Applying WAR to the NBA

Vivek Divakarla, Daniel Draymore, Keshav Iyer, Jacob Norris, Jason Stitt
Northeastern University, Boston, MA, USA
DS2500 - Rachlin/Strange
4/21/2021

Introduction

In basketball, the goal of a team is to win games. As analytics have evolved, data scientists have started to heavily influence the sports world, with managers attempting to use data to improve teams. Wins Above Replacement (WAR) is a statistic that originates from baseball, with a goal to measure how much better a player is than an average replacement player would be. This metric aims to encompass all of the contributions a player has to a team, and thus is a very complex formula. The sabermetric community has spent years refining the formula for the baseball world. However, the translation to the basketball world has not exactly happened. There are a few different WAR metrics, but they are very volume based, meaning the league's stars who take a lot of shots per game will be at the top of the list, and their efficiency is not heavily factored. Tying back to the first two sentences, we wanted to create our own Wins Above Replacement metric incorporating efficiency, with the goal of trying to optimize a team's assets/salary cap to create the best team they possibly can. After creating Efficiency WAR, we decided to model a Volume WAR that used popular NBA counting stats as opposed to efficiency. Finally, we combined the two WAR's with salary to understand the league's most undervalued and overvalued players. Below, you will find how we created these WAR statistics in addition to our analysis once they were created.

Data Sources and Methods

To start our project, we obtained our data from basketball reference [1], a website that tracks individual and team data in the NBA and has a wide variety of stats, both basic and advanced, to view. We then transferred the data we planned to use into .csv files and used these as the basis for our project. Using pandas, we combined multiple csv files, with basic stats, advanced stats, and salary into one dataframe. After we had our data in a singular df, we took a look at the statistics we had at our disposal and visualized some of our data in order to get a grasp on some correlations amongst the variables.

Using our prior basketball knowledge paired with these visualizations, along with some trial and error, we proceeded to create our first version of WAR, which we named "Efficiency WAR". This stat ignored volume-based statistics such as points, assists, and rebounds per game and instead relied upon % stats like 2 point field-goal percentage, assist percentage, and rebound percentage. We factored out players who had played under 5 games, had played under 50 minutes, and who had attempted less than 10 field goals so as not to skew the data on such a small sample size. We also had to decide on weights to give to each variable, since not all stats in the NBA are created equal. A statistic such as turnover percentage is weighted heavily, because retaining the ball and creating a scoring opportunity are two of the most important things in the modern NBA. Once we had our set of variables and associated weights, we created an "average" player in order to compare every player in the dataset to the average. This is consistent with our definition of WAR, which is a player's value above a replacement level player, which we interpreted as an average player. We then only needed to subtract the calculated player's statistics from the average player's statistics, then multiply by the respective weights, and sum up the individual's stats to come to our Efficiency WAR.

After creating Efficiency WAR, we decided to move into creating another form of WAR which we tagged as "Volume WAR". Unlike efficiency, this WAR used the 5 main statistics in basketball: Points, Rebounds, Assists, Blocks, and Steals. These are all statistics that are dependent upon pure volume, or how *many* of each stat a player gets in a game, hence the name of the WAR. They are heavily influenced on a player's playing time and their role on a team.

This is valuable because the numbers a player is producing in a game are extremely important to teams, sometimes even more so than their respective efficiencies. We again subtracted a player's volume stats from the league average volume stats in order to properly compare them to a replacement-level player. We then divided each by each stats standard deviation to demean the statistics. We summed the 5 differences together in order to come to our final Volume WAR.

As our final version of WAR, we created a simple formula to incorporate a player's salary into their value above an average player, which we titled "Salary WAR". We weighted Volume WAR and Efficiency WAR using some intuition regarding the value of a player's production vs their efficiency, and added the two together. After this, we took the combined WAR's and added the difference between a player's salary and the average salary in the NBA. We then scaled this number down by 1,000,000 in order to make the WAR more coherent to the viewer and effectively create a salary WAR that is portrayed in millions of dollars.

After we had our 3 calculated WAR statistics, we performed a variety of comparisons among them and also comparisons between the WAR and other statistics/information such as a player's position, their age, and their field goals attempted. Our analysis of these comparisons/plots, among other analysis, is contained in the Analysis and Results section below.

