The future of the NBA is on the Horizon*

Is it time for an expansion?

Yan Mezhiborsky

April 21, 2024

The NBA has not had an expansion in two decades. Using Data from the NBA and Basketball Reference we will investigate whether the NBA Should follow its peer the NHL and consider expanding, we will use multi-linear regression and other methods of statistical analysis to determine if the NBA should expand. Results show that the NBA has had their point per game increase a drastic amount in two decades and will continue to do so without intervention. We can conclude that the NBA should strongly consider expanding and adding a new team the the beautiful game of Basketball.

Table of contents

---

*Code and data are available at: https://github.com/Mezhi18/NBAExpansion .

1 Introduction

In this paper, we will be answering the question of whether it is time for an NBA expansion. We focused on data sets spanning from 1980 to 2024, featuring per-game statistics that serve as a foundation for our analyses. By examining variables such as points, assists, and rebounds, along with less common metrics, our research provides insights into the evolution of the NBA over nearly five decades and in some cases earlier. We also explored high-scoring games of players such as Kobe Bryant, and the impact of historical player performances, notably LeBron James, to contextualize changes in gameplay dynamics over time.

Utilizing R Core Team (2023) software, we meticulously combined and cleaned multiple data sets to ensure a clear, continuous thread for our analysis. This effort included assigning specific years to seasons based on playoff schedules, even during irregular seasons like the COVID-19 impacted 2019-2020 session. Moreover, we introduced new columns to better distinguish between different seasons and team expansions, enhancing our dataset's granularity and usefulness for predictive modeling. We will be using points per game as our estimand.

The resulting study not only revisits historical NBA statistics but also employs advanced statistical models to forecast future trends based on team expansion scenarios. Our predictive analysis seeks to understand the interplay between the number of teams and per-game scoring averages, providing a nuanced view of potential future league expansions. This research is pivotal for stakeholders interested in the strategic implications of NBA expansion and its influence on the game's competitive landscape.

This paper will follow the structure Section 3 where we will present the model, discuss our findings, and make predictions about the following few years. Second, Section 4 where we have some graphs and discuss past statistics as well as compare them to the model. Next, we have Section 5 where we have where we show some additional graphs from different data sets

and make our case for why there should be an expansion. Lastly, Section B where we have additional tables and data relevant to the paper but did not fit well in the middle.

We use R Core Team (2023), Alexander (2023), Wickham et al. (2019), Xie (2024a), Arel-Bundock (2024), Xie (2024b), Grolemund and Wickham (2024), Garnier, Ross, and Sciaini (2024), and Iannone, Cheng, and Schloerke (2024).

2 Data

The data for this research paper was gathered from Sports Reference LLC (2024), a subsection of Sports Reference LLC. Sports Reference hosts many other subsections such as Baseball, Football, Soccer, and several college sports. From Sports Reference LLC (2024), we have access to extensive data, including individual player statistics, playoffs, and direct comparisons on the website.

The first data set includes the season summary from 1980 to 2024, this past season. From the specific table where the data was gathered, the per-game stats of each team by year are listed, along with percentages and the league average for the year, which we will focus on from this data set. Some of the variables included in this data set are points, assists, personal fouls, 3-point percentage, and several more. The main focus of the data set was points, assists, turnovers, blocks, steals, and total rebounds (the option was available to divide total rebounds into offensive and defensive rebounds per game). As these summaries are on the website in individual links, they needed to be individually downloaded and combined in R Core Team (2023). While it was a tedious process, the retrieval of data was done carefully and meticulously. In regards to cleaning this data set, other than combining all data sets, the year needed to be added for each season so that they would be distinguishable from one another. As the NBA season is played through two calendar years for simplicity, the decision was made to assign the calendar year in which the playoffs of that season were played (e.g., the 1999-2000 NBA Season was saved under the year 2000). Next, we added an additional column that gave us the amount of teams there were in each season as that column did not exist in the original data set. From here, we created another data set solely focused on the league average of every NBA season. This data set was selected for many reasons, other than the fact that it contains the most important statistics such as points, assists, and rebounds as stated before, but it also includes some less-used statistics that can be used to analyze the NBA.

