Is it time for an NBA expansion?*

My subtitle if needed

Yan Mezhiborsky

April 20, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

Table of contents

---

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

1

1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

The remainder of this paper is structured as follows. Section 2….

Gebru et al. (2021)

2 Data

The data for this research paper was gathered from "Basketball Reference" (n.d.), a subsection of Sports Reference LLC. Sports Reference hosts many other subsections such as Baseball, Football, Soccer, and several college sports. From "Basketball Reference" (n.d.), 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). From here, we created another data set solely focused on the league average of every NBA season.

The next data set includes some of the highest scoring games in NBA history. The "Basketball Reference" (n.d.) 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, 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 "Basketball Reference" (n.d.) 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.

3 Model

The goal of our modelling strategy is twofold. Firstly, I want to analyze what how much other basic statistics such as assists and rebounds have on PPG as well as the affect the number of teams in the league in comparison with the year have on the same points per game. Next, I 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.

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

Predicted Points Across Different Team Scenarios

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

Figure 1: PPG Predecition Through the 2027 Season

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

We expect a positive relationship between the size of the wings and time spent aloft. In particular...
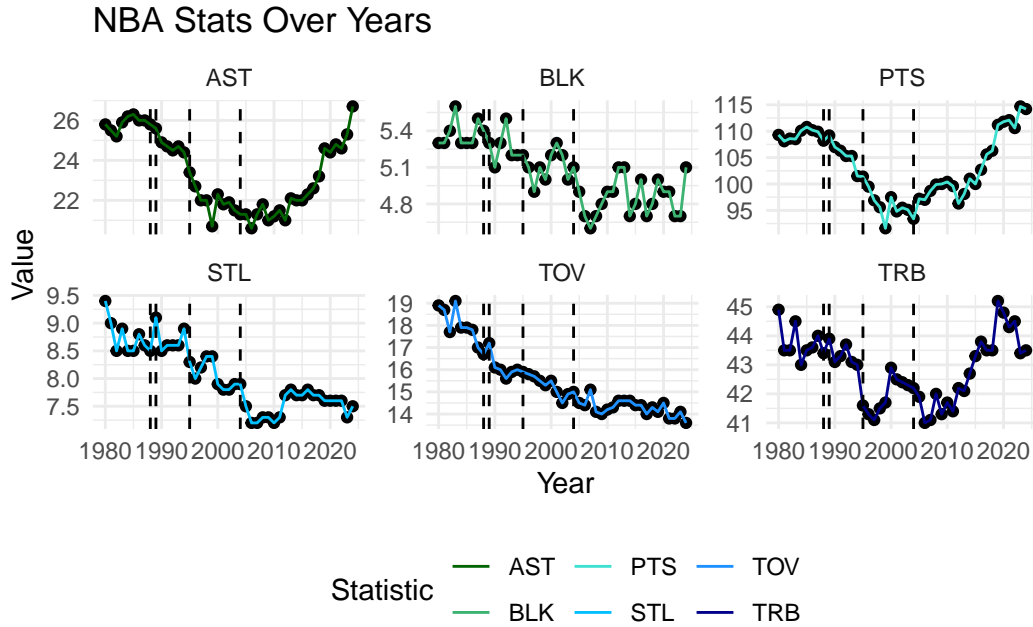
4 Results