Use Cases

The obvious target group for this project is NBA general managers and front office staff. The results of this analysis give insight into not just the production a player has, but the value they add to a team. Knowing both production and efficiency is the key to building a good team, as you need both players who can play big minutes and put up big numbers, and players who can step in for specific situations and deliver in a limited role. This program analyzes total production and efficiency, and it compares them, which helps reveal how different players help their team win.

While NBA front offices are the main target for this project, they are not the only group that can benefit from this analysis. For people who are fantasy sports enthusiasts, the volume analysis of this program can help identify players who have big statlines, and would therefore score highly in fantasy sports. In daily fantasy sports, where players have “salaries” that limit who a user can pick, the salary and efficiency WAR analysis can be used to identify cheaper players that can still produce.

With the emergence of online sports betting, more and more people have become interested in making picks on games. While our program doesn’t specifically help predict specific game outcomes, it shows trends in team and player play styles. Our results can be extrapolated to make specific player matchup predictions, and our production WAR metric can be used to predict a teams total production. Betting odds are available for almost any aspect of the game, from winners to individual player stats, and the results of our analysis can be used as a predictive model to make more informed picks.

Analysis and Results

We started our analysis off with some exploratory analysis comparing the relationships between all the statistics we deemed important. We used a heatmap to showcase and analyze these relationships to help decide/reinforce what graphical representations would be substantive and informative. First, we created a simple scatterplot of salary vs points per game. This chart shows a moderate positive correlation, which makes sense as the players who score more points should be paid more for their contributions to the team. The most meaningful graphs we created will be explained throughout this section.

Why Efficiency is Valuable

Before any major analysis was done, we had to answer the question of why an efficiency based WAR added any value. There are already volume based stats that many teams use. To answer this question, we actually compared players' efficiency WAR metrics to the number of shots they attempted, and what we found was a clear positive correlation between player volume and efficiency. This cemented the importance of efficiency. Seeing efficiency increase with volume led to the conclusion that players with low volume but high efficiency could realistically maintain or improve their efficiency with more playing time. Seeing this correlation reveals the added value of analyzing efficiency, and offers insight into what this analysis can add to the already vast array of advanced analytics that are available to teams. We've included the list of the top 4 players we identified through Efficiency WAR. Jokic, Doncic, and Curry are widely regarded as top players in the NBA, which helps to somewhat verify our Efficiency WAR. Ingles, however, is not known as a star player. It is players like Ingles we were hoping to identify in our Efficiency WAR creation.

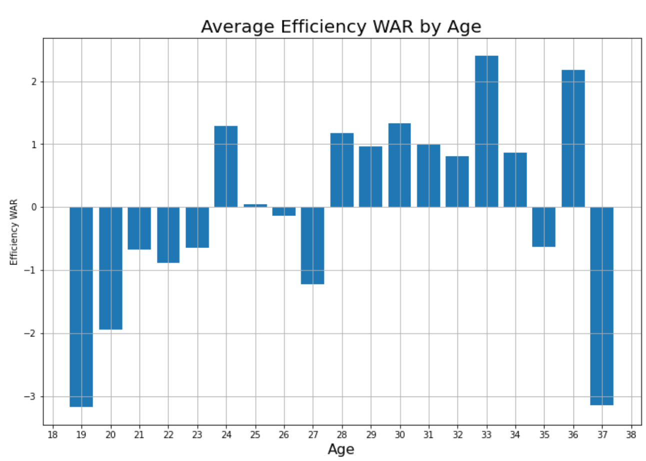
Player	**Efficiency WAR**
Nikola Jokić	12.19

Joe Ingles	8.98
Luka Dončić	8.90
Stephen Curry	8.80

Age vs Efficiency

We were very interested in determining how age impacted a players efficiency. We thought that older players would be more efficient because they have been around longer and in theory, have acquired more skills. Plotting the relationship between age and efficiency WAR did indeed show that older players were more efficient than younger players. (*Figure 1)* Upon further analysis, we hypothesised another reason for the upward trend could be that inefficient older players will get cut and forced out of the league, whereas an inefficient younger player will get more leeway because they are expected to make more mistakes. This could potentially be what is leading to a skew in the distribution of inefficient players between older and younger players.

