Discussing the 2024 NBA Defensive Player of the Year

Peter D. DePaul III

05-25-2024

Contents

1 Introduction

In the NBA, the Defensive Player of the Year award can often be controversial. This primarily is because defense is difficult to quantify and interpret in comparison to offensive production. The goal for this is to make the best effort to determine objectively who the most efficient and self-producing defender was this season. Keep in mind we are focusing on the individuals impact to their team.

2 Variable Table

When it came to choosing the variables for this analysis that I want to focus on, I arbitrarily will be focusing on the following individual player variables.

Table 1: Table of Variables

Variable	Description	Type
Deflections	Number of deflections	Integer
charges	Number of charges drawn	Integer
contested_shots	Number of contested shots	Integer
BLK	Number of blocks	Integer
STL	Number of steals	Integer
DBPM	Defensive Box Plus/Minus	Numeric
DRB	Defensive Rebounds	Integer
DFG_PCT	Defensive Field Goal Percentage	Numeric

I decided to use the variables above because of the available defensive statistics in basketball, these variables have the most direct explanation of a players defensive capabilities. I made the decision to include the contested_shots because I don't believe it's fair to compare players defensive qualities unless they contest and play a similar volume of shots. I will discuss further why I decided to include certain variables such as charges.

3 Exploratory Data Analysis

3.1 Position's Impact on Receiving DPOY Votes
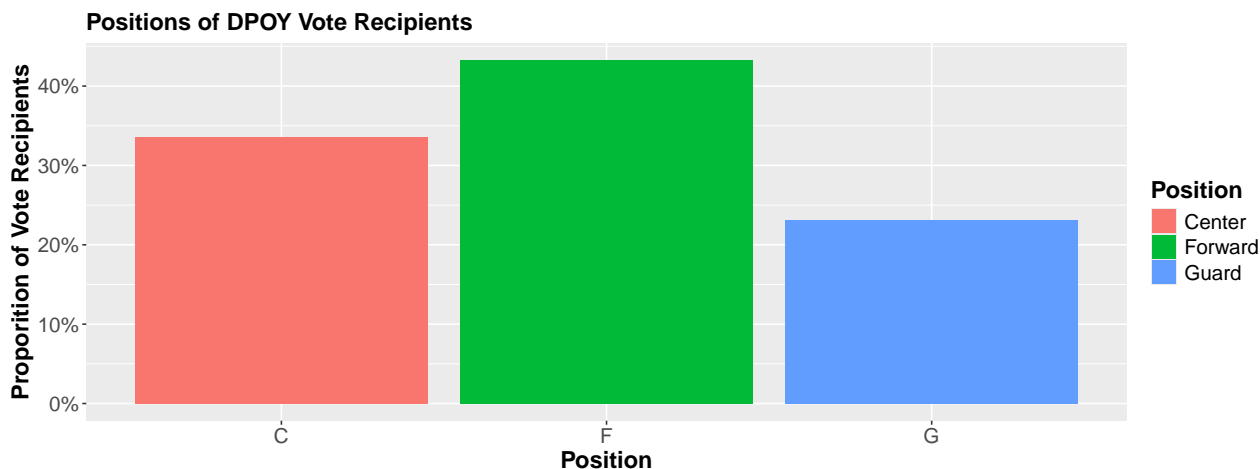
As we can see from the chart above, a majority of Defensive Player of the Year vote recipients are those whose primary position is Forward, followed by Centers, and Guards. This is reasonable as Forwards often have more involved all-around roles on both offense and defense, think Lebron James. Defensive centers can be praised for their "anchoring" of the defense, think Dikembe Mutombo or Ben Wallace.