The next data set includes some of the highest-scoring games in NBA history. The Sports Reference LLC (2024) link from which this data was gathered had a cutoff of 59 points in a game, so all the games recorded in this data set are from times when a player scored 59 points or more. In contrast to the previous data set, as these games were played on a single date, we do have the exact year they were played, but for consistency, the new date was again based on the year in which the playoffs were played, so even if the game was played in the fall of 2022 of the 2022-2023 NBA season, the year was saved as 2023. This needed to be carefully done as the postponement of the 2019-2020 NBA season due to the global pandemic, COVID-19,

3

required a carefully chosen date for some of the games so that they would not accidentally fall in the range of the wrong season.

Lastly, we have the data set which is a summary of LeBron James' career. For this data set, very little cleaning was needed as Sports Reference LLC (2024) creates rather useful data tables. The reason this data was chosen, aside from the fact that LeBron is one of the greatest players in the history of basketball, he entered the league around the time of its expansion and is still an active NBA player, making him not only the perfect choice but the only logical one as well.

As stated before the the data retrieved from Sports Reference LLC (2024) was already rather clean, and for the sake of reproducibility and further updates on this topic and paper, it was decided to leave most of the unused stats within the clean data even if it was not used at this time.

3 Model

The goal of our modeling strategy is twofold. Firstly, I want to analyze how much other basic statistics such as assists and rebounds have on PPG as well as the effect the number of teams in the league in comparison with the year have on the same points per game. Next, we would like to predict the PPG of the next three seasons depending on the number of teams that are in the league. Considering the NBA would likely expand only one team at a time we will be focusing on the potential PPG over the next three years with 30 teams, 31 teams, and 32 teams.

Here we briefly describe the Bayesian analysis model used to investigate the multiple linear regression model of PPG as described below.

3.1 Model set-up

Define y_i as the average number of points per game scored by a team through out the NBA season. Then α is the average assists per game, ρ the average rebounds per game, β is blocks per game, ψ is steals per game and lastly, τ is turnovers per game, ι is the year, and η is the number of teams.

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \alpha + \rho_i + \beta_i + \xi_i + \tau_i + \iota_i + \eta_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\rho \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\beta \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\psi \sim \text{Normal}(0, 2.5) \tag{6}$$
$$\tau \sim \text{Normal}(0, 2.5) \tag{7}$$
$$\iota \sim \text{Normal}(0, 2.5) \tag{8}$$
$$\eta \sim \text{Normal}(0, 2.5) \tag{9}$$
$$\sigma \sim \text{Exponential}(1) \tag{10}$$
$$\tag{11}$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

3.1.1.1 Positive Relationship

We expect a positive relationship between the points per game and the statistics above: assists, rebounds, steals, and blocks. The higher these variables are the higher the points per game should be. While some seem directly correlated to points such as assists, as points need to be scored for an assist to be recorded, others such as rebounds and defensive statistics, i.e. blocks and steals may not seem to be directly correlated to points but they are as from blocks and steals often point scoring opportunity are created as well as rebounds.

3.1.1.2 Negative Relationship

While for the previous variables, we expect a positive relationship between the number of teams and turnovers we expect a negative relationship, as of course turnovers would represent a point-scoring opportunity lost, and the number of teams, the variable we will be focusing on, represent the dilution of talent per team as there would need to be a NBA expansion draft and team would lose some talent and the average team would be overall worse.

3.2 Model Prediction

Taking the variables from the previous Season we will look to predict the average points per game of the NBA over the next three years. This will include no change in teams an additional team and two additional teams added to the NBA for the 2024-2025 NBA season. Table 1 will have the predicted data for the following three years.

Table 1

Predicted Points Across Different Team Scenarios

Teams	2025	2026	2027
30	115.7962	115.8151	115.8341
31	113.1331	113.1520	113.1710
32	110.4700	110.4889	110.5079

PPG Predecition Through the 2027 Season

As we can see in Table 1 the points per game drop as the number of teams increases and the points per game increase every year as predicted. This shows us that the number of teams will affect the points scored but as players get used to the new NBA and more talent enters the league through the draft and free agency(for those un-drafted and/or entering the league at an older age) the points per game will continue to rise again, or in the case of no teams being added the points per game will just rise.

We believe this is rather accurate when considering our historical data as we have a relatively high R^2 value at $R^2 = 0.961$. Other such details can be found in Section C of the appendix.

4 Results

In the following section, we will be looking at statistics since the 1979-80 NBA season. In particular, we will be looking at points, rebounds, assists, steals, blocks, and turnovers. We will be analyzing how they have changed over the years in between expansions, particularly since the last NBA expansion in 2004.

