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 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

Talk more about it.

And also planes (**?@fig-planes**). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Talk way more about it.

3 Model

The goal of our modelling strategy is twofold. Firstly,…

Here we briefly describe the Bayesian analysis model used to investigate… Background details and diagnostics are included in Appendix B.

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.

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

```r
# Fit a multiple linear regression model
nba_model <- lm(PTS ~ Year + AST + TRB + STL + BLK + TOV + Num_Teams, data = nba_data)

create_future_data <- function(additional_teams) {
  latest_data <- tail(nba_data, 1)
  adjustment_factor <- 1 - 0.03 * additional_teams  # Each new team reduces stats by 3%
  data.frame(
    Year = c(2024, 2025, 2026),
    Num_Teams = latest_data$Num_Teams + additional_teams,
    AST = rep(latest_data$AST * adjustment_factor, 3),
    TRB = rep(latest_data$TRB * adjustment_factor, 3),
    STL = rep(latest_data$STL * adjustment_factor, 3),
    BLK = rep(latest_data$BLK * adjustment_factor, 3),
    TOV = rep(latest_data$TOV * adjustment_factor, 3)
  )
}

# Create data frames for each scenario
data_no_new_teams = create_future_data(0)
data_one_new_team = create_future_data(1)
data_two_new_teams = create_future_data(2)

predict_points <- function(data, scenario_name) {
  predictions <- predict(nba_model, newdata = data)
  return(data.frame(Teams = scenario_name, Year = c(2025, 2026, 2027), Predicted_PTS = predi
}
```

```r
# Generate predictions for each scenario
predictions_no_new_teams <- predict_points(data_no_new_teams, "30 Teams")
predictions_one_new_team <- predict_points(data_one_new_team, "31 Teams")
predictions_two_new_teams <- predict_points(data_two_new_teams, "32 Teams")

# Combine predictions into a single data frame
all_predictions <- bind_rows(predictions_no_new_teams, predictions_one_new_team, predictions_

# Reshape the data frame
wide_predictions <- all_predictions %>%
  pivot_wider(names_from = Year, values_from = Predicted_PTS, values_fill = list(Predicted_P

# Print the reshaped data frame
wide_predictions
```

```
# A tibble: 3 x 4
  Teams    `2025` `2026` `2027`
  <chr>     <dbl>  <dbl>  <dbl>
1 30 Teams   116.   116.   116.
2 31 Teams   113.   113.   113.
3 32 Teams   110.   110.   110.
```

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`.