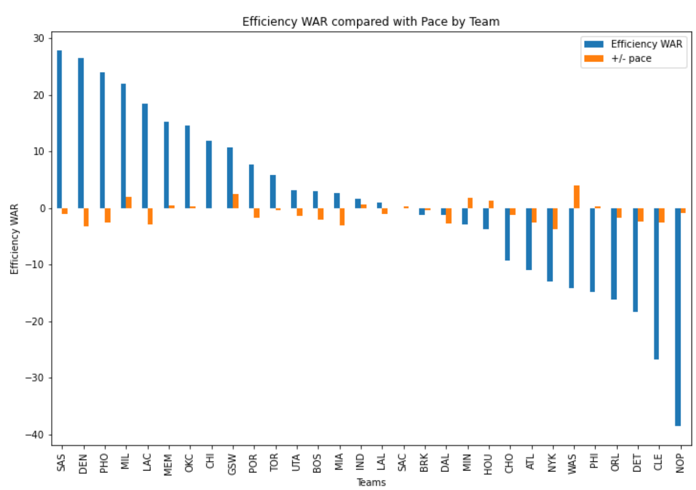
Figure 1



Pace

We wanted to get a better understanding of why there were some below average teams, record-wise, with higher efficiency WARs. We hypothesised that these worse teams played with a higher pace which is why they have higher WARs. However, when we compared the relationship, shown in *Figure 2*, we saw no clear association between the two variables. After analysing the data, we concluded that a higher pace was likely not the reason for these worse teams to have high efficiency WARs. Our original hypothesis was ignoring the fact that efficiency WAR is taking out the biases that high volume can introduce.

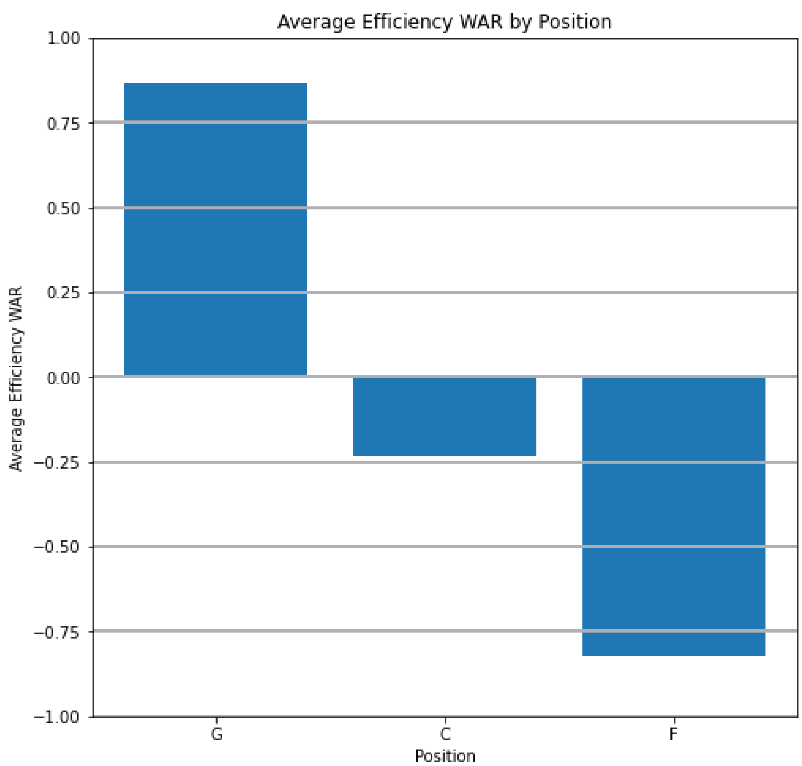
Figure 2



Position vs Efficiency War

We also decided to obtain the average efficiency war by position. We felt this would be an interesting analysis, because each team (normally) needs players at each position of guard, forward, and center, so simply choosing players based on their WAR without regard to positioning could be detrimental in a team building aspect. We identified through our analysis the fact that guards often had the highest Efficiency WAR by a wide margin. In fact, the other two positions had average Efficiency WAR's that were negative. One of the largest factors contributing to this margin could be because of the fact that we accounted for 3-point and 2-point field goals separately, each with their own weights. As most centers and forwards take significantly less 3-point shots, and their percentages are much lower on average, a guard's efficiency WAR is bound to be higher. In addition, guards often produce more assists than the other positions in the modern NBA game. We believe that for reasons like these, guards had much higher average efficiency in our version of WAR.

