Evaluating the Probability an NBA Player Increases the Chance of an Offensive Rebound by Missed Shot Distance

ID: 82454

1 Introduction

Identifying factors that influence a team's chances of winning a game typically begins by observing the flow and subsequent interactions that define the game. In general, a given game can be described as an ordered process of events or actions with varying resultant reactions that yield specific outcomes typically manifested in points scored. Given an objective to achieve, such as scoring more points than the opponent, and a detailed examination of the process within a game, players' actions can be marked as either value add or non-value add towards their team's chances of winning. This evaluation process is known as lean manufacturing, first introduced by Toyota. The theory requires a full understanding of the entire production process to produce one unit, the identification of all actions that are value and non-value add to then the redevelop the process aimed to reduce non-value and strengthen value add actions.

The same methodology and principles can be applied to sports by replacing the unit of production with a unit of points scored and, in this paper, is done for the game of basketball. In total, a game of basketball is a process of alternating offensive possessions between each team. The main objective is to maximize points, but to understand how this is accomplished the process must be examined at the possession level to identify value and non-value add actions. In a simplified basketball possession, an offensive team attempts to score by shooting the ball via a variety of ways. More often than not, the shot will be missed and the ability for a team to control the possession of the ball is dependent on their ability to rebound. The defensive team aims to rebound the missed shot in order to prevent the offense another scoring opportunity while the offense attempts to rebound the missed shot to create another scoring opportunity.

A team's ability to rebound the ball effectively directly effects their ability to control possession and indirectly, among many other factors not considered in this paper, their ability to increase their own and decrease their opponent's point total. As a result, offensive and defensive rebounding are value-add actions in the game of basketball. To analysis a player's impact and corresponding rebounding ability, this paper develops a binomial Rasch model to analysis and rank players' impact towards the occurrence of an offensive rebound while they are on the court during the NBA for the 2015 – 2016 season. To better understand a player's rebounding impact, the model takes into consideration the offensive and defensive players on the court, the distance of the missed shot the rebound has become available on and the type of rebound collected [offensive or defensive]. The paper's objective is to evaluate how a player's actions on the court impacts the probability of an offensive rebound off of a given missed shot attempt while on offense or defense.

2 Data

Data used for paper was the 2015 – 2016 NBA season play-by-play data acquired from NBAstuffers.com up to May 1st, 2016. Given the size of the seasonal play-by-play data, the file was synthesized in R by retaining only information pertinent to rebound events. A typical row of the data file used showed the following content:

Offensive Player 1 | O_2 | O_3 | O_4 | O_5 | Defensive Player 1 | D_2 | D_3 | D_4 | D_5 | Player Responsible | Rebound Type | Shot Distance |

The data was further analyzed in R using a *sparse matrix* object and the *glmnet* package to create a Rasch model.

3 Methodology

To evaluate NBA players' impact on the probability of an offensive rebound from the 2015 – 2016 season, the paper develops a binomial Rasch model, a statistical measurement model better suited for individual player evaluations. In general, a binomial Rasch model returns the probability of a specific outcome as a function of a person and specific item parameters. In this model, the probability is measured as a player obtaining an offensive rebound with specific item parameters of other players [offensive and defensive] on the court and the distance of the missed shot. In this paper the following Rasch model [see equation 1] was used:

Equation (1)
$$P(y_i = 1) = \frac{e^{\eta_i}}{1 + e^{\eta_i}}$$

where η_i is

$$\eta_i = \alpha + \theta_{R1} + \theta_{R2} + .. + \theta_{R7} \left(\sum_{j=1}^{5} \beta_{O\_R1_{ij}} + \sum_{j=1}^{5} \delta_{D\_R1_{ij}} \right) + \left(\sum_{j=1}^{5} \beta_{O\_R2_{ij}} + \sum_{j=1}^{5} \delta_{D\_R2_{ij}} \right) + \left(\sum_{j=1}^{5} \beta_{O\_R3_{ij}} + \sum_{j=1}^{5} \delta_{D\_R3_{ij}} \right) + \ldots + \left(\sum_{j=1}^{5} \beta_{O\_R7_{ij}} + \sum_{j=1}^{5} \delta_{D\_R7_{ij}} \right) + \epsilon_i$$

where α is the intercept, θ is the coefficient for a given missed shot range, β is the rebounding offensive rebound coefficient for each player and ε is the Gaussian error. The notation R_1 to R_7 signifies the missed shot distance ranges taken into consideration for the model [see appendix figure 2 for full visual detail]. The model was fit with ~111,470 instances of an individual offensive or defensive rebound. For each instance, the shot distance, players on the court and rebound type [OREB as 1 and DREB as 0] were marked in the model.

4 Impact on Team Offensive and Defensive Strategy

As mentioned above, the importance of rebounding by the offensive or defensive team is critical because of its ability to continue or end a given possession. A main factor in the model, the location of a given missed shot attempted has a significant impact on the chances that a team rebounds the ball. In coordination with figure 2 [see Appendix], figure 1 shows the expected probability that an offensive rebound event occurs given the distance of the missed shot attempt for an average set of players on the court during the 2015 – 2016 NBA season. The shot distance that yields the highest probability of an offensive rebound is between 0-3FT [dunks, layups, etc.] and 17-22FT [long 2PT or corner 3PT attempt] from the basket at 31.5% and 28.4% respectively. However, missed shots taken between 4-16FT from the basket reduce the probability of an offensive rebound by as much as 12.7% to 18.7% [9-12FT] to 21.2% [13-16FT]. Interestingly, an offensive team has a slightly greater chance at collecting a rebound off of a missed free throw attempt [15FT] despite the defense's better positioning to prevent an offensive rebound than a shot attempt between 4-16FT from the basket.

