

An Analysis of NBA statistics

Berkiel Molinard ,Aria Kajeh, Davis Westover

Primary Contributions:

1. Berkiel Molinard did Methods and Results
2. Aria Kajeh did Abstract, Aim, Background, Data Description
3. Davis Westover did Exploratory Analysis and Description

Abstract:

Two hoops and a ball, how competitive could it get? Basketball is one of the most popular sports all around the world and takes even greater importance in the United States where it is played at the highest level, the NBA. NBA teams are constantly doing whatever they can to gain an advantage in the league and win games, whether it be experimenting with different playstyles, changing offensive lineups, organizing harder workouts. However, a team's performance will almost always be determined by the sheer talent and skills of the players on its squad. Although it is not the entire story, a lot can be said about a player's skill level by looking at their different in game statistics. Whether it be three point shooting percentage, average assists, or even the amount of miles ran per game, there is plenty of value for teams to gain, especially when they are planning their next major player trade deals. Thus, our project will be focused on individual player statistics and looking at the actual numbers in an attempt to discern what trends across different variables correlate to better players that add significant value to their teams. We are interested and will be focusing on seeing if a player's assist to turnover ratio is capable of predicting their overall plus-minus in the game(The net change in the player's team's score relative to the opposing team's score while the player is on the court) and whether turnovers affect a player's value, for if we are able to, teams will have a better statistic to determine who needs to be on the floor for them to harness that championship

Background:

The NBA has tracked statistics like points, assists, and rebounds for decades, but as we have learned how much data can help us determine value and improve optimization, more and more statistics have been tracked. In 2022, even statistics like the average speed a player runs

and total distance traveled are estimated. In fact, both the NBA website and mobile application are essentially encyclopedias of different statistics for the current seasons, allowing fans to figure out whichever record they want from any game that season. Fans, social media, critics and most importantly teams are crazy about these different statistics and the findings they offer. This is because in competitive sports, any advantage that can be gained is extremely valuable and will most definitely be used (well, as long as it isn't steroids). Teams are ready to do whatever it takes to win the championship that they praise so greatly. Teams like the Oakland A's have found incredible success relying on data over any collection of human opinions and biases. We too, are NBA statistic crazy and will be focusing on trends or correlations, specifically whether there is a trend between a player's assist to turnover ratio and their overall plus-minus and how turnovers really affect a player's value, as these statistics may very well just give us the answers on how to win the prestigious championship trophy.

Aims:

The first direction we wanted to go for this project was to of course start with cleaning the data, but while doing so we wanted to analyze the variables and brainstorm about which variables carried the most importance when it came to a team's success. Well, a known major attribute of a championship winning team is they are not selfish when it comes to the ball. They are able to effectively move the ball across the floor to overcome the defense and create open shots, thus assists play a major role in the game. More assists usually means more points, but constant pass making can be risky. As the more the ball changes hands the more chance there is for potential turnovers and turnovers, well those are a players most dreaded statistic. For it represents the amount of times in which they failed to be effective. Fans, analysts, and critics will never let go of a high turnover game for a player and will constantly remind them in the form of social media trash talk and their overall respect for the player's ability to play. But, with this increase in potential turnovers with increased assists and all of the drama that surrounds turnovers in mind, we aimed to know just how much these turnovers affected a player's value to a team.

Data Description:

Our data set consisted of game averages of every player in the NBA beginning in the 2004 season to modern day. While we do not have information from throughout all NBA history, this

does gives us a huge sample of modern basketball. Our dataset was originally taken from Kaggle and was collected by Nathan Luga, from box scores of the actual NBA games, with our observational units being the actual NBA players playing from 2004 onward and our variables being countless different NBA in game statistics such as field goal percentage, rebounds and so forth. The data can be found at the following link:

[https://www.kaggle.com/datasets/nathanluga/nba-](https://www.kaggle.com/datasets/nathanluga/nba-games?resource=download&select=games_details.csv)

[games?resource=download&select=games_details.csv](https://www.kaggle.com/datasets/nathanluga/nba-games?resource=download&select=games_details.csv) With our data set containing countless different player statistics as variables, we decided to trim them down to include the basics along with only the statistics that we felt truly represented a player's performance during the game. These values being strictly numeric made it much easier to compare and contrast different statistics. Our finalized list of variables can be seen in the table below:

Units and observations: Specific NBA Players in Specific NBA Games starting in 2004.