Figure 3



Why Volume is Valuable

Volume statistics must be considered when a player is evaluated solely on efficiency. For example, young bench players often only play in “garbage time,” which is when the game becomes a blowout so the outcome of the game has already been decided, where it is easy for players to boost their per game stats to an unrealistic level. Substitutes play against other substitutes and neither team cares too much about the game anymore since the outcome is known. This is why if a player plays in 3 minutes per game but scores 4 points, it seems more impressive than it actually is and the player should be downgraded for it. Ingles, who has the second highest Efficiency WAR is not near the top five players in Volume WAR largely because of the bench role he serves for his team and deferring to the other star players around him.

Looking at the results for Volume WAR, some of the trends made sense while others were very intriguing. The top 5 in this category were Jokic, Giannis, Gobert, Harden, and Sabonis.

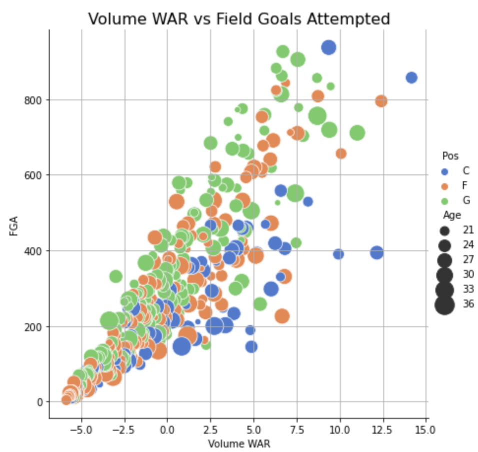
Player	Pos	Age	Tm	G	GS	MP	FG	FGA	3PP	...	AST%	STL%	BLK%	TOV%	ppg	Efficiency WAR	fga/m	+/- ppg	+/- fga	Volume WAR
Nikola Jokić	C	25	DEN	46	46	1640	489	857	0.428	...	39.9	2.2	1.8	12.6	26.93	12.192418	0.522561	1.638255	1.977009	14.174351
Giannis Antetokounmpo	F	26	MIL	43	43	1461	442	795	0.303	...	30.0	1.6	3.3	13.7	28.42	6.226098	0.544148	1.784226	1.761636	12.417811
Rudy Gobert	C	28	UTA	46	46	1414	256	394	0.000	...	5.8	0.8	7.6	12.5	14.72	-0.714382	0.278642	0.442076	0.368660	12.161678
James Harden	G	31	TOT	40	40	1527	329	711	0.362	...	44.6	1.6	1.7	16.7	26.08	8.354818	0.465619	1.554983	1.469840	11.026761
Domantas Sabonis	F	24	IND	45	45	1626	345	656	0.308	...	25.8	1.6	1.3	16.6	20.44	2.267138	0.403444	1.002448	1.278784	10.086903

Jokic, Giannis, and Harden are 3 of the league’s best players. They are all superstars putting up great all-around numbers and leading some of the best teams in the league. They also are doing this on very high efficiency, as Jokic has the top mark for Efficiency WAR while Harden and Giannis are among the top 20. Sabonis and Gobert are two interesting cases. Neither of the two put up super high scoring numbers, and their efficiency WAR scores are not near the top at all.

Gobert's contributions in rebounding and on the defensive end, and Sabonis' all-around game as a big man gets them to the top of this list. One constant between all of these players is that they all play big minutes, and have not missed out on many games. These players prove their value by not only being productive on the court, but also by being reliable and durable by playing almost every game. In a long season, durability is a valuable trait to have.

Looking farther than the top 5, there are a few trends to notice, particularly in the plot below.