3.2   Positional Impact on Winning Defensive Player of the Year

Positions of DPOY Winner
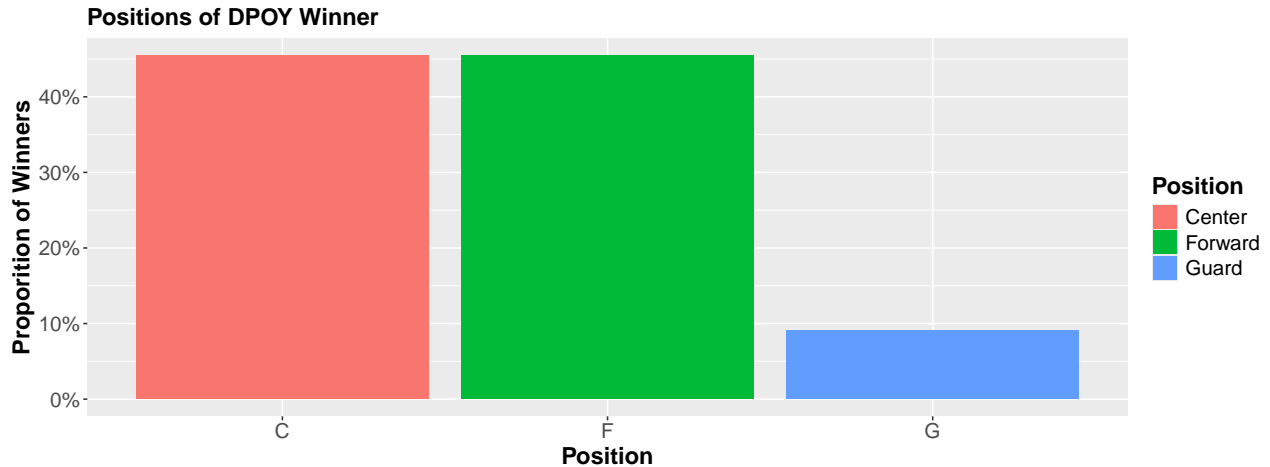
In the last 11 years, there has only been one Defensive Player of the Year who was a Guard and this was Marcus Smart in 2021-22. Perhaps this suggests that voters in recent years are biased towards Center and Forwards, and may inherently devalue the credible defending capabilities of guards. Perhaps this suggest that the voters don't value defensive performance statistics, and there is some extraneous factor they value highly.

3.3   Taking a Look at Defensive Rankings

To be clear for the following defensive graphs and metrics you will see concerning Defensive Rankings are created utilizing only the data for those who were in the top 80% of the NBA's DFGA season-by-season. I did this because I believe it's only fair for ranking purposes to compare only those players who see similar amounts of defensive volume.

3.3.1   Why does taking Charges Matter in Basketball?

One of the variables I'm suspecting some uproar about is the `charges` variable due to the fact that one might argue charges don't impact basketball that much. I'm hoping that argument doesn't arise, because to me that is nonsense. A charge results in a new offensive possession for your team strictly because of your ability to properly defend, that's meaningful in my eyes.

I am interested in investigating how Weight affects the ability to take charges. I figure another point that might be brought to light is the idea that the more you weigh the more difficult it will be to take a charge. I don't find this to necessarily be true. I think this is more circumstantial, and it depends on the quality of the defender. A good defender who is large can still draw charges, it might just take a little bit more effort but I can't comment on that.

Charge Leaders Over and Under 240 lbs
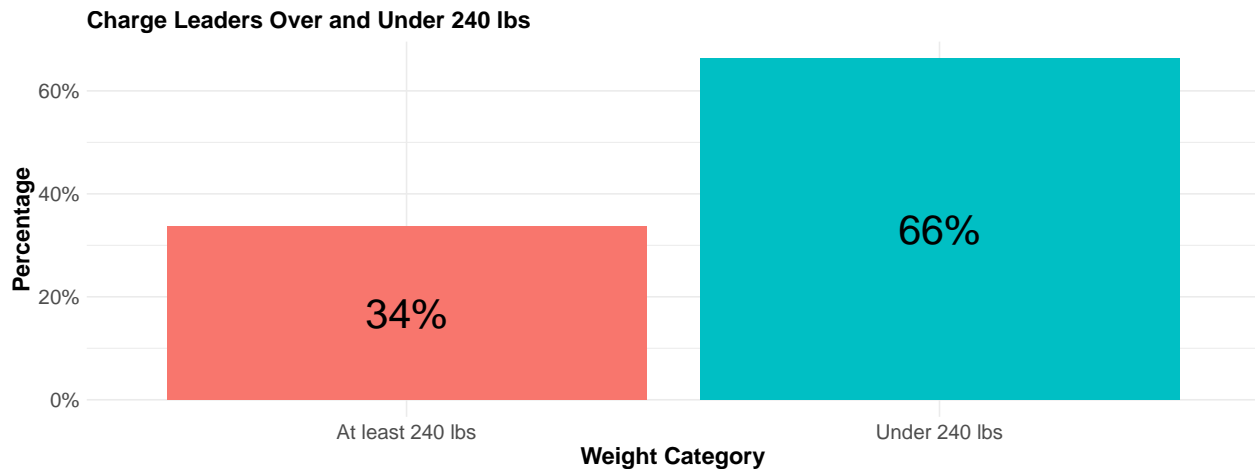
As we can see my point stands. Since 2016-17, one-third of the charges taken leaders each season in the NBA has been at least 240 lbs. For reference within our data set the mean and median weights were both right around 220 lbs. This highlights the importance of taking charges on defense as a metric to establish individual success.

