NBA Play Style Trends Since 1998

By: Kyle Hardy, Jinbeibei Huang, Nicholas Kraemer

Abstract

Sports analytics refers to the use of statistical analysis, data mining, and other quantitative techniques to gain insights and make informed decisions in the field of sports. This research investigates the impact of National Basketball Association game statistics such as defense efficiency, offense efficiency, average blocks per season, and average points per season on win percentages to discover which variables have the most impact for winning a game. We compiled a linear regression model to predict the win percentage with the significant variables. Through testing and remodeling, the model with average rebounds per game, average assists per game, average number of offensive rebound per game, average number of Field Goals Made (shots made) per game, average number of 3-pointers made per game, 3-point percentage (percentage of 3-pointers made), and free throw percentage (percentage of free throws made) best predicts the win percentage. After studying the individual coefficients associated with the variables, we decided to look more closely at the average number of 3-pointers made per game, average rebounds per game, and average number of offensive rebounds per game to see how differently the game is now (data received from 2022) compared to 1998.

Introduction

The National Basketball Association (NBA), since its inception in 1946, has not only been a showcase of some of the world's most talented athletes but also a reflection of the evolving dynamics of professional basketball. From changes in player styles and team strategies to adaptations in rules and regulations, the NBA has undergone significant transformation over the decades. This research focuses on the period from 1998 to the present, a particularly intriguing era marked by rapid technological advances, globalization of the sport, and shifts in player skills and tactics.

During this period, the NBA has seen a varied array of playstyles, from the dominance of towering centers and power forwards to the rise of versatile guards and forwards who excel beyond the three-point line. Coaches listening to analytics is a huge reason for this trend. They have tried to maximize their players' efficiency by as much as they can. This has led to trends of specific NBA stats. These trends/shifts will be analyzed in this report. By analyzing comprehensive game statistics, this study aims to uncover underlying trends and shifts in playstyles, offering insights into how and why the game has evolved in the manner it has. Key to this analysis is a dataset encompassing a range of variables including team seed, total games, wins and losses, win percentage, and various per-game statistics like points, rebounds, assists, steals, blocks, and efficiency metrics. This data not only provides a quantitative basis for understanding changes in the game but also helps in correlating these changes with specific events or eras in NBA history.

Main question:

How have playstyles in the NBA evolved from 1998 to the present, as evidenced by trends in key statistical categories such as offensive and defensive efficiency, average points, and average blocks per game?

Objective:

The aim of this research is not only to document these changes but also to provide a deeper understanding of their implications. By doing so, it seeks to contribute to the broader discussion on the evolution of professional basketball in the NBA, offering a historical perspective that may inform future discussions about the direction in which the sport is heading.

Methodology

Data Collection:

The primary dataset for this research was obtained from https://data.world/etocco/nba-team-stats, a website for datasets. This data set contains a wide range of NBA statistics on team performance throughout the season, dating from 1998. Average yearly data, for the time series plots, was calculated by grouping the teams by year and taking the mean of each statistic. We also used the "ggridges" package to produce joy plots for some of the most valuable stats.

Linear Regression Analysis with Win Percentages as the Dependent Variable:

The first part of the analysis process is meant to understand which statistics most significantly impact a team's success, measured in terms of win percentages. In this analysis, win percentages are treated as the dependent variable. This model aims to uncover which factors (like offensive or defensive efficiency, turnovers, etc.) have the most substantial correlation with winning games.

Time Series Analysis of Significant Predictors:

For each variable identified as a significant predictor in the above analysis, a detailed time series plot will be developed. These plots provide a visual representation of how each significant statistic has evolved over the years, offering a clearer picture of trends and shifts in playstyles.

Joy Plots for Trend Analysis

Finally, the study incorporates joy plots (also known as ridgeline plots) to examine the distribution and evolution of different statistical measures over the years. Joy plots are particularly useful in visualizing changes in distributions.