Figure 2: NBA Statistics from 1980

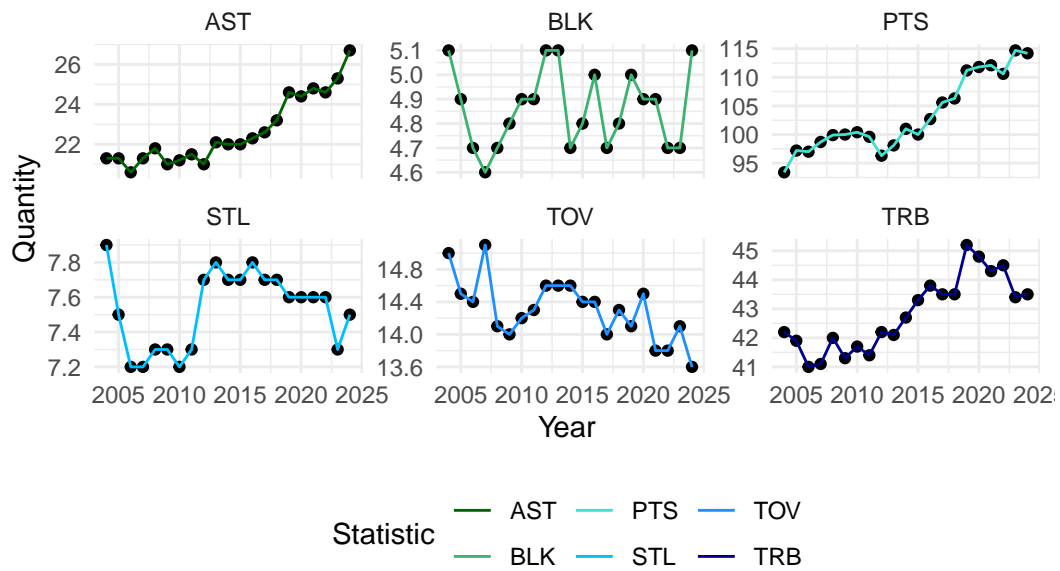## NBA Stats Over Years (Post–2004)