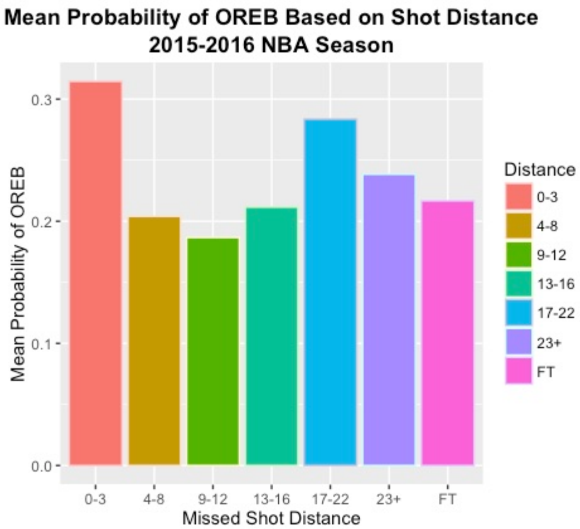

Figure 1: Shows the mean probability that an offensive rebound is collected given the missed shot attempt's distance for the 2015 - 2016 NBA season.

Despite its effort, an offense will not make every shot attempt. As a result, a team's offensive strategy should aim to minimize the probability that the defense rebounds a missed shot, ending a possession, while maximize the probability that they maintain possession by rebounding the missed shot, increasing their expected point value per possession. As highlighted in figure 1, an offensive strategy should prioritize shots taken either close to the basket [4FT] or from longer range [17-22FT]. Naturally, shorter shots are typically made at a higher efficiency while longer shots are worth more points if taken beyond the 3PT line. Yet, more importantly, if missed, these shots yield higher probabilities of an offensive rebound and thus continuation of the possession. Defensively, a team should aim to strengthen their interior presence that discourages close shots attempts and utilize a perimeter defense that forces the offense out of long shots into midrange shot attempts. Such efforts will significantly increase their probability of rebounding a missed shot, ending the possession and reducing the offense's expected point value.

5 Players on Offense – Why *Maniac Russ* Works

It is important to note that the model used in this paper does not rate a player's ability to individually collect an offensive rebound, but does measure the impact that a given player has while on the court towards the probability that an offensive rebound event occurs. Instead of looking at the box score for insight, the model highlights how a given player's style of play impacts the game with respect to an offensive rebound event. Drawn from the model's results [see Appendix table 2 -3 for summarized results], table 1 indicates that point guard Russell Westbrook's impact on the Oklahoma City Thunder's chance of collecting an offensive rebound while on the court during an offensive possession is significantly positive compared to an average NBA player, most notably on missed shots between 0-8FT and 23+FT while ranking in the top-12 in four shot distance categories. Known for a style of play that is often criticized as reckless and detrimental towards the team while on offense, the model's results suggest that "Maniac Russ's" style of play actually increases the chance of an offensive rebound and consequentially increases the Thunder's expected point value during a possession. On closer review, this should not be surprising and highlight, despite high volume shot attempts, the value Westbrook's style of play adds to the Thunder.

In general, Westbrook is one of the more aggressive dribble-drive guards in the NBA and excels at penetrating and pressuring defenses to react to his actions. This style of play can result in several actions that, although may lead to a missed shot, actually put the offense in a better position to capture an offensive rebound off a missed shot. As stated earlier, driving into the lane provokes help defenders, which can result in 1) clear paths to the rebound for teammates if Westbrook attempts and misses the shot or 2) puts Westbrook in the immediate vicinity of the ball making one more body the defense must overcome to collect the rebound. Additionally, by pressuring and collapsing the defense, Westbrook can opt to pass out towards the perimeter to likely open teammates for a shot that, even if missed, has the second highest probability of an offensive rebound.

Table 2: Shows Russell Westbrook's expected probability that an offensive rebound event occurs while he is on the court on offense given a missed shot attempt. Values are compared to each shot distance mean and Westbrook's rank amongst other NBA players by shot distance category.

Player	Russell Westbrook		OKC	
Missed Shot Distance	**Mean P(OREB)**	**R. Westbrook P(OREB)**	**+/-**	**Rank**
0 - 3 FT	31.5%	33.6%	2.2%	6
4 - 8 FT	20.4%	21.7%	1.3%	10
9 - 12 FT	18.7%	18.8%	0.1%	161
13 - 16 FT	21.2%	22.1%	0.9%	12
17 - 22 FT	28.4%	29.3%	0.9%	34
23+ FT	23.8%	26.4%	2.6%	3

Despite great additive values close to the basket, the addition of Westbrook on the court with midrange missed shot attempts yields the least amount of positive impact most notably shots between 9-12FT. This result is significant for both the Thunder and opposing teams. First, following the general trend of the model, the Thunder's expected offensive points per possession is greater when shot attempts are close to the basket or from deep while Westbrook is on the court. As result, Westbrook's style of play is more impactful if actions on the court are focused on creating inside while potentially opening outside opportunities. A sign that the offense is potentially under-performing is if shots are at midrange distances. Defensively, the converse is true as forcing the Thunder offense into midrange attempts while Westbrook is on the court reduces his advantage that he brings to the Thunder offense in terms of rebounding.