Offensive Analysis:

Linear Regression Analysis with Win Percentages as the Dependent Variable (Full Offensive Stats Model):

Wins = -625.0703 + 0.2205(Points) + 3.6075(Rebounds) +

0.7268(Assists)+1,3282(Offensive Rebounds) + 2.2580(Field Goals Attempted) - 11.3137(Field

Goals Made) + 1294.7228(Field Goal Percentage) + 4.3670(3's Made Per Game) - 1.2273 (3's

Attempted Per Game) + 79.6486 (3-Point Percentage) - 1.6851(Free Throws Made) +

1.1539(Free Throws Attempted) + 105.5347(Free Throw Percentage)

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -625.0703   233.2409  -2.680  0.00753 **
Pts             0.2205     5.6718   0.039  0.96900
Reb             3.6075     0.2926  12.329  < 2e-16 ***
Ast             0.7268     0.2676   2.716  0.00677 **
Oreb            1.3282     0.5044   2.633  0.00864 **
FGA             2.2580     2.6571   0.850  0.39573
FGM           -11.3137    10.8829  -1.040  0.29889
FG.          1294.7228   481.7862   2.687  0.00737 **
X3m_per_game    4.3670     6.4001   0.682  0.49525
X3a_per_game   -1.2273     1.0006  -1.227  0.22039
X3pt.          79.6486    55.6272   1.432  0.15263
Ftm            -1.6851     6.7705  -0.249  0.80352
Fta             1.1539     3.3719   0.342  0.73230
FT.           105.5347   110.1630   0.958  0.33839
---
     Residual standard error: 10.77 on 711 degrees of freedom
                  Multiple R-squared:  0.5207,
                  Adjusted R-squared:  0.5119,
               F-statistic: 59.41 on 13 and 711 DF,
                        p-value: < 2.2e-16
```