4.1 Per Game Stats Since 1980

Since the 1979-80 NBA season, there have been four expansions. The first, was in 1980 when the Dallas Mavericks entered the league as the 23rd team. The subsequent expansion was in 1988 when the Charlotte Hornets(now New Orleans Pelicans) and the Miami Heat entered the league as the 24th and 25th teams. The following year in 1989, The Minnesota Timberwolves

and The Orlando Magic brought the total up to 27 teams. Next, in the Canadian expansion in 1995, the Toronto Raptors and the Vancouver Grizzlies(now Memphis Grizzlies) were created and there were 29 teams. Lastly, the most recent expansion the Charlotte Bobcats(now Charlotte Hornets) were inducted into the league. This means it has been two decades since the last time a team was added to the NBA and the so-called 'Modern Expansion' era concluded.
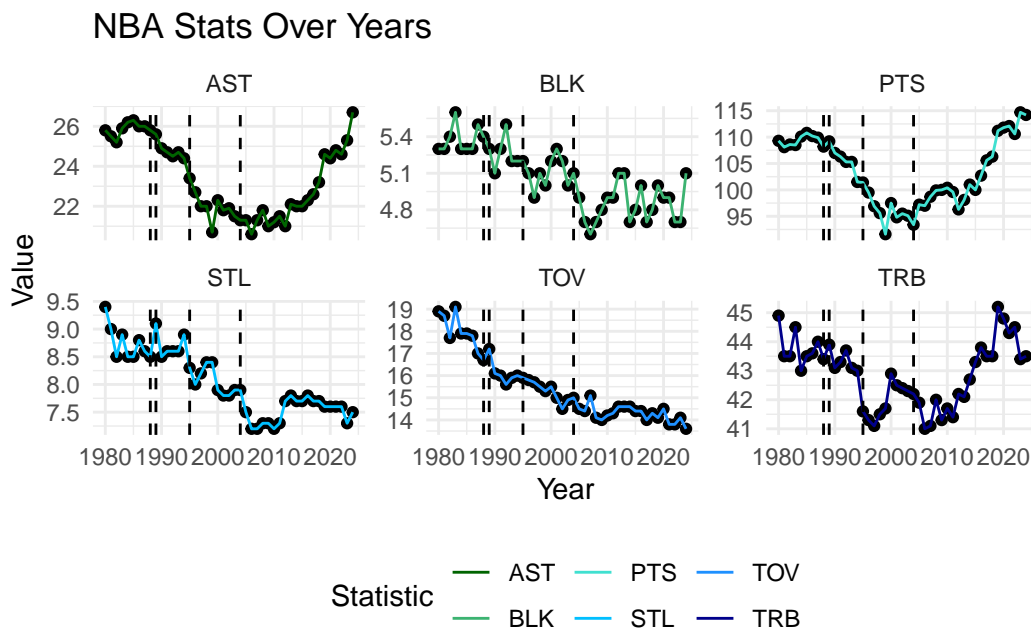

Figure 1NBA Statistics since 1980

In Figure 1 we see the graphs of the previously mentioned statistics every year from 1980. The dashed vertical lines represent every year there was an expansion. We noticed that nearly all the statistics would drop after an expansion year. Indicating that our model likely made a correct prediction. While we look at steals and turnovers, which are highly related statistics have been trending on somewhat of a decline since 1980, indicating that either open-court defense has gotten weaker over time or ball handling has gotten progressively better. In either scenario, we must recall that the NBA is a source of entertainment that needs to generate money, and while fans love offenses, better defenders would create a more competitive and therefore entertaining game. Lastly, if we look at points and assists in Figure 1 they seem to have a 'U' shape starting rather high in the 1980s taking a dive in the 1990s and early 2000s and coming back to a similar level in the modern NBA. Rebounds have a somewhat similar shape with a significant spike in the early 2000s.