6 Players on Defense

Additionally, the model examines the probability that an offensive rebound occurs while a given player is on defense. The results of the model are summarized in the Appendix [see table 3] highlighting the best, worst and notable player's percentages. In this case lower percentages are valued greater because a low probability that an offensive rebound occurs while a player is on defense is good for his team, a probability they are trying to minimize. Examining player actions that reduce offensive rebounding probabilities could include a player's box-out ability or inclination to pursue missed shots on defense as oppose to running out on transition

7 Impact and Further Development

The main effort of this model is to highlight how a player's time on the court impacts a meaningful part of every basketball possession, rebounding. As in the case with Russell Westbrook, the model can help bring insight to why a particular player's actions benefit or inhibit a team's ability to rebound effectively. Understanding this part of the process can help inform decision makers on how to better utilize their players or team strategy with respect to rebounding. A limiting component of the model is that it does not taken into consideration if a player actually rebounded the ball or not and further development of the model should take into account this component. Potential figures that plot player probability of an offensive rebound event against his offensive rebounds collected per possession could add further insight to how much a player directly contributes to the increased probability of an offensive rebound.

8 Appendix

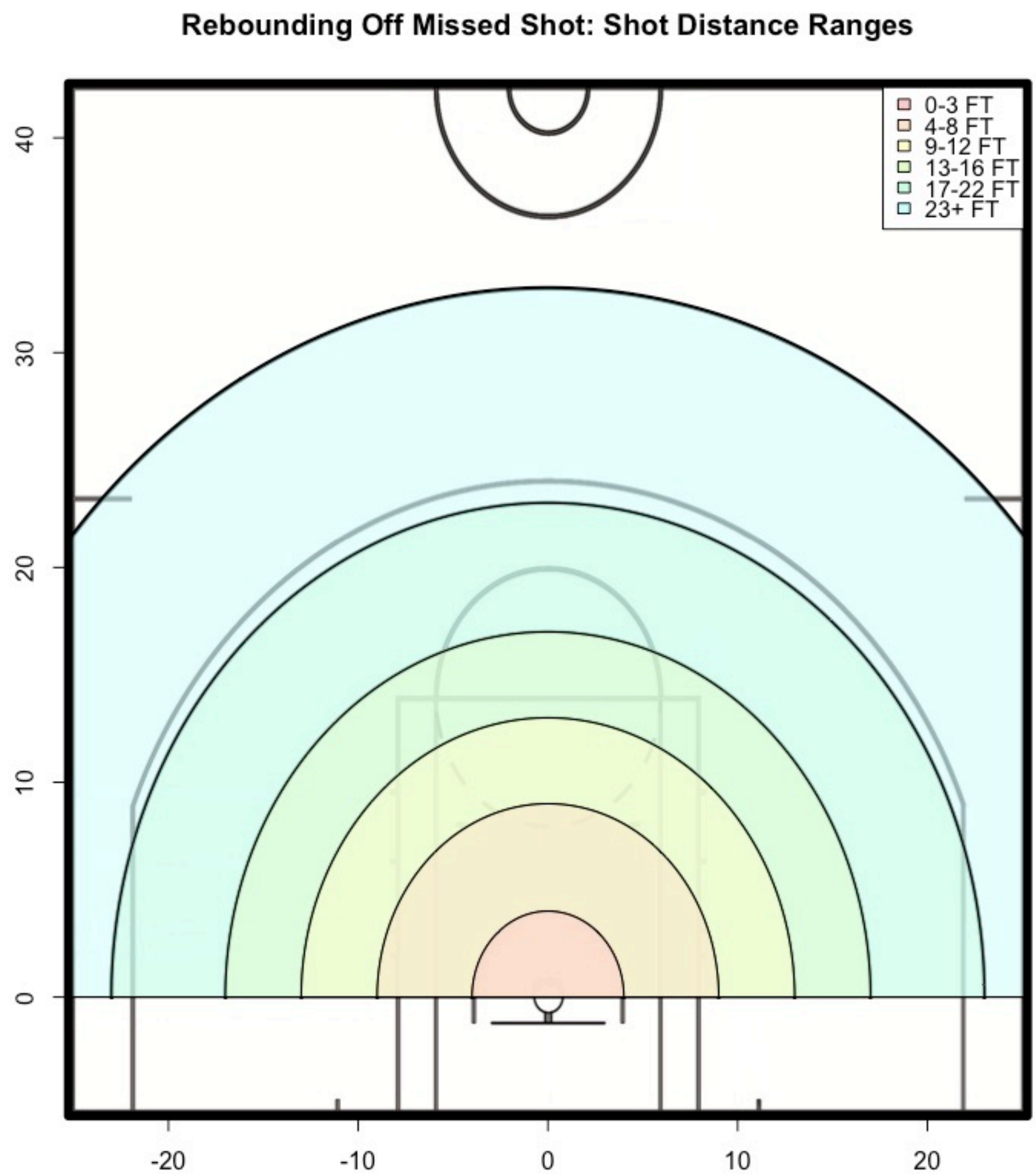

Figure 2: Is a visual representation of the various missed shot distances considered in the Rasch model with respect to an NBA court. The color regions correspond to the figure 2 in the analysis section that shows the mean probability that an offensive rebound event occurs off of a missed shot from one of the given distance ranges.