This model is the first model we created that included all of the variables given in the dataset. From the adjusted R-squared value of .5119, this model is not accurate since it only predicts the win percentage 51.19% of the time. The purpose of this model is to see which variables have a statistically significant coefficient. This output helps us determine which variables to keep for our reduced model.

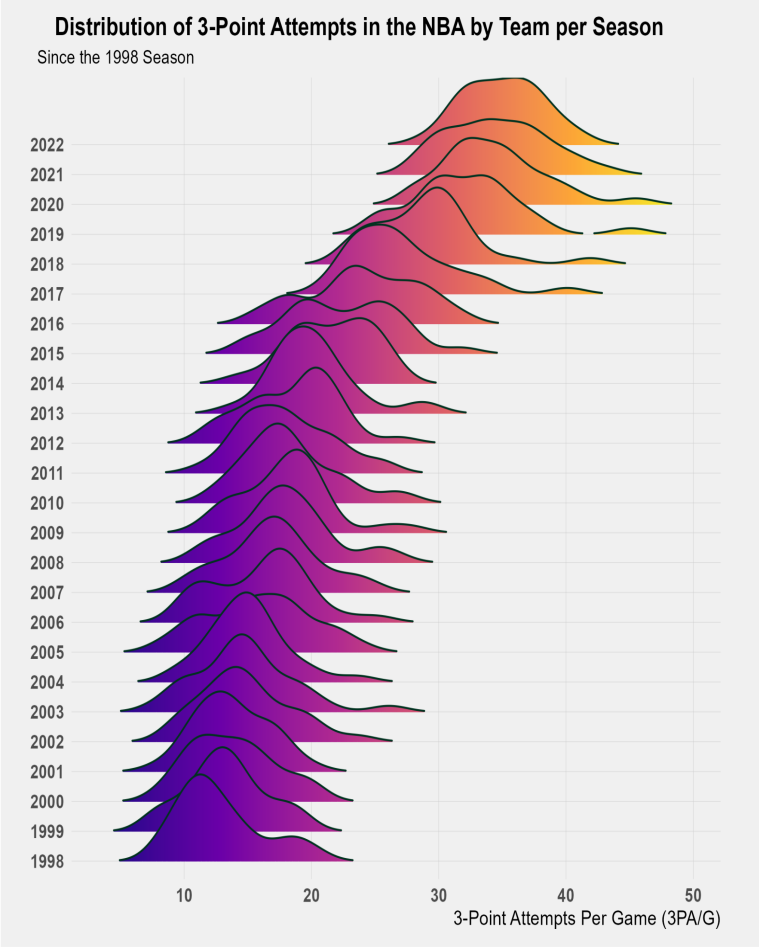
Linear Regression Analysis with Win Percentages as the Dependent Variable (Reduced Offensive Stats Model):

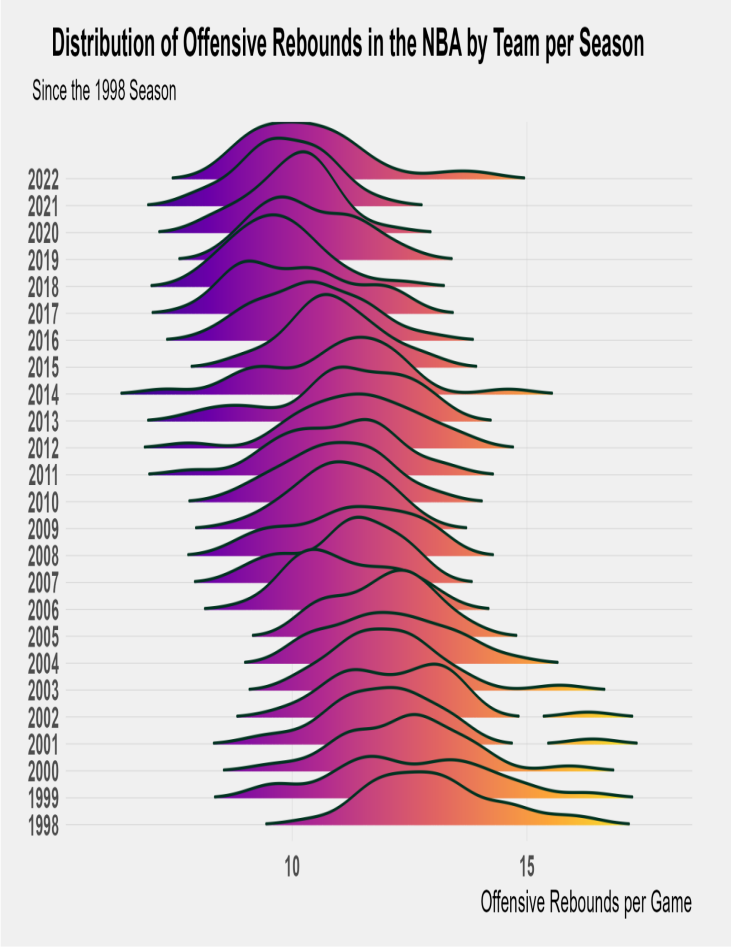
Wins = -440.1274 + 3.6039(Rebounds) + 0.7165(Assists) + 1.4968(Offensive Rebounds) - 6.0131(Field Goals Made) + 906.5749(Field Goal Percentage) +1.1954(3's Made Per Game) + 138.8127(3-Point Percentage) + 70.9188(Free Throw Percentage)

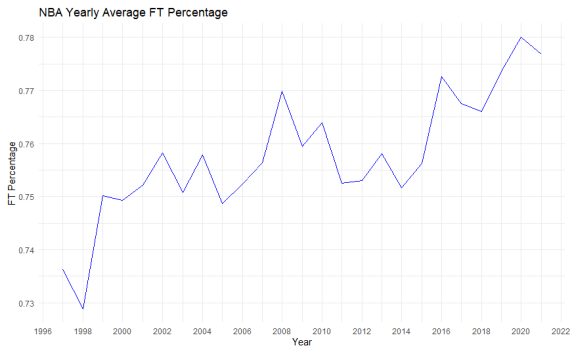
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -440.1274    21.9904 -20.015  < 2e-16 ***
Reb             3.6039     0.2886  12.490  < 2e-16 ***
Ast             0.7165     0.2646   2.708  0.00694 **
Oreb            1.4968     0.4660   3.212  0.00138 **
FGM            -6.0131     0.4199 -14.320  < 2e-16 ***
FG.           906.5749    47.9805  18.895  < 2e-16 ***
X3m_per_game    1.1954     0.2880   4.150 3.72e-05 ***
X3pt.         138.8127    26.8662   5.167 3.09e-07 ***
FT.            70.9188    15.0422   4.715 2.91e-06 ***
---
     Residual standard error: 10.75 on 716 degrees of freedom
                  Multiple R-squared:  0.5193,
                  Adjusted R-squared:  0.5139,
               F-statistic: 96.68 on 8 and 716 DF,
                       p-value: < 2.2e-16
```