4.2 Per Game Stats Since 2004

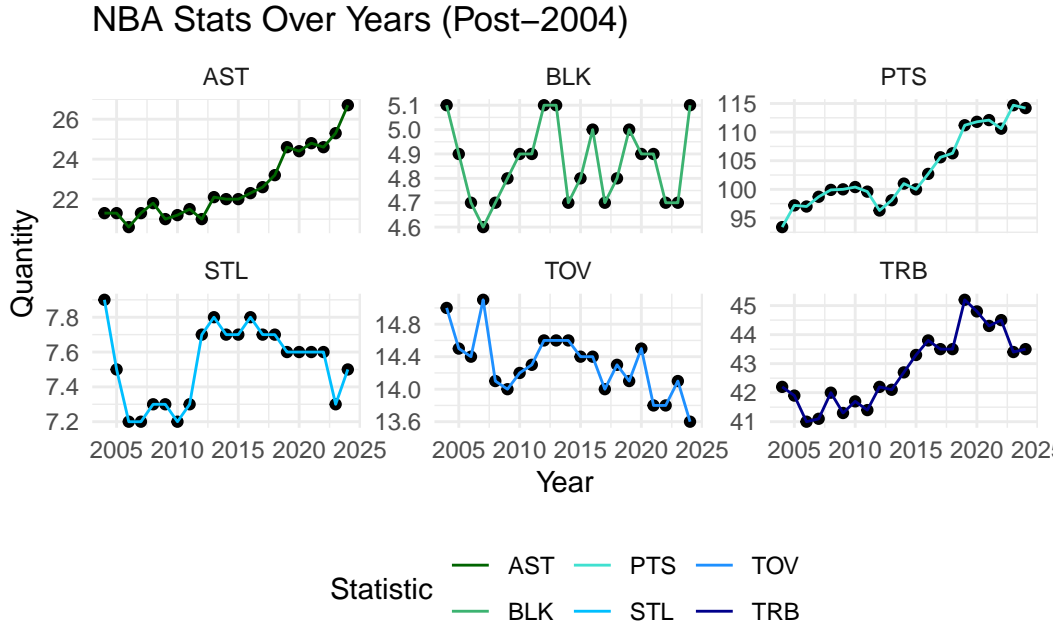In our next graph, Figure 2, we will be looking at the same graphs only after 2004, the most recent expansion.

Figure 2NBA Stats Since 2004

While blocks and steals don't seem to have a particularly strong trend in Figure 2, steals are staying rather steady between 7.2 and 7.8 steals per game while blocks seem to be oscilating between 4.6 and 5.1 blocks per game. Points, assists, and rebounds all seem to have a strong upward trend over the past 20 years with some variation from year to year. With points rising from below 100 points per game to a monstrous 115 points per game in the span of twenty years and no expansions. When we compare this to Figure 1 we can see especially how this compares to the late 1990s and the early 2000s.

From Figure 2 and our model Table 1, we have no reason to believe that any of these statistical categories show any sign of stopping in the foreseeable future.

5 Discussion

5.1 High Scoring Games

In the following three graphs, Figure 3, Figure 4, and Figure 5, we will be taking a look at the occurrences of some of the highest scoring games of all time.

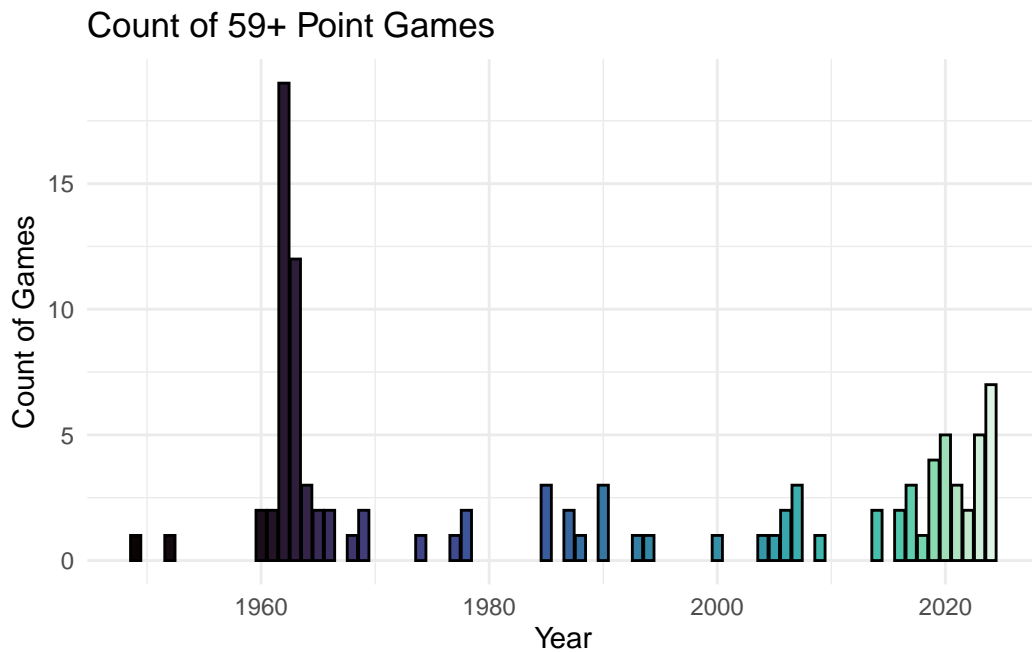### 5.1.1 High Scoring Games (All-Time)

Count of 59+ Point Games

Figure 3High Scoring games(59+ pts)

While Figure 3 is not incredibly relevant in the context of considering an NBA expansion in the 2020s it brings into context how much more common incredibly high-scoring games were in the early days of the NBA, where we can see that massive spike around the 1960 mark where one year, 1962, there was more than 59+ point game within a single season. Now when we look to the 2020s we can see the number of high-scoring games rise again.