3.3.2 Why include Defensive Rebounding?

Rebounding is an important aspect of NBA success when looking at the overall totals. Is it more important on offense versus when on defense? I'm not really sure, but anyways this article is about defense so we'll talk about defense. My main interests for including `DRB` as a variable are Rudy Gobert's elite defensive rebounding capabilities, and to highlight the flaws of the `DWS` statistics which accounts for `DRB` in its calculations.

Frequency of Top 10 for Defensive Rebounding
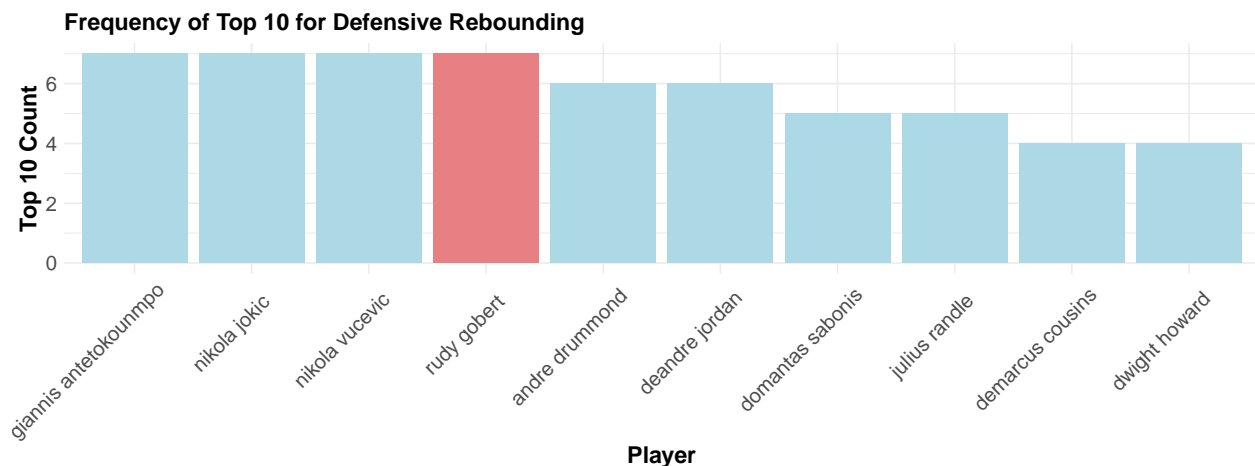
One of Gobert's best arguments for his DPOY candidacy most years is his incredible Defensive Rebounding skills. As we can see from the plot above Rudy Gobert is one of only 4 players who have the distinction of finishing top 10 in defensive rebounding 7 out of the past 11 seasons. This is incredible, and beats the number of elite rebounders such as Andre Drummond and DeAndre Jordan but this number is based off totals whereas those two players do not log large minutes in recent years.

The reason I'm even addressing Defensive Rebounds is to try to give credit to Gobert's strengths. We all know he's great with blocks, and Defensive field goal percentage but Defensive Rebounds are important for defense. Are they as important as advanced metrics give them credit? No, other stats such as Deflections, Blocks, charges etc. have a more meaningful explanation for individual defensive quality in the game of basketball.