Table 2: Shows the probability of an offensive rebound occuring given the player is on the court as a member of the offensive team and also given a specific missed shot attempt. The top-5, bottom-5 and notable players are listed for each shot distance category. A postive +/- indicates that the player increases the chance that an offensive rebound is recovered (by his own team) compared to the league average given the missed shot attempt distance and the player is on the court as a member of the offensive team. A negative indicates that the player decreases the probability of an offensive rebound (by his own team) compared to the league average given the shot attempt distance and the player is on the court as a member of the offensive team.

Probability of OREB Given Player is on Offense

Rank	Missed Shot Distance / Mean P(OREB)	0 - 3 FT / 31.5%		
	Player	Team	P(OREB)	+/-
1	Andre Drummond	DET	34.1%	2.7%
2	Kevin Love	CLE	34.0%	2.6%
3	Aaron Gordon	ORL	33.7%	2.3%
4	Kentavious Caldwell-Pope	DET	33.7%	2.2%
5	Jason Terry	HOU	33.7%	2.2%
-	LeBron James	CLE	32.5%	1.1%
-	Stephen Curry	GSW	31.5%	0.0%
-	Kawhi Leonard	SAS	30.9%	-0.5%
5	Robert Covington	PHIL	29.7%	-1.7%
4	Nik Stauskas	PHIL	29.7%	-1.7%
3	Al Horford	ATL	29.7%	-1.8%
2	Kyle Korver	ATL	29.5%	-1.9%
1	Nene	WSH	28.9%	-2.5%

Rank	Missed Shot Distance / Mean P(OREB)	4 - 8 FT / 20.4%		
	Player	Team	P(OREB)	+/-
1	Clint Capela	HOU	22.9%	2.5%
2	Nikola Jokic	DEN	22.2%	1.8%
3	Zach Randolph	MEM	22.2%	1.8%
4	Steven Adams	OKC	22.0%	1.6%
5	Ed Davis	POR	21.8%	1.4%
-	LeBron James	CLE	20.4%	0.0%
-	Stephen Curry	GSW	20.4%	0.0%
-	Kawhi Leonard	SAS	19.6%	-0.8%
5	Austin Rivers	LAC	19.4%	-1.0%
4	Mario Hezonja	ORL	19.4%	-1.0%
3	Andrew Nicholson	ORL	19.3%	-1.1%
2	Donald Sloan	BKN	19.3%	-1.1%
1	Eric Gordon	NO	19.2%	-1.2%

Rank	Missed Shot Distance / Mean P(OREB)	9 - 12 FT / 18.7%		
	Player	Team	P(OREB)	+/-
1	Tony Allen	MEM	20.0%	1.4%
2	Lavoy Allen	IND	19.9%	1.3%
3	Trey Burke	UTAH	19.7%	1.1%
4	DeAndre Jordan	LAC	19.7%	1.0%
5	James Harden	HOU	19.6%	1.0%
-	LeBron James	CLE	18.5%	-0.1%
-	Stephen Curry	GSW	18.1%	-0.6%
-	Kawhi Leonard	SAS	18.8%	0.2%
5	Chandler Parsons	DAL	17.8%	-0.9%
4	Jared Dudley	WSH	17.8%	-0.9%
3	Ish Smith	PHIL	17.7%	-0.9%
2	Myles Turner	IND	17.6%	-1.0%
1	Jrue Holiday	NO	17.6%	-1.1%

Rank	Missed Shot Distance / Mean P(OREB)	13 - 16 FT / 21.2%		
	Player	Team	P(OREB)	+/-
1	Kevin Durant	OKC	22.7%	1.5%
2	Josh Richardson	MIA	22.6%	1.4%
3	Hassan Whiteside	MIA	22.4%	1.3%
4	Rudy Gobert	UTAH	22.4%	1.3%
5	E'Twaun Moore	CHI	22.4%	1.2%
-	LeBron James	CLE	21.8%	0.7%
-	Stephen Curry	GSW	21.5%	0.3%
-	Kawhi Leonard	SAS	20.5%	-0.7%
5	Nicolas Batum	CHA	20.1%	-1.0%
4	Al Jefferson	CHA	20.1%	-1.1%
3	Kobe Bryant	LAL	20.1%	-1.1%
2	John Wall	WSH	20.0%	-1.1%
1	Frank Kaminsky	CHA	19.9%	-1.3%