5.1.2 High Scoring Games (Since 1980)

Count of 59+ Point Games Since 1980

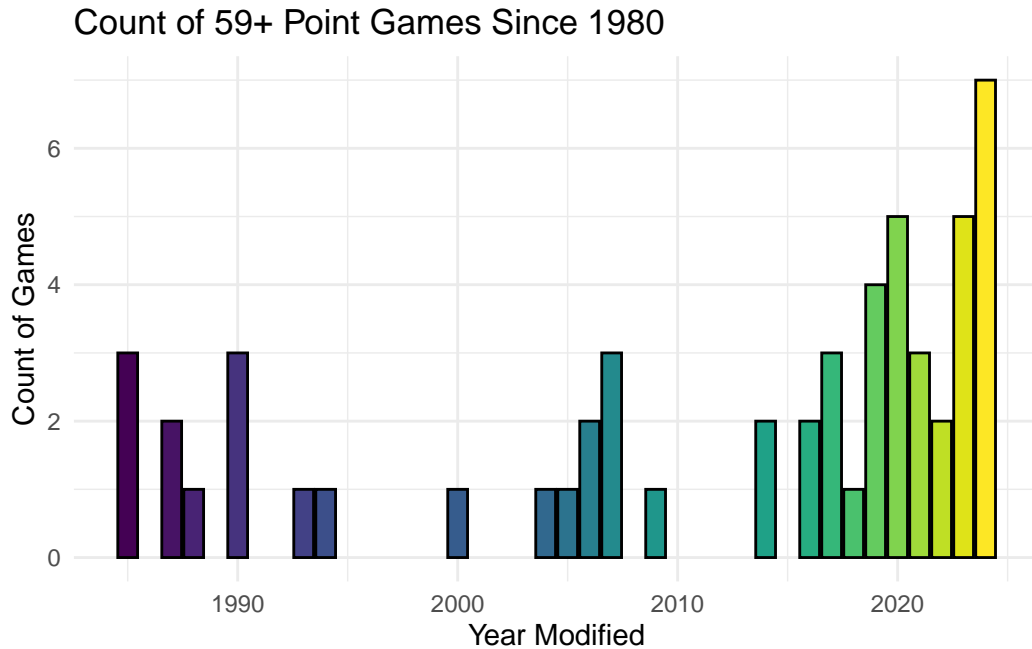
Figure 4High Scoring games(59+ pts) Since 1980

Figure 4 and Figure 3 are quite similar, but Figure 4 focused on the NBA since 1980 as our data set in Figure 1 was focused on the same years. This graph puts into perspective how many high-scoring games have been played in the past ten years when we bring in the greater context of looking at the 45 history from the beginning of the expansion era til now. High-scoring games have become so much more common in the last ten years that they have lost some but not all of their significance. From a time when only a handful in a generation type of player could produce such an incredible accomplishment, these games have become rare but not out of the ordinary.