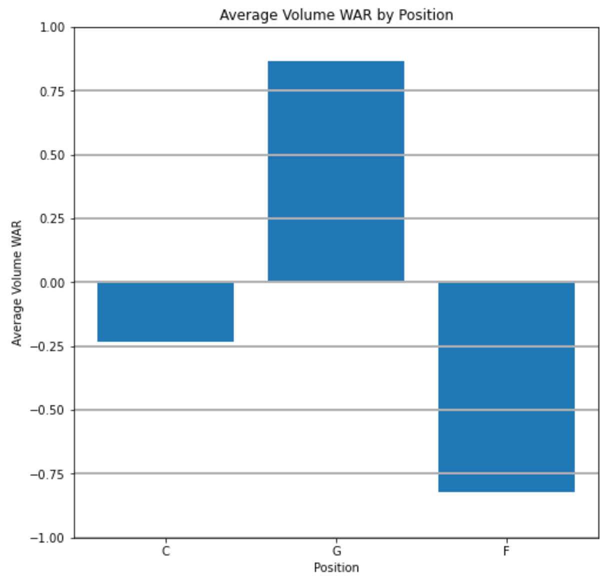
Figure 4



Viewing this graph, it's clear there is a positive correlation between Volume WAR and field goals attempted, which would make sense here. There are a few outliers of players with a high volume WAR and not many FGA, but they are mainly big men who contribute in other areas of the game. Another trend based on the color of the bubbles is that Guards and Forwards are the players which attempt most of the field goals, and have many of the higher marks for Volume WAR. This shows a larger trend in the entire league. With the exception of 2-3 superstar centers, Guards and Forwards are the ones who are the focal points of most offenses, and expected to carry the load. The majority of the centers do not shoot as much, but provide their value in other categories, such as defense and rebounding.

Position vs. Volume WAR

Figure 5



The above graph shows the average production WAR by position. Compared to the efficiency metric, the volume data is largely similar. Guards tend to put up the biggest numbers, which makes sense as they have the most well rounded stat lines. The big difference is with centers. They had extremely bad efficiency but they have a close to neutral volume. This suggests that centers operate at high volume and low efficiency, which makes sense as they don't shoot threes as often and rely on rebounds for stats (rebounds aren't weighted heavily in efficiency metrics). This data offers insight into what positions are better for volume, and the efficiency data can help find the most valuable players within each position.

Why Salary is Valuable

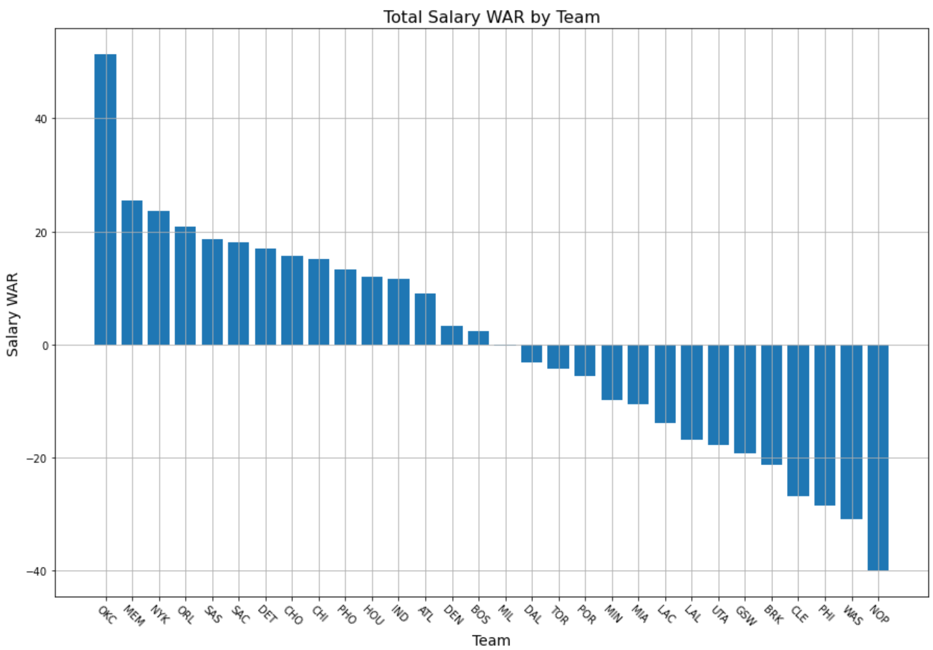
A player's salary is one of the most important factors to NBA teams. Each team only has a certain amount of money they can spend on players each year. It is a debate for whether managers should strive for a team comprised of only solid players, a couple of superstars paired with a bunch of subpar players, or anywhere in between. One of the most important and desired qualities of managers is the ability to find game changing players who will not break the bank to acquire. Championship teams always have players that leave us baffled by what a bargain their contract was. This is why we are trying to find players with the perfect combination of impressive efficiency with adequate volume who are under an affordable contract. These players should be a target for most teams in the next transfer window. The top 5 players according to Salary WAR are included below.