Variable descriptions:

Name	Variable description	Type	Units of measurement
PLAYER_NAME	Player Name	Numeric	N/A
PLAYER_ID	Player ID	Numeric	N/A
FGA	Field Goals Attempted	Numeric	Field Goals/game
FG_PCT	Field Goal Percentage	Numeric	Field Goals made/Field Goals Attempted
FG3A	3 Point Field Goals Attempted	Numeric	3Point Field goals attempted/game
FG3A_PCT	3 Point Field Goal Percentage	Numeric	3Point Field goals made/ 3Point Field goals attempted
FTA	Free Throws Attempted	Numeric	Free throws attempted/game
FT_PCT	Free Throw Percentage	Numeric	Free throws made/free throws attempted
REB	Rebounds	Numeric	Rebounds/game
AST	Assists	Numeric	Assists/game
TO	Turnovers	Numeric	Turnover/game
PTS	Points	Numeric	Player's points/game
PLUS_MINUS	The net change in the player's team's score relative to the opposing team's score while the player is on the court	Numeric	Change in points/game

Methods:

We had to tackle the problems of size and relevancy of the data set. We had two solutions for this.

First, we filtered field goal attempts to players that made more than three attempts in order to keep the list of players relevant and to minimize the risk of negative outliers. We reason that if a player has little to no attempts that they have negligible impact on the game.

Then, do not differentiate between seasons/years. This is because we do not place emphasis on the improvement/worsening of player performance across the years. As a result, we group each player's years into a singular player entry and take the mean. E.g., LeBron James' records from 2007, 2008 and 2009 are all averaged into a row that contains the means for LeBron's statistics.

By doing this, we managed to shorten the data set and keep the entries in the dataset relevant for the analysis.

It is important to note that the values used to graph and analyse do not fully represent the reality of the situation. For the sake of statistical and graphical analysis we are going to assume that these values are continuous. In the real world it is impossible to have one half of an or one and one half of an assist, or half of a point for that matter.

Results:

