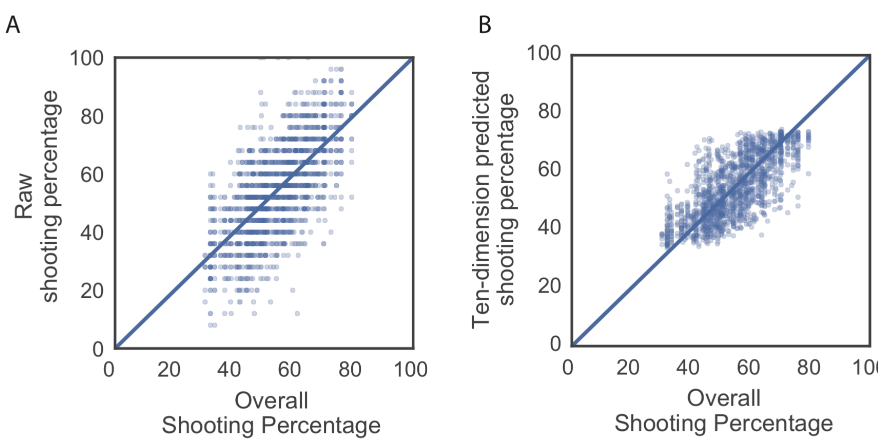
As mentioned in Section 3.2, there is a striking correlation between shooter movement during a session and shooting percentage. In order to assess the relationship between shooter movement and shooting percentage, we looked across all players with over 1,000 shots in the 18-22 foot range. For each day a player shot more than 25 shots, we split their shots into 25-shot sessions by the shot time stamps. For example, if 25 shots were taken, only one session was extracted. If 50 shots were taken, two sessions were extracted, and so on. Then, we defined movement as the average (mean) distance between consecutive shots in each session and calculated this value for each player. For each player, we calculated the Spearman correlation between movement and shooting percentage across all of their sessions (**Appendix Figure 1**). The vast majority of players have a significant negative correlation between these two variables. Thus, we excluded all sessions with an average movement of greater than two feet in order to isolate a single shot type and increase the prediction power.




Appendix Figure 1: A histogram of Spearman rho correlations for player movement between shots and shooting percentage in a session for all players.

## Section 2

As mentioned in Section 3.3, the prediction range of the ten-dimension prediction model is limited by the input data. Due to the level of players tested, the training data only has players with overall shooting percentages between 30% and 90%; thus, the ten-dimension prediction model will only predict between those values. As a result, the poorer shooters will have predicted shooting percentages that are skewed upward and the better shooters will have predicted shooting percentages that are skewed downward (**Appendix Figure 2**).




Appendix Figure 2: Prediction algorithm results. (A) Scatter plot of raw shooting percentage with overall shooting percentage. (B) Scatter plot of ten-dimension predicted shooting percentage with overall shooting ability.