Player	Pos	Age	Tm	G	GS	MP	FG	FGA	3PP	...	ppg	Efficiency WAR	+/- ppg	+/- fga	Volume WAR	eWAR v vWAR	Total WAR	Salary WAR
Chris Silva	F	24	TOT	12	0	84	10	14	1.000	...	2.67	8.249858	-0.738428	-0.951367	-5.169574	13.419431	4.224028	11.455332
Drew Eubanks	C	23	SAS	27	0	351	48	95	1.000	...	4.63	7.134658	-0.546412	-0.669993	-2.365657	9.500315	4.284563	11.413284
Tre Jones	G	21	SAS	21	0	91	15	31	1.000	...	1.71	6.895218	-0.832476	-0.892314	-5.486677	12.381894	3.180649	11.031624
Robert Williams	C	23	BOS	40	3	721	135	189	0.000	...	7.70	3.493858	-0.245653	-0.343460	4.821014	-1.327156	3.892004	10.611369
Tyrese Haliburton	G	20	SAC	41	10	1237	208	427	0.423	...	13.17	6.272018	0.290227	0.483294	4.300590	1.971427	5.680589	10.598034

As one can see from the table, the top 5 players have started very few of their team's games; these players are all being paid below the average salary, and are highly efficient according to Efficiency WAR. Their Volume WAR is low, which can be attributed to the fact that all, aside from Haliburton, have played well under 50% of their team's total minutes. These are the players we feel are undervalued in the current market and are examples of players teams should attempt to either trade for or sign. Salary WAR's biggest strength is to help teams find bargains, which are extremely important if a team has championship aspirations.

Total Salary WAR for each team is included in figure 6, in descending order:

Figure 6



This visualization is informative, especially from a team's perspective, to see how their players and corresponding contracts match up against the rest of the league. A player has a high Salary WAR if they play with high efficiency and/or volume and have a team-friendly contract. Therefore, a team's total Salary WAR is high if they have many players that fit this category. It is important to note that Salary WAR should be used to find players that are "hidden gems," not super stars because their high contracts will significantly decrease their score. Therefore, this graph showcases more so which teams have a lot of hidden gems rather than which teams are better. For example, the Oklahoma City Thunder have the highest total Salary WAR according to the graph, but are currently in 13th place in the West (not good). The Thunder have an extremely young and cheap roster full of huge potential, but not many of them have reached it yet. Considering their roster and how they do not have stars, that is why their Salary WAR is high and they do not win games. Other teams should consider trading for or acquiring players on this team. On the other hand, the New Orleans Pelicans are in last in the graph above and 11th in the West (also not good). They have a bunch of underperforming players on huge salaries, resulting in horrible Salary WARs. Other NBA teams probably should look elsewhere for potential trade and acquisition candidates.

Conclusion

Through creating new and improved WAR metrics and applying them to the sport of basketball, we think teams can significantly increase their odds of winning a championship by drawing insights from this project. Our Efficiency WAR calculates which players win the most games for their team based on their efficiency. Our Volume WAR filters out the pool of players who only have efficient statistics because of the role they play on their team and upgrades those who put up high numbers, regardless of efficiency. Our Salary WAR highlights the best players at the cheapest price by combining the first two WARs. NBA teams should use our algorithms to address their needs, by identifying and acquiring players whose contracts fit their budgets and who will help them secure more wins. Teams should also give their current players with higher efficiency WAR's more playing time. Taking these actions has the potential to vault a team up the standings and make them a legitimate annual contender to win the NBA Championship.

Author Contributions

Keshav - Helped analyze strengths and weaknesses of prior successful teams to identify what stats were important for different WAR metrics and helped determine weights for our WAR equation. Worked especially on efficiency WAR.

Jason- Gathered original data, helped determine proper stats and weights when creating the formulas for our WAR's. Aided in determining the graphs to be created using WAR. Analyzed results of WAR against other variables and the results of graphs produced.

Daniel - Assisted in coming up with ideas regarding the four different WAR metrics we would calculate and how/why we would calculate them. Figured out which types of graphs would best showcase our results and looked up their documentation for how to code them.

Jacob - Wrote code to clean the data, apply the WAR formulas, and analyze the data. Created different visualizations to show the relationships between variables. Also assisted with analyzing our WAR metrics.

Vivek - Contributed to the discussion when it came to deciding which stats to use/how to utilize them in each of our WAR metrics. Focused especially on the Salary WAR and Production WAR and helped with the debugging process.

References

[1] "Basketball Statistics and History." *Basketball*, www.basketball-reference.com/.