Rank	Missed Shot Distance / Mean P(OREB)	17 - 22 FT / 28.4%		
	Player	Team	P(OREB)	+/-
1	Alex Len	PHX	31.1%	2.7%
2	Robin Lopez	NYK	30.5%	2.2%
3	Sasha Vujacic	NYK	30.5%	2.1%
4	Rudy Gobert	UTAH	30.1%	1.8%
5	Hassan Whiteside	MIA	30.0%	1.7%
-	LeBron James	CLE	29.2%	0.8%
-	Stephen Curry	GSW	28.8%	0.4%
-	Kawhi Leonard	SAS	27.9%	-0.5%
5	Brook Lopez	BKN	26.8%	-1.5%
4	Gary Neal	WSH	26.8%	-1.5%
3	Donald Sloan	BKN	26.8%	-1.6%
2	Ish Smith	PHIL	26.6%	-1.8%
1	Courtney Lee	CHA	26.2%	-2.2%

Rank	Missed Shot Distance / Mean P(OREB)	23+ FT / 23.8%		
	Player	Team	P(OREB)	+/-
1	Enes Kanter	OKC	27.4%	3.5%
2	Lavoy Allen	IND	26.6%	2.8%
3	Russell Westbrook	OKC	26.4%	2.6%
4	Nick Collison	OKC	26.1%	2.2%
5	Ed Davis	POR	25.9%	2.1%
-	LeBron James	CLE	24.9%	1.1%
-	Stephen Curry	GSW	24.3%	0.4%
-	Kawhi Leonard	SAS	25.0%	1.2%
5	Al Horford	ATL	22.1%	-1.7%
4	Paul Pierce	LAC	22.1%	-1.8%
3	Dante Cunningham	NO	22.0%	-1.9%
2	Andrew Bogut	GSW	22.0%	-1.9%
1	Chandler Parsons	DAL	21.9%	-2.0%

Rank	Missed Shot Distance / Mean P(OREB)	Free Throw / 21.7%		
	Player	Team	P(OREB)	+/-
1	Steven Adams	OKC	22.8%	1.1%
2	Kevin Durant	OKC	22.7%	1.0%
3	Enes Kanter	OKC	22.5%	0.8%
4	Alex Len	PHX	22.5%	0.8%
5	Russell Westbrook	OKC	22.4%	0.8%
-	LeBron James	CLE	21.1%	-0.6%
-	Stephen Curry	GSW	21.1%	-0.6%
-	Kawhi Leonard	SAS	21.6%	0.0%
5	DeMar DeRozan	TOR	20.7%	-0.9%
4	Kent Bazemore	ATL	20.7%	-1.0%
3	Justise Winslow	MIA	20.7%	-1.0%
2	James Harden	HOU	20.5%	-1.2%
1	DeAndre Jordan	LAC	20.5%	-1.2%

Table 3: Shows the probability of an offensive rebound occuring given the player is on the court as a member of the defensive team and also given a specific missed shot attempt. The top-5, bottom-5 and notable players are listed for each shot distance category. A negative +/- indicates that the player decreases the chance that an offensive rebound is recovered (by the opposing team) compared to the league average given the missed shot attempt distance and the player is on the court as a member of the defensive team. A postive +/- indicates that the player increases the probability of an offensive rebound (by the opposing team) compared to the league average given the shot attempt distance and the player is on the court as a member of the defensive team.

Probability of OREB Given Player is on Defense

Rank	Missed Shot Distance Mean P(OREB)	**0 - 3 FT** **31.5%**		
	Player	Team	P(OREB)	+/-
1	RJ Hunter	BOS	29.8%	-1.7%
2	Elton Brand	PHIL	29.8%	-1.6%
3	Rondae Hollis-Jefferson	BKN	29.8%	-1.6%
4	TJ Warren	PHX	29.8%	-1.6%
5	Danilo Gallinari	DEN	29.9%	-1.6%
-	LeBron James	CLE	31.4%	0.0%
-	Stephen Curry	GSW	31.2%	-0.2%
-	Kawhi Leonard	SAS	31.3%	-0.2%
5	CJ McCollum	POR	33.3%	1.8%
4	Keith Appling	ORL	33.4%	1.9%
3	JaKarr Sampson	DEN	33.4%	2.0%
2	Austin Rivers	LAC	33.5%	2.1%
1	LaMarcus Aldridge	SAS	33.8%	2.3%

Rank	Missed Shot Distance Mean P(OREB)	**4 - 8 FT** **20.4%**		
	Player	Team	P(OREB)	+/-
1	Anthony Morrow	OKC	19.0%	-1.4%
2	Rakeem Christmas	IND	19.2%	-1.2%
3	Leandro Barbosa	GSW	19.4%	-1.0%
4	Trey Lyles	UTAH	19.4%	-1.0%
5	Kevin Garnett	MIN	19.4%	-1.0%
-	LeBron James	CLE	20.5%	0.1%
-	Stephen Curry	GSW	20.4%	0.0%
-	Kawhi Leonard	SAS	20.5%	0.0%
5	Cole Aldrich	LAC	21.7%	1.3%
4	Mario Hezonja	ORL	21.8%	1.4%
3	Alexis Ajinca	NO	21.8%	1.4%
2	Nene	WSH	22.0%	1.6%
1	Iman Shumpert	CLE	22.3%	1.9%