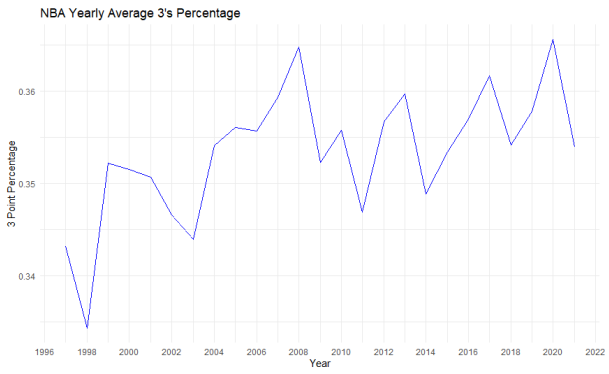
This model is created after multiple trials with only the statistically significant variables left. Compared to the other models, this one has the most number of statistically significant variables. However, the adjusted R-squared value is still only .5139, only a slight increase from the previous model.

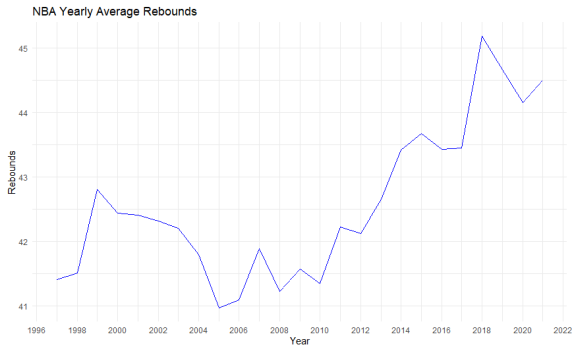
The point of these models is to show which statistics have the greatest impact on a teams' win percentage. As we can see from the results, rebounds, assists, offensive rebounds, field goals made, field goal percentage, 3 points made, 3's attempted per game, 3 point percentage, free throw percentage, defensive rebounds, steals and blocks are all significant predictors of a teams win percentage. An interesting observation is that points are not a significant predictor of win percentage which will be looked into further.

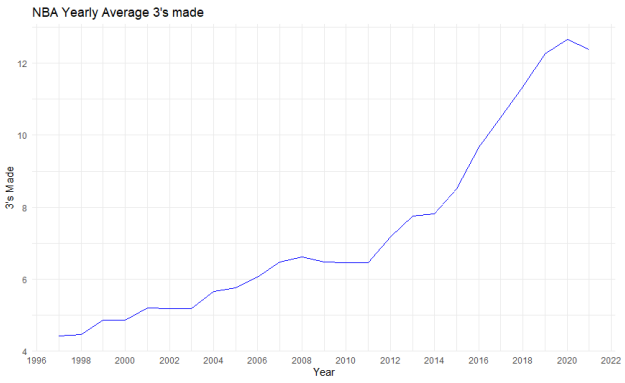

Distribution of 3-Point Attempts in the NBA by Team per Season
Since the 1998 Season
2022
2021
2020
2019
2018
2017
2016
2015
2014
2013
2012
2011
2010
2009
2008
2007
2006
2005
2004
2003
2002
2001
2000
1999
1998
10
20
30
40
50
3-Point Attempts Per Game (3PA/G)

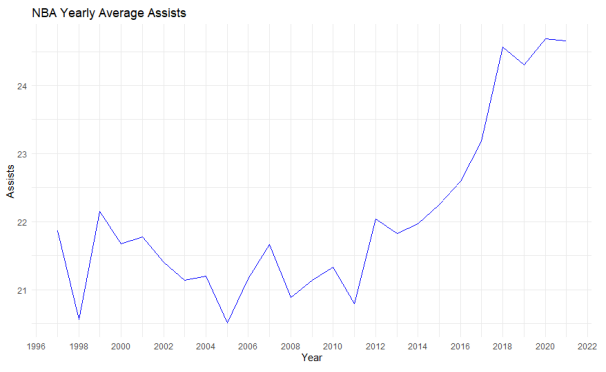

Distribution of Offensive Rebounds in the NBA by Team per Season
Since the 1998 Season
2022
2021
2020
2019
2018
2017
2016
2015
2014
2013
2012
2011
2010
2009
2008
2007
2006
2005
2004
2003
2002
2001
2000
1999
1998
10
15
Offensive Rebounds per Game