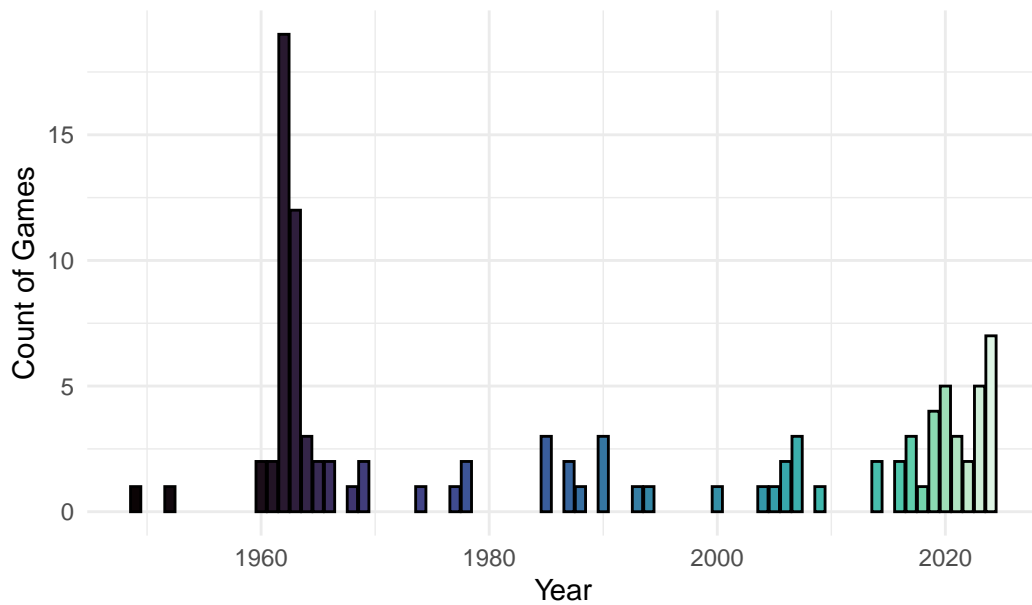Figure 3: NBA Stats Since 2004

Count of 59+ Point Games

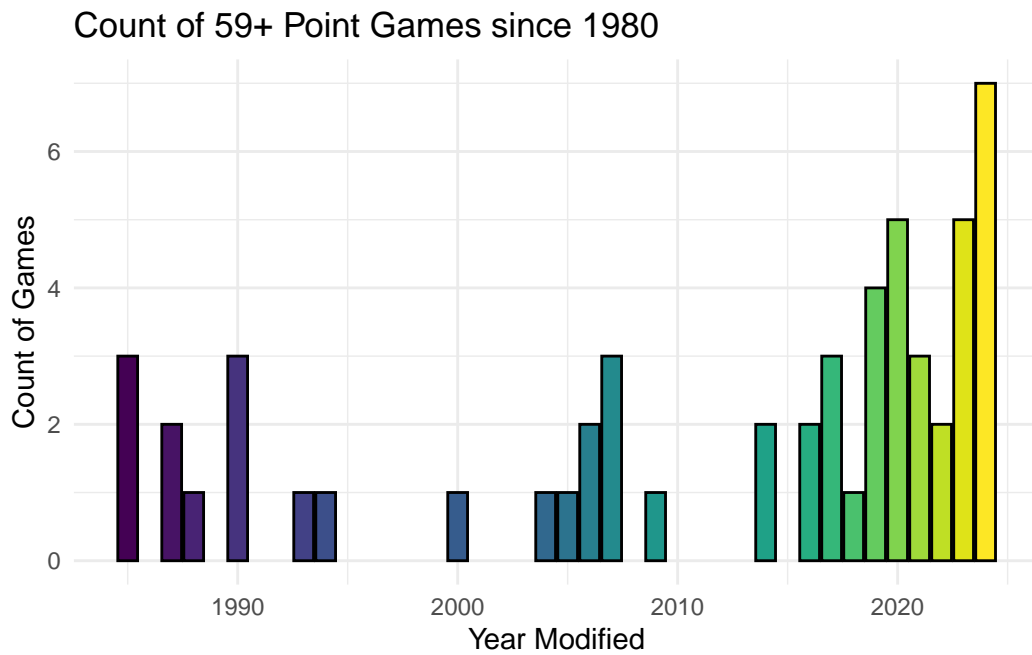Figure 4: High Scoring games(59+ pts)

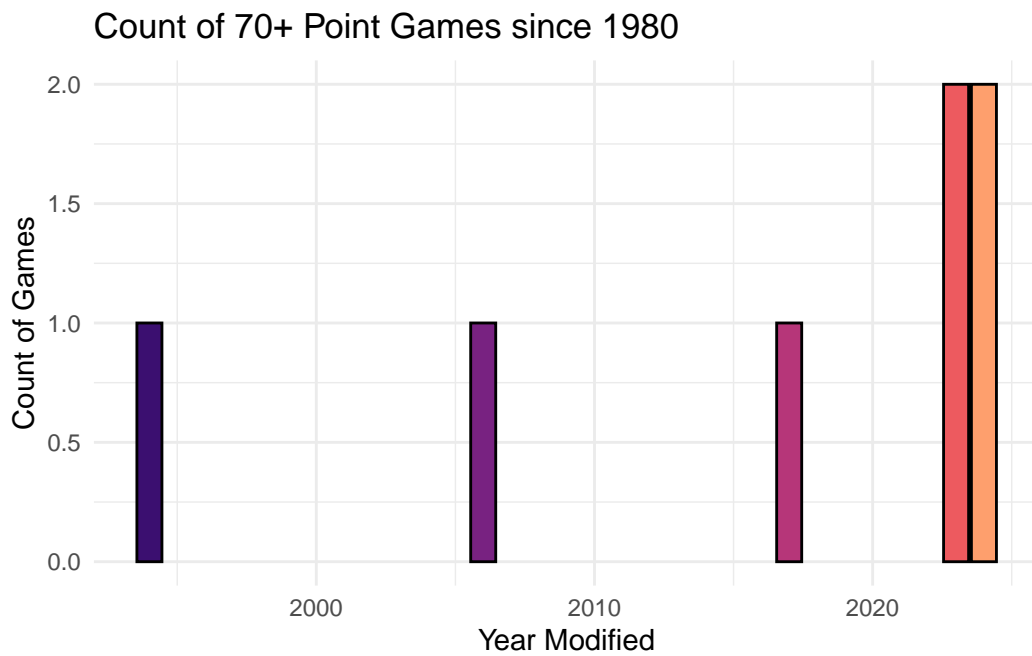Figure 5: High Scoring games(59+ pts) Since 1980