Rank	Missed Shot Distance Mean P(OREB)	**9 - 12 FT** **18.7%**		
	Player	Team	P(OREB)	+/-
1	Matt Bonner	SAS	17.4%	-1.3%
2	Deron Williams	DAL	17.6%	-1.1%
3	Goran Dragic	MIA	17.7%	-0.9%
4	Danny Green	SAS	17.8%	-0.9%
5	Marco Belinelli	SAC	17.8%	-0.9%
-	LeBron James	CLE	18.9%	0.3%
-	Stephen Curry	GSW	18.4%	-0.3%
-	Kawhi Leonard	SAS	18.7%	0.0%
5	Derrick Favors	BKN	19.5%	0.9%
4	Tony Wroten	MEM	19.6%	0.9%
3	Jusuf Nurkic	DEN	19.6%	0.9%
2	Seth Curry	SAC	19.6%	0.9%
1	Austin Rivers	LAC	19.6%	0.9%

Rank	Missed Shot Distance Mean P(OREB)	**13 - 16 FT** **21.2%**		
	Player	Team	P(OREB)	+/-
1	Garrett Temple	WSH	20.0%	-1.2%
2	Jahlil Okafor	PHIL	20.1%	-1.1%
3	Karl-Anthony Towns	MIN	20.1%	-1.0%
4	Trevor Ariza	HOU	20.2%	-1.0%
5	Kendrick Perkins	NO	20.2%	-1.0%
-	LeBron James	CLE	21.2%	0.1%
-	Stephen Curry	GSW	20.7%	-0.4%
-	Kawhi Leonard	SAS	21.1%	0.0%
5	Patrick Beverley	HOU	22.1%	1.0%
4	Ray McCallum	SAC	22.2%	1.1%
3	Jason Smith	ORL	22.3%	1.1%
2	Ersan Ilyasova	ORL	22.3%	1.2%
1	Tayshaun Prince	MIN	22.3%	1.2%

Rank	Missed Shot Distance Mean P(OREB)	**17 - 22 FT** **28.4%**		
	Player	Team	P(OREB)	+/-
1	Nick Young	LAL	26.6%	-1.8%
2	Kyle Korver	ATL	26.6%	-1.8%
3	Ian Clark	GSW	26.7%	-1.7%
4	Marc Gasol	MEM	26.7%	-1.7%
5	Jeremy Lamb	CHA	26.8%	-1.6%
-	LeBron James	CLE	28.1%	-0.3%
-	Stephen Curry	GSW	28.8%	0.4%
-	Kawhi Leonard	SAS	28.2%	-0.1%
5	Jeff Ayres	LAC	30.0%	1.6%
4	Wayne Ellington	BKN	30.0%	1.7%
3	Kevon Looney	GSW	30.1%	1.7%
2	J.R. Smith	CLE	30.2%	1.9%
1	Jordan Adams	MEM	30.3%	1.9%

Rank	Missed Shot Distance Mean P(OREB)	**23+ FT** **23.8%**		
	Player	Team	P(OREB)	+/-
1	O.J. Mayo	MIL	21.7%	-2.2%
2	James Young	BOS	22.1%	-1.7%
3	Evan Turner	BOS	22.3%	-1.6%
4	DeMar DeRozan	TOR	22.3%	-1.5%
5	Wesley Johnson	DAL	22.4%	-1.5%
-	LeBron James	CLE	23.8%	-0.1%
-	Stephen Curry	GSW	23.7%	-0.1%
-	Kawhi Leonard	SAS	23.7%	-0.1%
5	Derrick Rose	CHI	25.4%	1.6%
4	JaKarr Sampson	DEN	25.4%	1.6%
3	Paul Millsap	ATL	25.5%	1.6%
2	Devin Booker	PHX	25.9%	2.1%
1	Arron Afflalo	NYK	26.1%	2.3%

Rank	Missed Shot Distance Mean P(OREB)	**Free Throw** **21.7%**		
	Player	Team	P(OREB)	+/-
1	Zaza Pachulia	DAL	12.6%	-9.1%
2	Karl-Anthony Towns	MIN	20.6%	-1.0%
3	Delon Wright	TOR	20.7%	-1.0%
4	David West	SAS	20.7%	-1.0%
5	Nikola Pekovic	MIN	20.8%	-0.9%
-	LeBron James	CLE	21.6%	0.0%
-	Stephen Curry	GSW	21.4%	-0.3%
-	Kawhi Leonard	SAS	22.3%	0.6%
5	Goran Dragic	MIA	22.3%	0.7%
4	Jorge Gutierrez	CHA	22.4%	0.7%
3	Rudy Gay	MEM	22.4%	0.7%
2	Jrue Holiday	NO	22.5%	0.8%
1	Robert Sacre	LAL	22.6%	1.0%