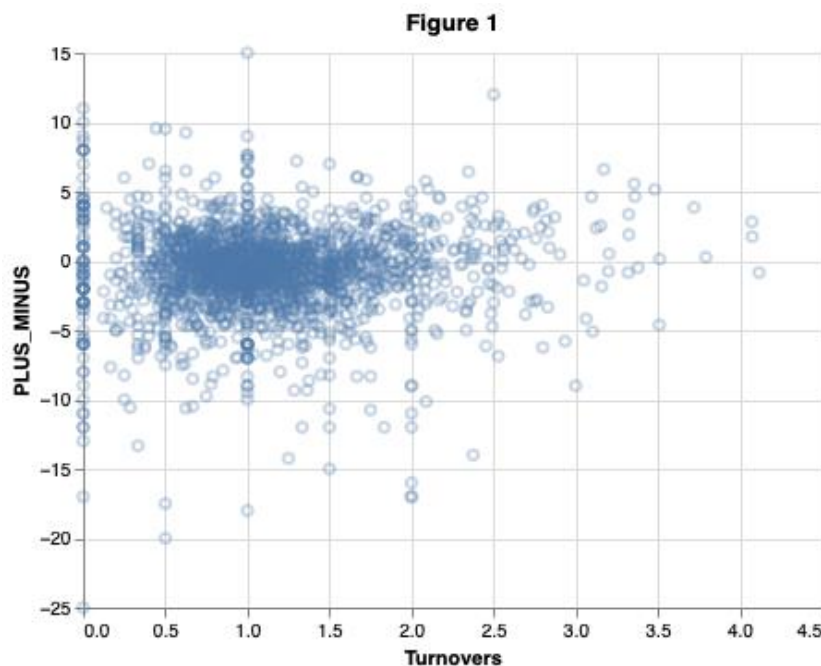


Figure 2

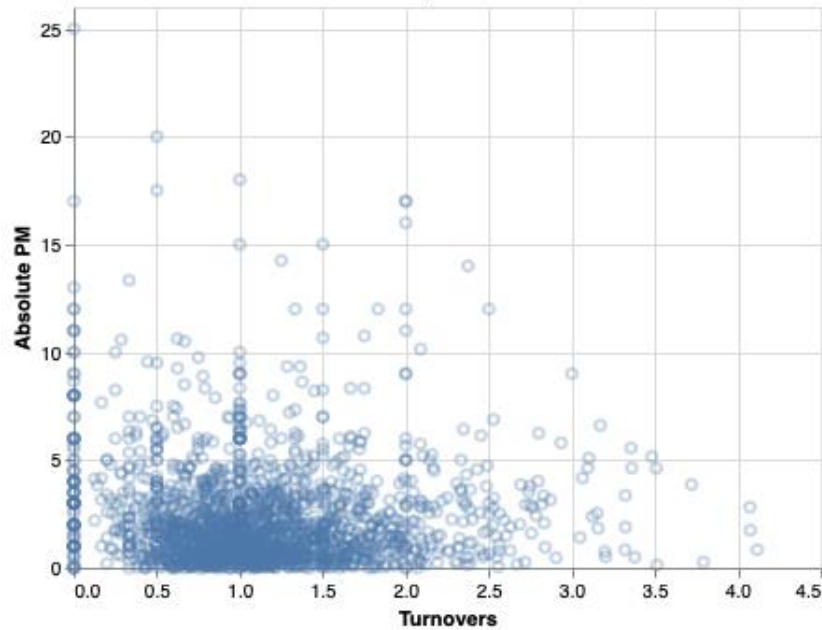
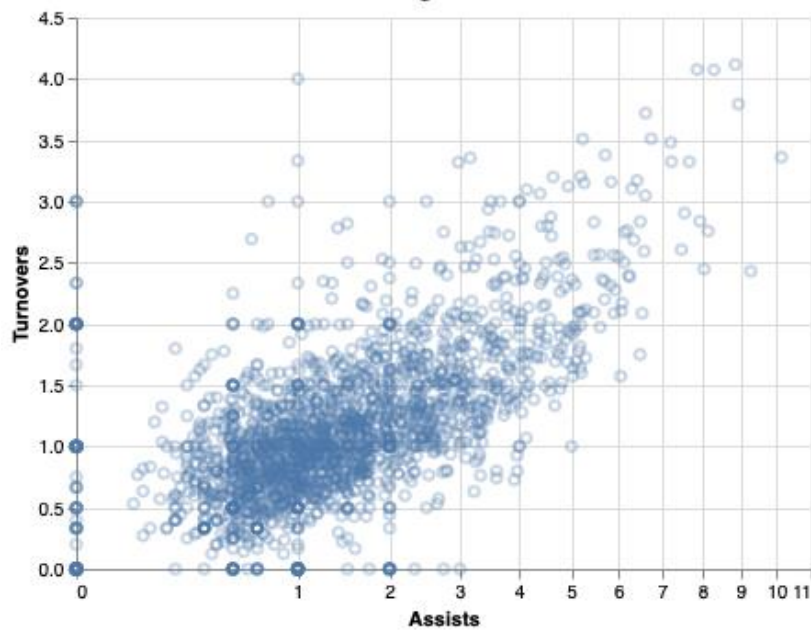


Figure 3



The graph of the change in score and turnovers are interesting. We see in Figure one that there is evidence of the values are bound between five and -five. Additionally, if we adjust the plus and minus parameter to an absolute change of points (making the negative points into positive) as we see in figure two. We see that there is no clear relationship between turnovers and the change in points.

In figure three, we observe that there seems to be a linear relationship between Turnovers and Assists. It is to be stated however that the majority of data points leads to a large concentration between the ranges of .5 to 1.5 for turnovers and .5 and two for assists.

Discussion:

We analyzed player box scores for NBA games from 2014 to 2021 (Table 1). By averaging their stats over this time span (Table 2) we were able to analyze trends and use plus-minus (which is not always reliable in single game samples) to understand how different statistics affect performance. Our research led us to notice an increase in turnovers does not correlate with a worse average plus-minus (Figure 1). Plus minus absolute values in Figure 2 show there is no significant correlation in any regard to turnovers and plus minus. To find an explanation, we provided Figure 3, which shows a clear positive correlation between assists and turnovers.

The analysis suggests that higher turnovers do not influence a player's value to his team. Most players converge to a plus-minus of 0 regardless of turnovers. In fact, there is more variation (and worse plus-minus rates) at lower turnover rates. Players who average 3 or more turnovers have a near 0 or positive plus minus. Further, by graphing the absolute value of plus-minus to turnovers, we see there is no clear relationship between the two for great or bad players. The reason this negative stat does not necessarily have a negative correlation is likely because of the relationship between turnovers and assists. We found a positive correlation between the two, meaning as players average more assists, they average more turnovers. To get high assists, which helps the team, a player has to throw more passes. Furthermore, great passers like Russell Westbrook and Magic Johnson were known for high risk high reward passes. This analysis suggests a player averaging high turnovers is more emblematic of their role on the team than them being a bad player.

No matter how detailed modern NBA stat-taking is, there still are components of the game that our data doesn't show. For example, sometimes the player that creates a shot opportunity passes to the passer and does not get the assist. Also, plus-minus is far from a perfect scale of a player's value. It is heavily impacted by a player's teammates, especially when teams

keep the same 'Starting lineups' mostly playing together, and there are more complicated statistics nba statisticians use to try to give a one number grade to players. However, it does give a general idea of value, and the consistency of trends in the analysis as well as player examples seem to validate the findings.