Figure 6: High Scoring games(70+ pts) Since 1980
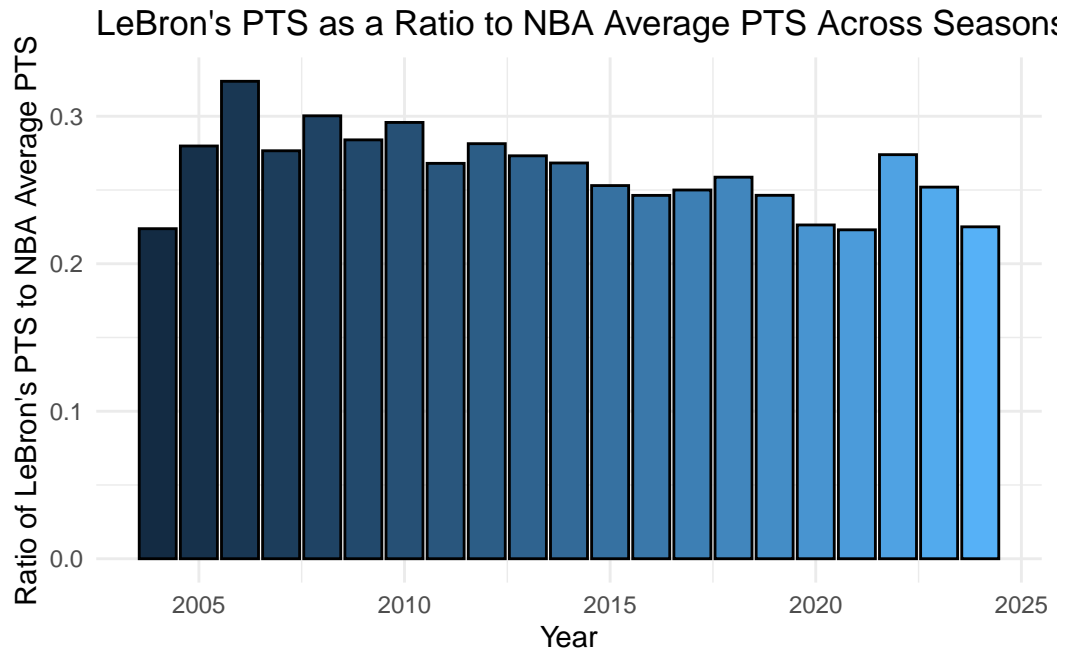
Figure 7: Lebrons PPG as a Ratio of the League Average
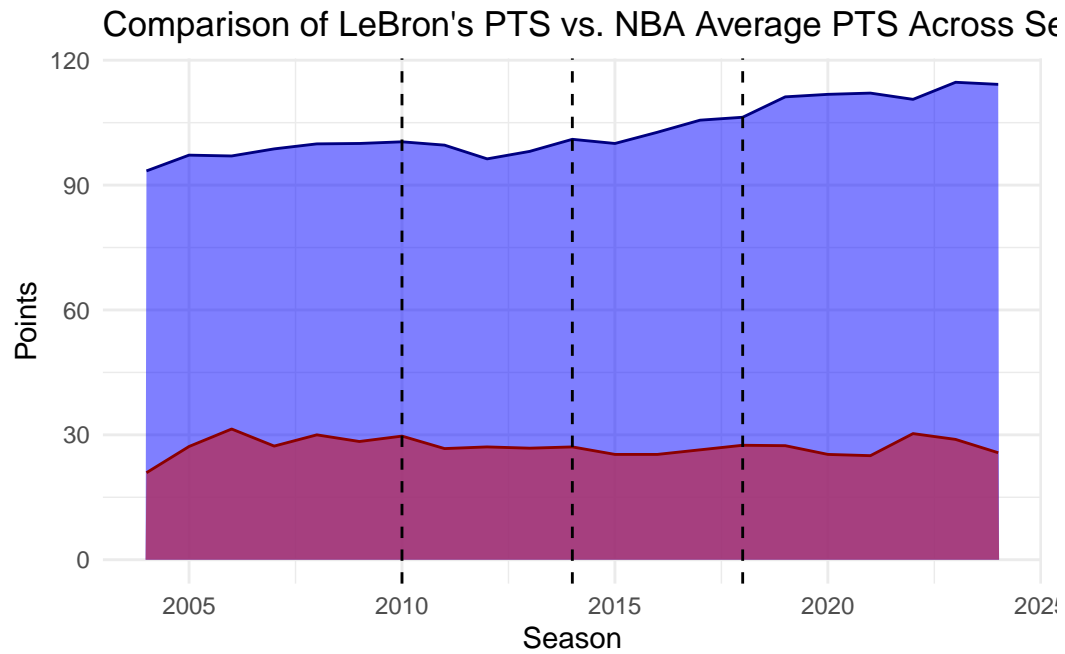
Figure 8: League Average PPG vs Lebrons PPG

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

Linear NBA Model

B.1 Posterior predictive check

B.2 Diagnostics

	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

"Basketball Reference." n.d.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92.

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

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

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.