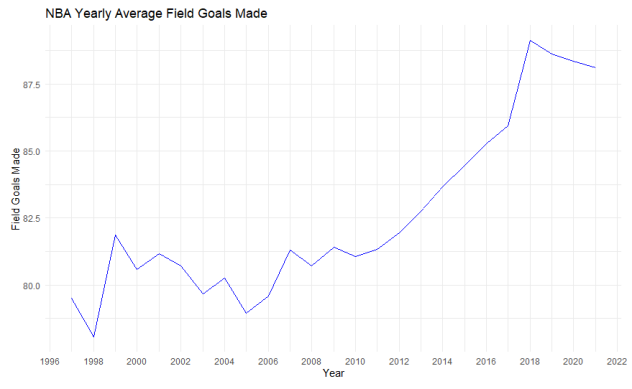

NBA Yearly Average FT Percentage
FT Percentage
0.78
0.77
0.76
0.75
0.74
0.73
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
Year

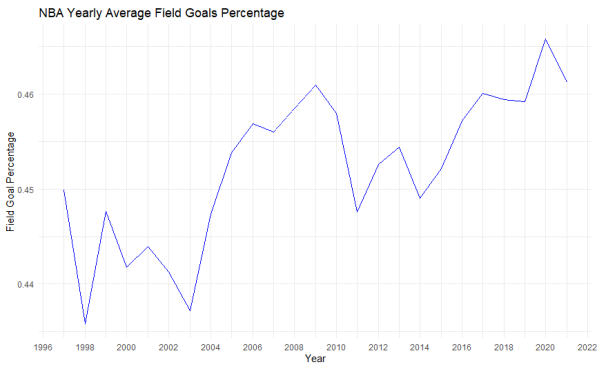

NBA Yearly Average 3's Percentage
3 Point Percentage
0.36
0.35
0.34
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
Year

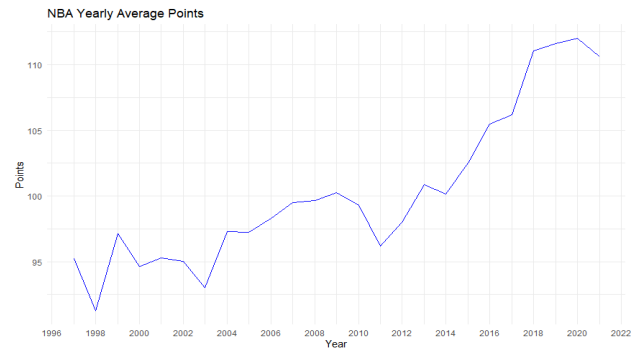

NBA Yearly Average Rebounds
Rebounds
45
44
43
42
41
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
Year


NBA Yearly Average 3's made
3's Made
12
10
8
6
4
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
Year


NBA Yearly Average Assists
Assists
24
23
22
21
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
Year


NBA Yearly Average Field Goals Made
Field Goals Made
87.5
85.0
82.5
80.0
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
Year


NBA Yearly Average Field Goals Percentage
Field Goal Percentage
0.46
0.45
0.44
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
Year


NBA Yearly Average Points
Points
110
105
100
95
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
Year

In reviewing the time series data from 1998 to 2022, a consistent upward trend is evident across various NBA statistics. Notably, the 'Points' metric, while not a significant predictor of win percentage, also shows a rising trend. This suggests that regardless of a team's success in the league, the overall scoring has increased.

The analysis of joy plots presents further intriguing trends. The first joy plot reveals a notable increase in the number of 3-point attempts since 1998, reflecting a strategic shift in gameplay. Contrary to initial expectations, the second joy plot indicates a slight decrease in rebound attempts during the same period. This could be attributed to the evolving nature of the game, where the emphasis on physical play and rebounding has diminished in favor of 3-point shooting strategies.

These trends underscore a transformation in the NBA's playing style over the past two decades, where the focus has shifted towards more dynamic and less physical strategies. Such a shift not only affects team dynamics and player roles but also has broader implications for coaching tactics and game analytics.