5.1.3 Kobe's 81-Point Game

Kobe Bryant, a Hall of Fame player whose career spanned from 1996-2016 set a record for 2nd most points to ever be scored in a game and the most point scored in the game in the modern NBA.
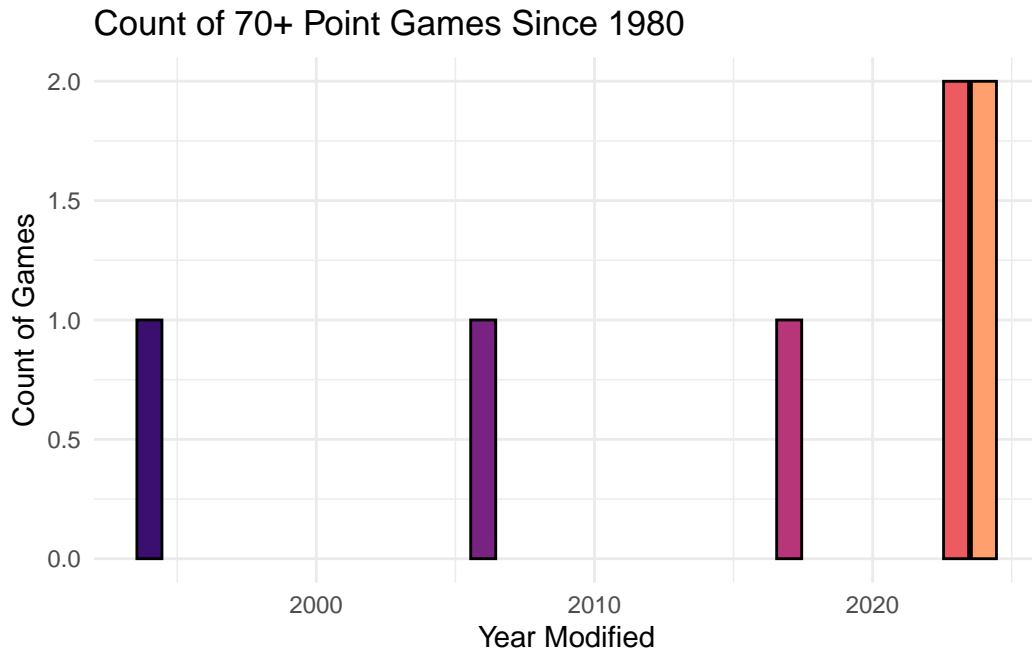
10

Count of 70+ Point Games Since 1980

Figure 5High Scoring games(70+ pts) Since 1980

As can be seen in Figure 5 since 1980 there have been a total of seven 70+ point games, five of which have been since 2010, and 4 of them have been in the past two seasons.

On January 22nd, 2006, Kobe Bryant scored his infamous 81 over the Toronto Raptors. This game was so unique there hasn't been an 80+ point game before or since, barring Wilt's 100-point game. As previously mentioned, even 70 points gamed were a once-in-a-decade occurrence until the 2020s as can be seen in Figure 5. To make an additional comment on Kobe's behalf while this was a one-of-a-kind game in his career, he did score 62 points in three quarters before refusing to play the fourth.

For some context, legendary basketball player Wilt Chamberlain scored the only ever recorded 100-point game in NBA history a record that is likely never to be broken, this was in a time the the points per game were much higher and the skill difference between the best players and the mediocre players was more than palpable. Considering these high-scoring games were so incredibly rare until recently we can conclude that the NBA today is vastly different than it used to be and this should be considered, one of the ways we might be able to reserve these games for generational players is to have an expansion and dilute the talent in the average NBA team.

5.2 LeBron James vs. The NBA

LeBron James is debatably the greatest player to ever play basketball and fortunately for us, he has stayed relatively consistent throughout his entire career and he entered the NBA as a rookie right around the time of the most recent expansion.
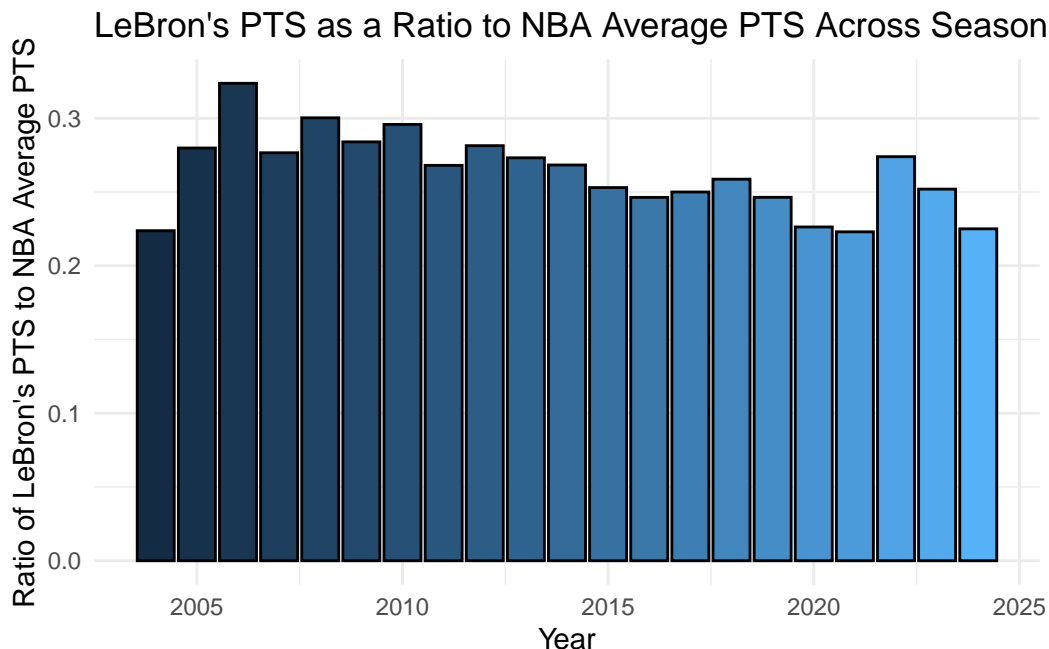