Code

```r
###################### - Parse PBP Data into Rebound ONLY Data - ######################
# Set directory to correct location. Read in data.
setwd("~/Desktop/NBA PBP Data/CSV/2015-16")
data_pbp = read.csv("[10-20-2015]-[05-01-2016]-combined-stats.csv", header = TRUE)

# Get all players out of data.
players = unlist(data_pbp[, 4:13])

# Get all row instances of a rebound.
num.rebounds = sum(data_pbp$type == "rebound offensive") + sum(data_pbp$type == "rebound defensive")

# Setup matrix to collect synthesized rebound data.
new.data = matrix(nrow = num.rebounds, ncol = 13)
colnames(new.data) = c("O1", "O2", "O3", "O4", "O5", "D1", "D2", "D3", "D4", "D5", "responsible", "type", "shot_distance")

counter = 1
for(i in 1:length(data_pbp$event_type)){
  # Check OREB case.
  if(data_pbp$type[i] == "rebound offensive"){
    # Collect the players on the court for that event.
    away.players = as.character(unlist(data_pbp[i, 4:8]))
    home.players = as.character(unlist(data_pbp[i, 9:13]))

    # If the player is an away player.
    if(data_pbp$player[i] %in% away.players){
      for(j in 1:5){
        # Away team on offense.
        new.data[counter, j] = away.players[j]

        # Home team on defense.
        new.data[counter, j + 5] = home.players[j]
      }
      new.data[counter, 11] = as.character(data_pbp$player[i])
      new.data[counter, 12] = "OREB"

      # FT have no shot distance, else store shot distance.
      if(is.na(data_pbp$shot_distance[i - 1])){
        new.data[counter, 13] = "FT"
      } else{
        new.data[counter, 13] = data_pbp$shot_distance[i - 1]
      }

    } else{
      for(j in 1:5){
        # Home team on offense.
        new.data[counter, j] = home.players[j]

        # Away team on defense.
        new.data[counter, j + 5] = away.players[j]
      }
      new.data[counter, 11] = as.character(data_pbp$player[i])
      new.data[counter, 12] = "OREB"

      if(is.na(data_pbp$shot_distance[i - 1])){
        new.data[counter, 13] = "FT"
      } else{
        new.data[counter, 13] = data_pbp$shot_distance[i - 1]
      }
    }
    counter = counter + 1

  # Check DREB case.
  } else if(data_pbp$type[i] == "rebound defensive"){
    # Collect the players on the court for that event.
    away.players = as.character(unlist(data_pbp[i, 4:8]))
    home.players = as.character(unlist(data_pbp[i, 9:13]))

    # If the player is an away player.
    if(data_pbp$player[i] %in% away.players){
      for(j in 1:5){
        # Home team on offense.
        new.data[counter, j] = home.players[j]

        # Away team on defense.
        new.data[counter, j + 5] = away.players[j]
      }
      new.data[counter, 11] = as.character(data_pbp$player[i])
      new.data[counter, 12] = "DREB"

      if(is.na(data_pbp$shot_distance[i - 1])){
        new.data[counter, 13] = "FT"
      } else{
        new.data[counter, 13] = data_pbp$shot_distance[i - 1]
      }
    } else{
      for(j in 1:5){
        # Away team on offense.
        new.data[counter, j] = away.players[j]

        # Home team on defense.
        new.data[counter, j + 5] = home.players[j]
      }
      new.data[counter, 11] = as.character(data_pbp$player[i])
      new.data[counter, 12] = "DREB"

      if(is.na(data_pbp$shot_distance[i - 1])){
        new.data[counter, 13] = "FT"
      } else{
        new.data[counter, 13] = data_pbp$shot_distance[i - 1]
      }
    }
    counter = counter + 1
  } else{
    # Do nothing, skip to next row
  }
}
# Store data into new .csv file.
write.csv(new.data, file = "NBA_Rebound_Data_2015-2016.csv")
######################### - Rasch Model Setup - ########################
# Install and require necessary packages, set directory and read in data.
install.packages('Matrix')
```

```r
require(Matrix)
require(glmnet)
require(ggplot2)

setwd("~/Desktop/NBA PBP Data/CSV/2015-16")
data = read.csv("NBA_Rebound_Data_2015-2016.csv", header = TRUE)

# Collect all players on court during a rebound instance.
players = unlist(data[,2:11])
num.players = length(unique(players))

# Create OFF and DEF tag for each player.
prefix = rep(c("OFF", "DEF"), each = 5*nrow(data))

# Create shot distance tag for each rebound instance.
distance_range = rep("", nrow(data))

distance_range[data$shot_distance %in% c(0:3)] = "0-3"
distance_range[data$shot_distance %in% c(4:8)] = "4-8"
distance_range[data$shot_distance %in% c(9:12)] = "9-12"
distance_range[data$shot_distance %in% c(13:16)] = "13-16"