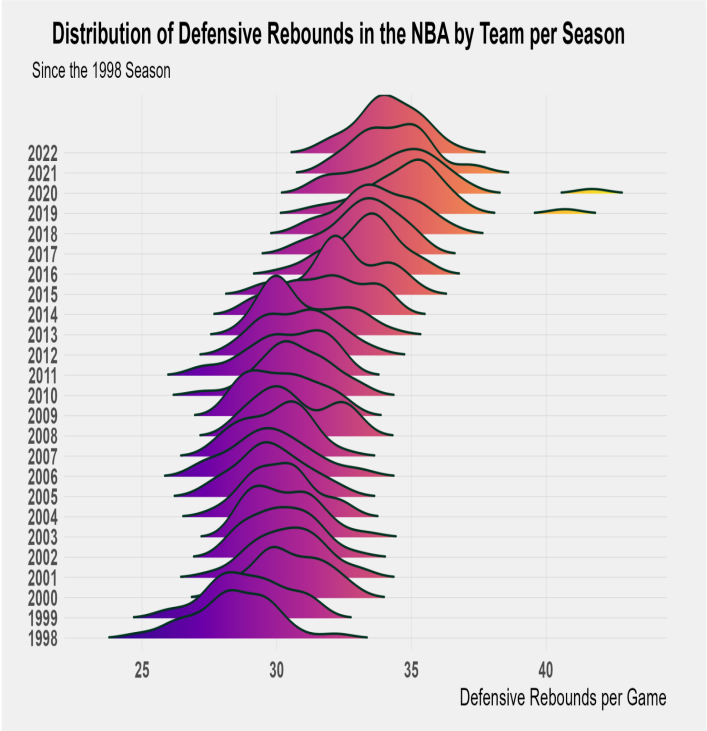
Defensive Analysis

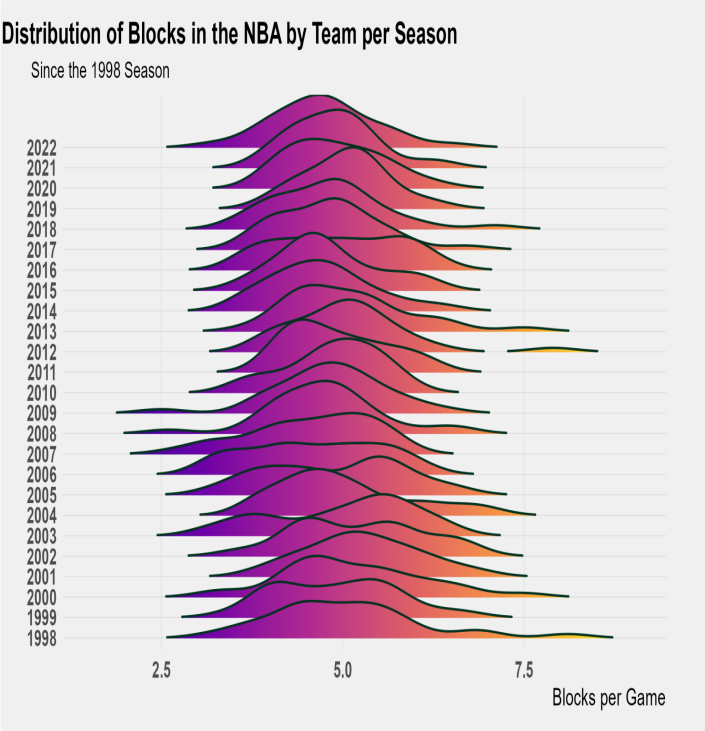
Linear Regression Analysis with Win Percentages as the Dependent Variable (Full Defensive Stats Model):

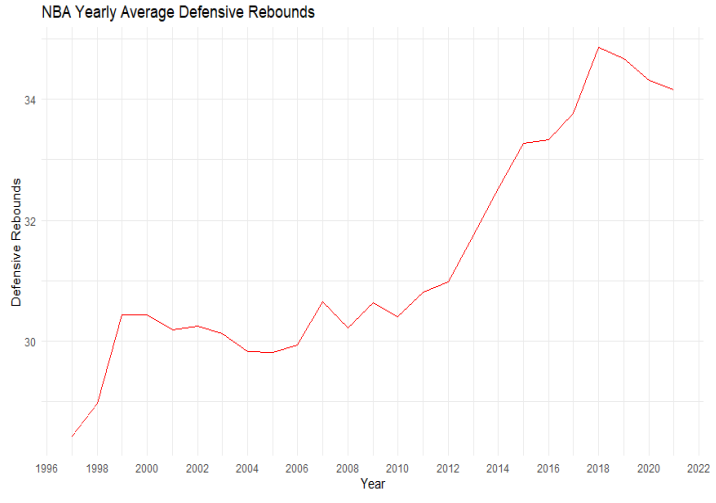
Wins = -42.2348 + 1.8826(Defensive Rebounds) + 1.3843(Steals) + 3.1406(Blocks)