Figure 6Lebrons PPG as a Ratio of the League Average

In Figure 6 we see that when comparing his points per game to the league average points per game he has maintained a rather steady approximately 25% of the league average, especially in his older age, where we can see a mild jump in comparison to the league average. When Considering this we can see that as he maintains his points per game in comparison to the league average we can conclude that the amount of points being scored is somewhat inflated, as if there were just more points being we would be able to see a more significant decline in percentage. While there is a minor gradual decrease as he ages it is also obvious he is mainlining his scoring habits showing that not only is he scoring more as the league average increases, but he is mostly maintaining his points per game in comparison. So a 25-point-per-game season in 2024 is much less impressive than one in 2005 or even the 1990s or 1980s. In Figure 7 we can make the same conclusion, where LeBron's points per game are in red and the team average points per game is blue and the vertical dashed lines represent a change in teams.
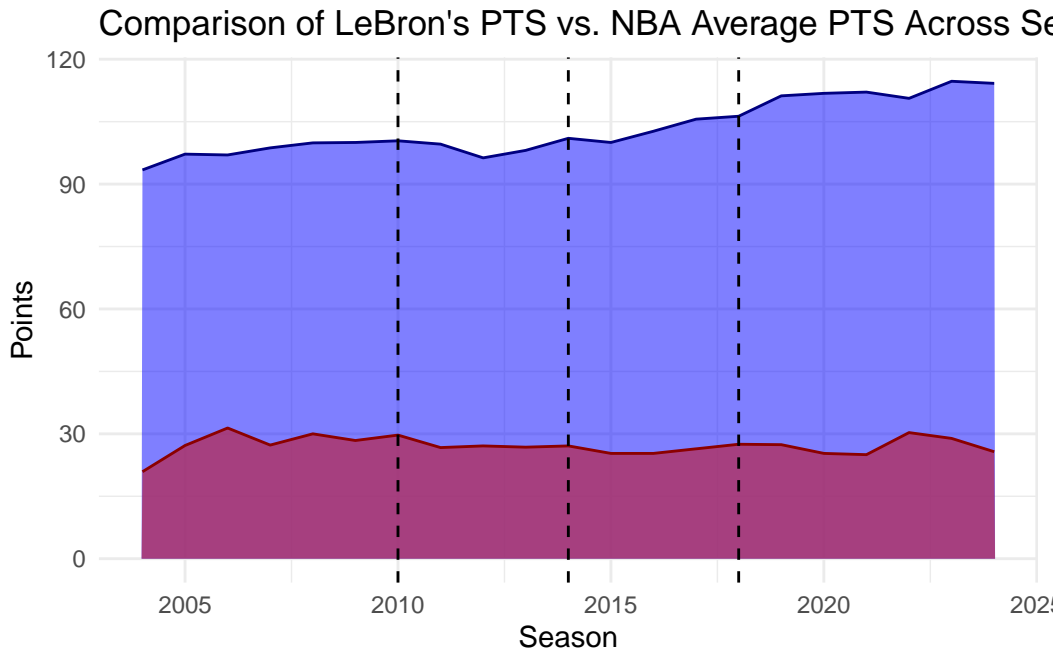
Comparison of LeBron's PTS vs. NBA Average PTS Across Se

Figure 7League Average PPG vs Lebrons PPG

5.3 NFL and NHL

The NFL and NHL both have 32 teams, while the NFL has not had a recent expansion like the NHL it was the first out of the four major sports leagues in north America to have 32 teams. As previously mentioned, the NBA is a business and they are there first to make money, there are plenty of metropolitan areas that could be mutually beneficial to the NBA with the expansion into a new city. Cities that don't have an NBA team yet could truely benefit from one. They've got fans waiting and venues ready, making them ideal spots for the league's expansion. Adding teams there could be a smart business move, giving the NBA new sources of revenue like ticket and merchandise sales. New teams in new cities would bring in not only existing fans but new fans into the basketball community, and as a "long-term investment," the more children hat love basketball will in turn bring more talent to the NBA in the future.