4 How does a player get more DPOY Votes?

For this process I decided to utilize the `xgboost` library along with the `tidymodels` interface to create a Gradient Boosted Decision Tree to analyze the important variables for predicting a DPOY `vote_getter` status (either "Yes" or "No").

4.1 Model Formula

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7$$
$$+ \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} + \beta_{12} x_{12} + \beta_{13} x_{13} + \beta_{14} x_{14}$$

Where:

Table 2: Explanation of Variables in the Regression Equation

variable	variable name	variable	variable name
$\hat{y}$	Predicted variable (vote_getter)	x_7	charges
β_0	Intercept term	x_8	def_LB
x_1	Season	x_9	Deflections
x_2	BLK	x_{10}	Height
x_3	DWS	x_{11}	Weight
x_4	DBPM	x_{12}	MP
x_5	DRtg	x_{13}	w_L_pct
x_6	DFG_PCT	x_{14}	DRB

4.2 Data Cleaning process

Most of my data cleaning process was performed utilizing `Python` prior to the development of the actual report. Details for this can be found on the GitHub for this project [INSERT LINK HERE].

The most critical thing I want to discuss for the data cleaning process for model development was the data imputation. I decided to utilize the Multiple Imputation by Chained Equations process (`mice` library in R). If you would like to read more about it [INSERT MICE LINK]. I found this to be a possibly good use for this process because without imputation we still have a little more than $1,700$ observations in our dataset. This is a substantial amount of data for the process to base its decision on. The imputation used is important to keep in the back of your mind for the model, and it's interpretations as for 2013-14 through the end of 2015-16 season there are imputations contained. Prior to 2016-17, there were no "hustle" defensive stats measured which is why these values were missing.

4.3 Hyperparameter Tuning process

I utilized the `tune` interface of `tidymodels` to Hyperparameter tune the Gradient Boosted Tree in order to ensure the best performance given our training data. I utilized a 500 row random search matrix created using the `grid_latin_hypercube()` function, which searched for the ideal parameters of the `boost_tree` model with the exceptions of `sample_size` and `stop_iter` which I restricted to 1 and 5 respectively.

4.3.1 Tuned Model Parameters

Table 3: Hyperparameter Values

hyperparameter	value
mtry	6
trees	138
min_n	3
tree_depth	8
learn_rate	$7.990\,051 \times 10^{-2}$
loss_reduction	$6.486\,839 \times 10^{-3}$
sample_size	1
stop_iter	5

4.4 Model Metrics and Performance

4.4.1 10 Fold Cross-validation Performance

Table 4: Cross-Validation Metrics

.metric	.estimator	mean	std_err
accuracy	binary	0.9568	0.0023
roc_auc	binary	0.9556	0.0089
specificity	binary	0.3825	0.0387

From the cross-validation metrics we are able to see the model has a high performance accuracy of 0.9568, which is outstanding performance. The `roc_auc` being 0.9556 also indicates that we are relatively close to a near perfect predictor for the data set. However the one thing of important note is the relatively poor performance of `specificity`. This is likely due to the limited occurrences of those receiving DPOY votes, with about 10-15 people per year receiving votes.

4.5 Performance on Test Data

Table 5: Table of Test Data Performance

Metric	Value
Accuracy	0.964
AccuracyPValue	0.011
Sensitivity	0.982
Specificity	0.659
Pos Pred Value	0.980
Neg Pred Value	0.675
Balanced Accuracy	0.820

On the `test` data set we achieved a high Accuracy of 96.4% along with a high Sensitivity of 98.2% pointing to the fact that the model excels at predicting when players will not receive DPOY votes. However I want

to highlight a few lower-performing metrics. The Specificity achieved is 65.9%, which outperformed the cross-validation significantly. Additionally, the `AccuracyPValue` is less than 0.05. This indicates that at a 95% confidence level there is significant evidence to indicate that the model's Accuracy is better than a model achieved from randomly guessing.

This table suggests that the model we have created using this imputed data performs at a high level overall, despite my desire for higher Specificity.