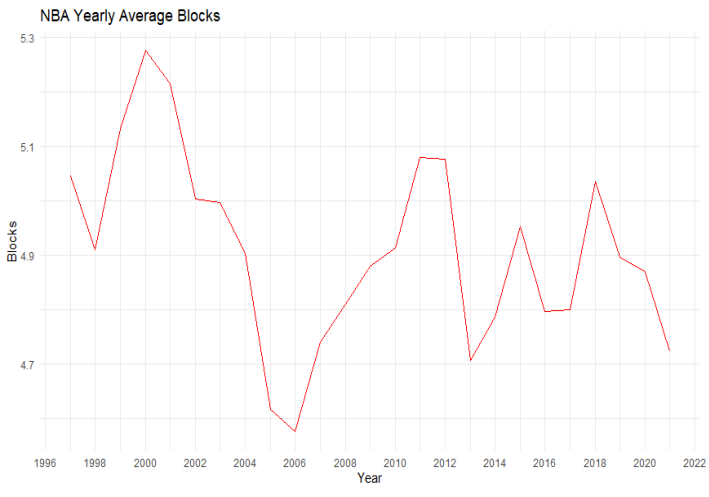
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -42.2348     9.8604  -4.283 2.09e-05 ***
Dreb          1.8826     0.2419   7.783 2.46e-14 ***
Stl           1.3843     0.6337   2.185   0.0292 *
Blk           3.1406     0.6848   4.586 5.32e-06 ***
---
     Residual standard error: 14.52 on 721 degrees of freedom
                  Multiple R-squared:  0.1163,
                  Adjusted R-squared:  0.1126,
               F-statistic: 31.62 on 3 and 721 DF,
                       p-value: < 2.2e-16
```

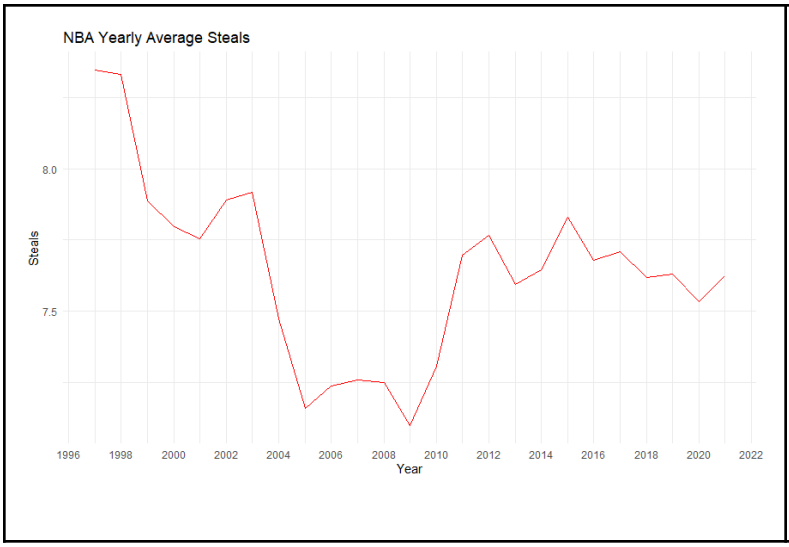
The positive coefficients for defensive rebounds (1.8826), steals (1.3843), and blocks (3.1406) suggest that increases in these defensive metrics are associated with higher win percentages. This underscores the importance of a strong defensive strategy in the game.

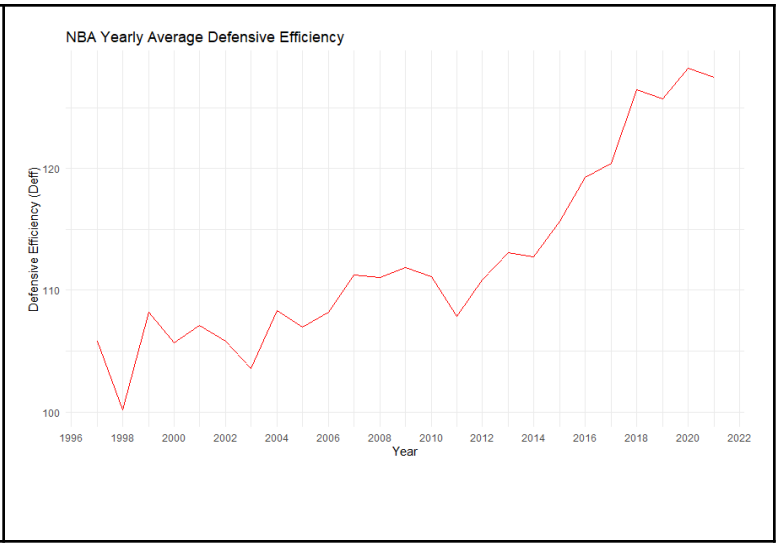

Distribution of Defensive Rebounds in the NBA by Team per Season
Since the 1998 Season
2022
2021
2020
2019
2018
2017
2016
2015
2014
2013
2012
2011
2010
2009
2008
2007
2006
2005
2004
2003
2002
2001
2000
1999
1998
25
30
35
40
Defensive Rebounds per Game


Distribution of Blocks in the NBA by Team per Season
Since the 1998 Season
2022
2021
2020
2019
2018
2017
2016
2015
2014
2013
2012
2011
2010
2009
2008
2007
2006
2005
2004
2003
2002
2001
2000
1999
1998
2.5
5.0
7.5
Blocks per Game


NBA Yearly Average Defensive Rebounds
Defensive Rebounds
34
32
30
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
Year


NBA Yearly Average Blocks
Blocks
5.3
5.1
4.9
4.7
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020
2022
Year





Upon delving into NBA defensive statistics, the time-series plots unveil a notable uptrend in defensive efficiency and average rebounds per game from 1998 to 2022. However, a contrasting trend emerges when examining the average steals and blocks per game, which have experienced a decline over the seasons. Our investigation extends to the impact of these defensive metrics on win rates, revealing a significant influence across the board. Notably, steals exhibit the least impact on win percentage and demonstrate minimal fluctuation over the years.

The joy plots, strategically applied to metrics with significant predictors of win percentage, specifically blocks and defensive rebounds, provide insightful visualizations. The analysis shows an ascending trend in defensive rebounds from 1998 to 2022, while the total number of blocks exhibits a slight decline, though not by a very large amount.