5.4 Weaknesses and Next steps

When writing this paper it is important to be aware of any weaknesses in our arguments and data. While all the data is true and accurate it does not necessarily tell the full story, plenty of other variables could be used including advanced statistics, total statistics, per 100 possessions, and plenty more, one is unlikely to find a shortage of basketball statistics. Within the model in mind, there may have been a more fitting regression model when considering points per game, as we can see in Figure 1 and Figure 2 not all of the statistics would be well represented with

a linear regression model. As said previously the NBA is a different game than it used to be it is not just a matter of inflation of talent and there being more good players this is true due to many factors such as an increase in population, increase in popularity in the sport especially internationally outside the USA, and rule changes that make different types of players good in different eras. We can also consider the way the game is played, Stephen curry essentially revolutionized the game in the mid-2010s when the amount of three-point shots being taken and made skyrockets, this is just one example of a player that changed the game in its long history. Lastly, when mentioning how an expansion could be economically beneficial, we fail to mention which cities would be best and make a recommendation for the best choice or two for the NBA expansion as it would require much more research and data analysis.

The next steps would be to continue monitoring the NBA in the following years and updating our data sets as more seasons pass, in addition to doing more research into more complex models that could represent the NBA points per game even more accurately than were able to achieve. Lastly, we would look into more complex and arbitrary variables that could make further interesting remarks about the NBA. Finally, further research into which cities would make the best new location to host an NBA expansion Team.

A Appendix

B Additional data details

Rank	Player	PTS	Date	Tm	Year
1	Wilt Chamberlain*	100	-2862	PHW	1962
2	Kobe Bryant*	81	13170	LAL	2006
3	Wilt Chamberlain*	78	-2946	PHW	1962
4	Luka Dončić	73	19748	DAL	2024
4	David Thompson*	73	3020	DEN	1978
4	Wilt Chamberlain*	73	-2603	SFW	1963
4	Wilt Chamberlain*	73	-2910	PHW	1962
8	Wilt Chamberlain*	72	-2616	SFW	1963
9	Damian Lillard	71	19414	POR	2023
6	Donovan Mitchell	71	19359	CLE	2023
6	David Robinson*	71	8879	SAS	1994
6	Elgin Baylor*	71	-3334	LAL	1961
13	Joel Embiid	70	19744	PHI	2024
7	Devin Booker	70	17249	PHO	2017
7	Wilt Chamberlain*	70	-2489	SFW	1963

Figure 8Top 15 Scoring Games

C Model details

Linear NBA Model

	Points Model
(Intercept)	4.82
	(137.00)
Year	0.02
	(0.07)
AST	3.26
	(0.24)
TRB	1.09
	(0.38)
STL	−3.16
	(0.93)
BLK	−6.47
	(1.53)
TOV	0.18
	(0.56)
Num_Teams	−0.26
	(0.35)
Num.Obs.	45
R2	0.961
R2 Adj.	0.953
AIC	164.7
BIC	181.0
Log.Lik.	−73.374
RMSE	1.24

References

Alexander, Rohan. 2023. *Telling Stories with Data: With Applications in r and Python.* https://tellingstorieswithdata.com/.

Arel-Bundock, Vincent. 2024. *Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready.* https://CRAN.R-project.org/package=modelsummary.

Garnier, Simon, Noam Ross, and Marco Sciaini. 2024. *Viridis: Default Color Maps from 'Matplotlib'.* https://CRAN.R-project.org/package=viridis.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Grolemund, Garrett, and Hadley Wickham. 2024. *Lubridate: Make Dealing with Dates a Little Easier.* https://CRAN.R-project.org/package=lubridate.

Iannone, Richard, Joe Cheng, and Barret Schloerke. 2024. *Gt: Easily Create Presentation-Ready Display Tables.* https://CRAN.R-project.org/package=gt.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sports Reference LLC. 2024. Basketball-Reference.com - Basketball Statistics and History. https://www.basketball-reference.com/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2024a. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://CRAN.R-project.org/package=knitr.

———. 2024b. *Tinytex: Helper Functions to Install and Maintain TeX Live, and Compile LaTeX Documents.* https://CRAN.R-project.org/package=tinytex.