4.6   Variable Importance Plot

Variable Importance Plot
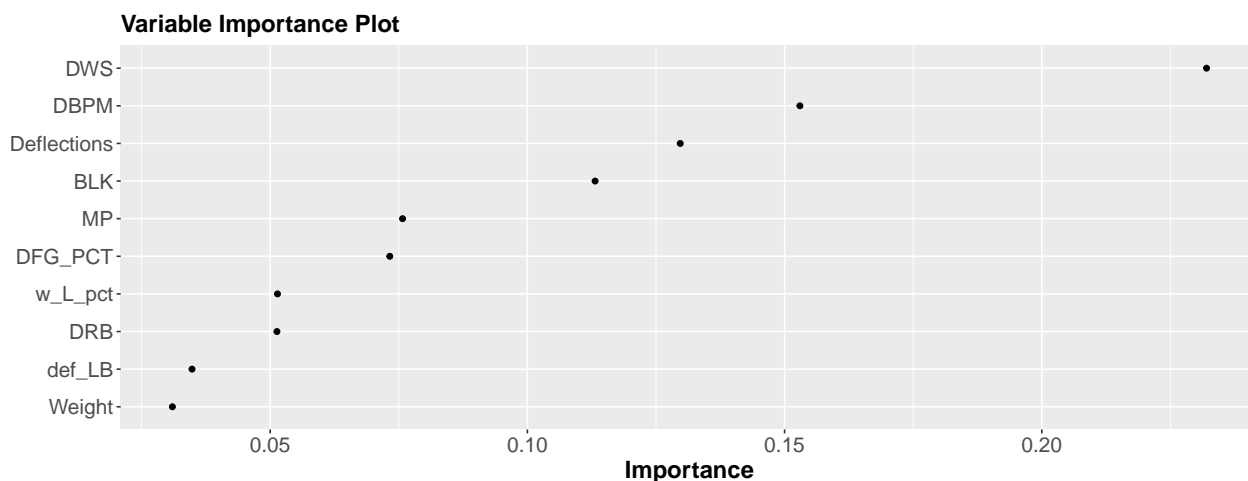
From the Variable Importance Plot of the model above we can highlight the 5 most important variables for determining who will receive a DPOY vote. These 5 variables are

- `DWS` - Defensive Win Shares
- `BLK` - Player's Total Blocks
- `MP` - Minutes Played
- `Deflections` - Pass Deflections
- `DBPM` - Defensive Box Plus-Minus

This goes along with what I would assume. To me these are 4 of the 5 best metrics we have for defining an NBA player's individual defensive impact. The only one of these variables I find problematic is `DWS` due to the method used for calculating it which I will discuss further.

4.6.1   Calculating Defensive Win Shares (DWS)

Defensive Win Shares are calculated using the formula:

$$\text{DWS} = \frac{\text{marginal defense}}{\text{marginal points per win}}$$

For further understanding of the formula I'd suggest visiting [INSERT LINK TO BASKETBALL REFERENCE] to have a better idea of marginal defense and marginal points per win specifically. However to quickly address the flaws I find is the dependence on both Defensive Rating and a team's defensive possessions per game for marginal defense. Marginal points per win also utilizes PACE for part of the equation. The problems with Pace and Defensive Rating in my mind are the fact they don't really indicate defensive performance. Their both per 100 possession stats, and all they imply is how fast your team plays and to some degree how much your defense can slow down the other team. Anyone who watches basketball knows these stats are flukey and rather questionable interpretation.

4.7 Interpretation of the Model

My goal with creating this model is to try and quantitatively understand what persuades an NBA DPOY voter to choose certain players. I believe I have mildly accomplished this goal. The model was able to achieve a moderately high Specificity while achieving an extraordinary Sensitivity on the testing data. While it could perform poorly on future data, for what we have right now I would say it's a pretty amazing start.