distance_range[data$shot_distance %in% c(17:22)] = "17-22"
distance_range[data$shot_distance %in% c(23:94)] = "23+"
distance_range[data$shot_distance == "FT"] = "FTA"

# Tag all players with OFF/DEF and shot distance.
tag = paste(prefix, rep(distance_range, 10), players, sep = "_" )
tag.factor = as.factor(tag)

# Create X matrix.
i.distance = rep(1:nrow(data))

i.players = rep(1:nrow(data), 10)

i = c(i.distance, i.players)
j = as.numeric(as.factor(c(distance_range, tag)))
X = sparseMatrix(i, j)


# Create Y matrix. OREB is 1, DREB is 0.
Y = rep(0, nrow(data))
Y[data$type == "OREB"] = 1

###################### - Analysis of Rasch Model Data - ######################
time = Sys.time()
rasch = cv.glmnet(X, Y, alpha = 0, standardize = FALSE, family = "binomial", lambda = exp(seq(-10,0,length=100)))
print(Sys.time() - time)
plot(rasch)

coef = coef(rasch, s = "lambda.min")[, 1]
names(coef) = c("Intercept", sort(unique(distance_range)), sort(unique(tag)))

alpha = coef[1]
D_1 = coef[2]
D_2 = coef[3]
D_3 = coef[4]
D_4 = coef[5]
D_5 = coef[6]
D_6 = coef[7]
D_7 = coef[8]

###################### - Calculate OREB Probabilities - #########################
OFF_D1 = sort(exp(alpha + D_1 + coef[grep("OFF_0-3", names(coef))])/(1 + exp(alpha + D_1 + coef[grep("OFF_0-3", names(coef))])))
DEF_D1 = exp(alpha + D_1 + coef[grep("DEF_0-3", names(coef))])/(1 + exp(alpha + D_1 + coef[grep("DEF_0-3", names(coef))]))
OFF_D1_AVG = exp(alpha + D_1)/(1 + exp(alpha + D_1))

OFF_D2 = sort(exp(alpha + D_2 + coef[grep("OFF_4-8", names(coef))])/(1 + exp(alpha + D_2 + coef[grep("OFF_4-8", names(coef))])))
DEF_D2 = exp(alpha + D_2 + coef[grep("DEF_4-8", names(coef))])/(1 + exp(alpha + D_2 + coef[grep("DEF_4-8", names(coef))]))
OFF_D2_AVG = exp(alpha + D_2)/(1 + exp(alpha + D_2))

OFF_D3 = sort(exp(alpha + D_3 + coef[grep("OFF_9-12", names(coef))])/(1 + exp(alpha + D_3 + coef[grep("OFF_9-12", names(coef))])))
DEF_D3 = exp(alpha + D_3 + coef[grep("DEF_9-12", names(coef))])/(1 + exp(alpha + D_3 + coef[grep("DEF_9-12", names(coef))]))
OFF_D3_AVG = exp(alpha + D_3)/(1 + exp(alpha + D_3))

OFF_D4 = sort(exp(alpha + D_4 + coef[grep("OFF_13-16", names(coef))])/(1 + exp(alpha + D_4 + coef[grep("OFF_13-16", names(coef))])))
DEF_D4 = exp(alpha + D_4 + coef[grep("DEF_13-16", names(coef))])/(1 + exp(alpha + D_4 + coef[grep("DEF_13-16", names(coef))]))
OFF_D4_AVG = exp(alpha + D_4)/(1 + exp(alpha + D_4))

OFF_D5 = sort(exp(alpha + D_5 + coef[grep("OFF_17-22", names(coef))])/(1 + exp(alpha + D_5 + coef[grep("OFF_17-22", names(coef))])))
DEF_D5 = exp(alpha + D_5 + coef[grep("DEF_17-22", names(coef))])/(1 + exp(alpha + D_5 + coef[grep("DEF_17-22", names(coef))]))
OFF_D5_AVG = exp(alpha + D_5)/(1 + exp(alpha + D_5))

OFF_D6 = sort(exp(alpha + D_6 + coef[grep("OFF_23+", names(coef))])/(1 + exp(alpha + D_6 + coef[grep("OFF_23+", names(coef))])))
DEF_D6 = exp(alpha + D_6 + coef[grep("DEF_23+", names(coef))])/(1 + exp(alpha + D_6 + coef[grep("DEF_23+", names(coef))]))
OFF_D6_AVG = exp(alpha + D_6)/(1 + exp(alpha + D_6))

OFF_D7 = sort(exp(alpha + D_7 + coef[grep("OFF_FT", names(coef))])/(1 + exp(alpha + D_7 + coef[grep("OFF_FT", names(coef))])))
DEF_D7 = exp(alpha + D_7 + coef[grep("DEF_FT", names(coef))])/(1 + exp(alpha + D_7 + coef[grep("DEF_FT", names(coef))]))
OFF_D7_AVG = exp(alpha + D_7)/(1 + exp(alpha + D_7))

###################### - Create Mean P(OREB) Comparison Table - #########################
bar.plot = data.frame(Distance = c("0-3", "4-8", "9-12", "13-16", "17-22", "23+", "FT"),
                      average = c(OFF_D1_AVG, OFF_D2_AVG, OFF_D3_AVG, OFF_D4_AVG, OFF_D5_AVG, OFF_D6_AVG, OFF_D7_AVG))
bar.plot$Distance = factor(bar.plot$Distance, levels = bar.plot$Distance)

ggplot(data = bar.plot, aes(x = Distance, y = average, fill = Distance)) +
  geom_bar(stat = "identity", col = c(rgb(255/255,204/255,204/255),
                                      rgb(255/255,229/255,204/255),
                                      rgb(255/255,255/255,204/255),
                                      rgb(204/255,255/255,204/255),
                                      rgb(204/255,204/255,229/255),
                                      rgb(204/255,255/255,255/255),
                                      rgb(229/255,204/255,255/255))) +
  xlab("Missed Shot Distance") +
  ylab("Mean Probability of OREB") +
  ggtitle("Mean Probability of OREB Based on Shot Distance\n2015-2016 NBA Season") +
  theme(plot.title = element_text(lineheight=1, face="bold"))
```