In sum, both time-series and joy plots collectively unveil a defensive evolution in the NBA since 1998, characterized by increased defensive efficiency and increased defensive rebounds, despite minimal shifts in steals and blocks.

Conclusion:

In conclusion, our comprehensive analysis of NBA statistics from 1998 to 2022 paints a vivid picture of the league's evolving dynamics and strategies. The key findings are as follows:

Despite the overall enhancement in defensive efficiency, the analysis shows a decline in the average steals and blocks per game. This suggests a shift in defensive tactics, possibly moving away from aggressive steal and block strategies.

The increase in 3-point attempts, as revealed by the first joy plot, marks a significant strategic shift. This aligns with the observed decrease in rebound attempts, pointing towards a transformation in gameplay that favors less physical, more skill-focused strategies like 3-point shooting.

All in all, the NBA has undergone a remarkable evolution since 1998, characterized by strategic adaptations in both offensive and defensive play. This evolution reflects a broader shift in basketball philosophy, embracing efficiency and skill over traditional physicality, reshaping the very fabric of the game.

Limitations:

The report mainly focuses on statistical trends within the NBA and does not account for external factors that could influence these trends. Factors such as changes in league rules, advancements in player training and sports science, or shifts in the broader cultural and economic landscape of sports might have significant impacts.

The analysis identifies correlations between various statistics and win percentages. However, it is important to note that correlation does not imply causation. The relationships observed may be influenced by other, unaccounted-for variables. In addition, the variables most likely possess co-linearity. For example, field goals made and points scored can definitely be assumed to be collinear but this was not looked into.

While quantitative data is crucial, the report might not adequately consider qualitative aspects such as team chemistry, player morale, coaching style, and other intangibles that significantly impact performance and trends in sports.

By acknowledging these limitations, the report can provide a more balanced and nuanced view of the trends in NBA basketball. It also opens avenues for future research to explore these areas more deeply.