5 Who Deserved the DPOY Awards?

5.1 The Highest Ranked Defenders from Each Season

Table 6: Table Highest Ranked Defenders

Player	Season	average_percentile	DPOY	diff_DPOY	diff_2nd
victor wembanyama	2023-24	87.35	0	24.49	13.06
draymond green	2022-23	83.04	0	25.89	10.42
nikola jokic	2021-22	76.83	0	13.65	6.35
draymond green	2020-21	81.57	0	3.69	3.69
anthony davis	2019-20	89.80	0	13.47	13.06
marc gasol	2018-19	77.51	0	1.82	1.22
anthony davis	2017-18	86.59	0	28.86	5.54
draymond green	2016-17	91.60	1	0.00	5.60

From this table above we are able to see the top ranked defenders by `average_percentile` variable since 2013-14 season.

I want to highlight the interesting part is the estimation suggests only Draymond Green's 2016-17 Season to be the only year in which the voters correctly determined the objectively best defender through these metrics. I find it important to highlight that often in these rankings too the highest rater defender is substantially higher ranking than the person who did win DPOY. This includes 2017-18, 2022-23, 2023-24 where all the highest ranked defenders did not win DPOY yet were almost 25 percentile points better on average than the DPOY winner for that year.

5.2 Let's talk about 2023-24

Table 7: Rudy and Wemby

Player	Deflections %ile	charges %ile	BLK %ile	STL %ile	DBPM %ile	DRB %ile	DFG_PCT %ile	average %ile
victor wembanyama	88.57	97.14	100.00	80.00	97.14	85.71	62.86	87.35
rudy gobert	37.14	34.29	88.57	17.14	74.29	91.43	97.14	62.86

Now remember these `%tile` variable are their percentile rankings among high volume defensive players similar to their roles. This "high volume" refers to those players who were in the top 20% DFGA by season. Now as we can see from this Wembanyama was only worse at Defensive Rebounding, and significantly worse at Defensive field goal percentage. The defensive field goal percentage is problematic for sure, but it likely has to do with his immaturity as a defender which he'll grow out of with more experience.

I want to highlight how Wemby's defensive game is significantly more versatile. Wemby is doing everything, he's in the top 20% of these high volume defenders for deflections, charges, total blocks, total steals, defensive box plus-minus, and even defensive rebounding. Again his weakness comparatively is his defensive field goal percentage.

Gobert on the other hand is pretty much just the king of the interior. He's great at getting blocks, and preventing people from making shots but he doesn't contribute much outside of this.

I want to acknowledge that both of these guys have a tremendous effect on their respective team's defenses as Wemby is in the 97th percentile for Defensive Box Plus-Minus, and Gobert is in the 74th percentile. This is stellar for the both of them. I'll admit defensive box plus-minus is not a flawless statistic, as it neglects lineup combinations. For example, one of the reasons Wemby is in such a high percentile for DBPM is because the Spurs were terrible at defense overall this year. When Wemby was off the court they might as well not even have played defense. Once Wemby stepped onto the court his impact was felt immediately, which leads to his high defensive box plus-minus.

6   Looking Forward and Future Considerations

Overall I don't think there's much more to be said regarding how voters choose for the Defensive Player of the Year award. There's no perfect method for determining who gets votes, and often we're nowhere close to choosing them correctly (based on my rankings).

Looking forward I hope analysis like this can be utilized to perhaps build a higher performing and more sound model. Perhaps it could even be used as a suggestion for future decisions concerning the future defensive player of the year. It may not be the perfect set up and rankings, but I believe it's better than whatever methods the voters are utilizing. The so called "eye test" often fails to meet the mark, and this is shown in the rankings. It shouldn't come about that so often these players who are doing all they can on defense to better their team lose to players who are simply highly specialized in one aspect of defense.

7   Conclusion

From the data I have presented we were able to come to the following conclusions. Rudy Gobert likely should not have been Defensive Player of the Year in the 2023-24 season, it should have been awarded to Rookie of the Year Victor Wembanyama. I discovered that this is not the only time it has happened in recent memory. In fact, it is the 6th time in the last 7 seasons that the winner of the award was not the highest ranked defender. I was able to build a model for predicting those who would simply receive Defensive Player of the Year votes, which performed moderately well. I think the model's importance is greater than one simply trying to predict the Defensive Player of the Year, because it's up to the voters and from what I've shown they're hard to predict.

Hopefully further development can come along in this field of basketball analytics, as I believe defense needs the most attention from a statistical standpoint.