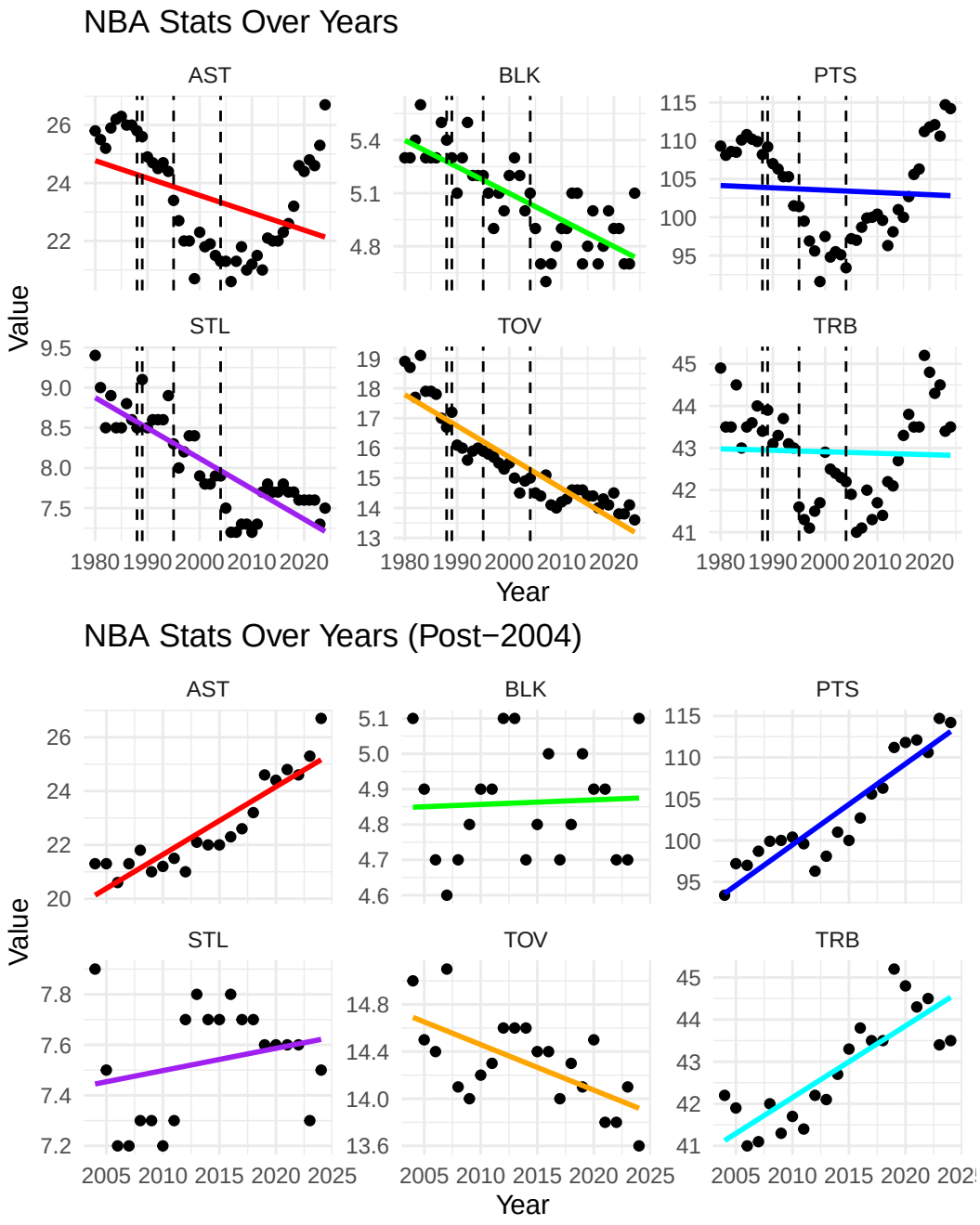
3.1.1 Model justification

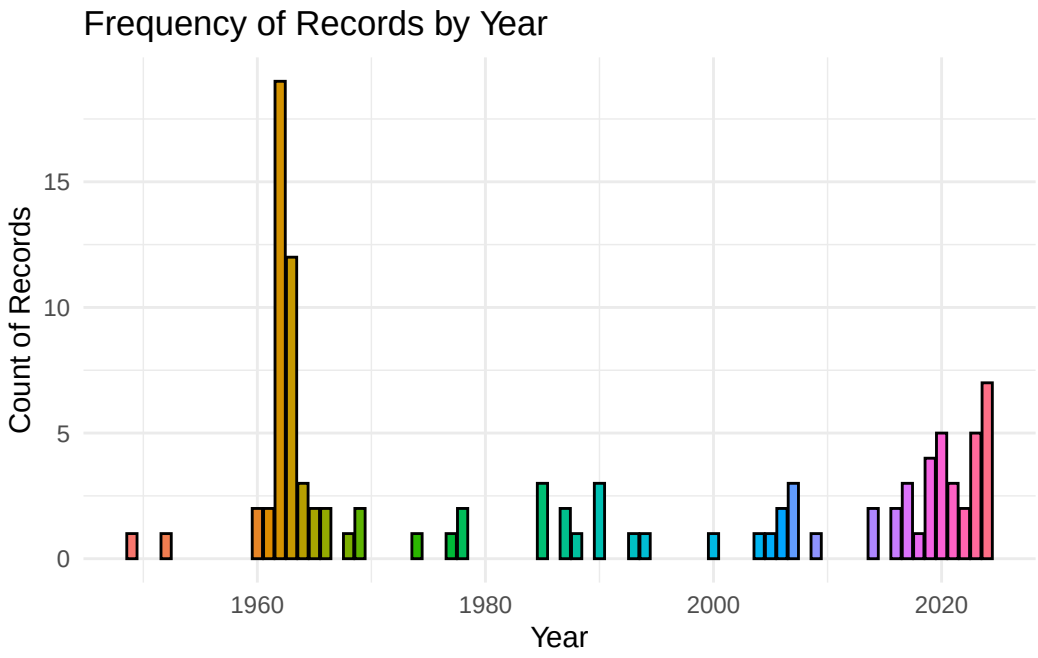
We expect a positive relationship between the size of the wings and time spent aloft. In particular...

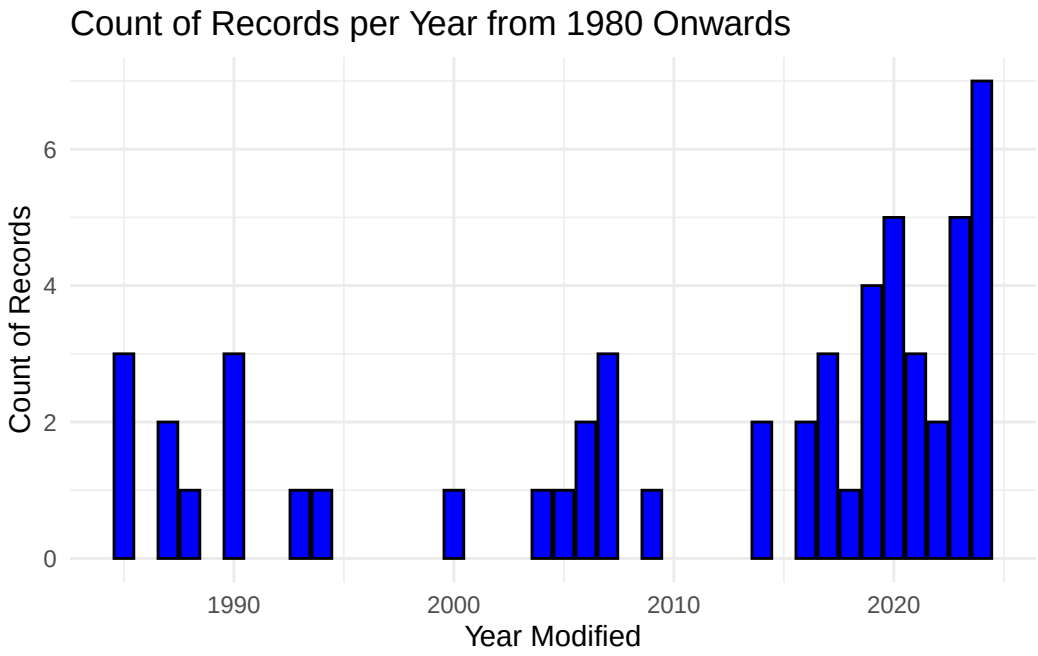
We can use maths by including latex between dollar signs, for instance θ.

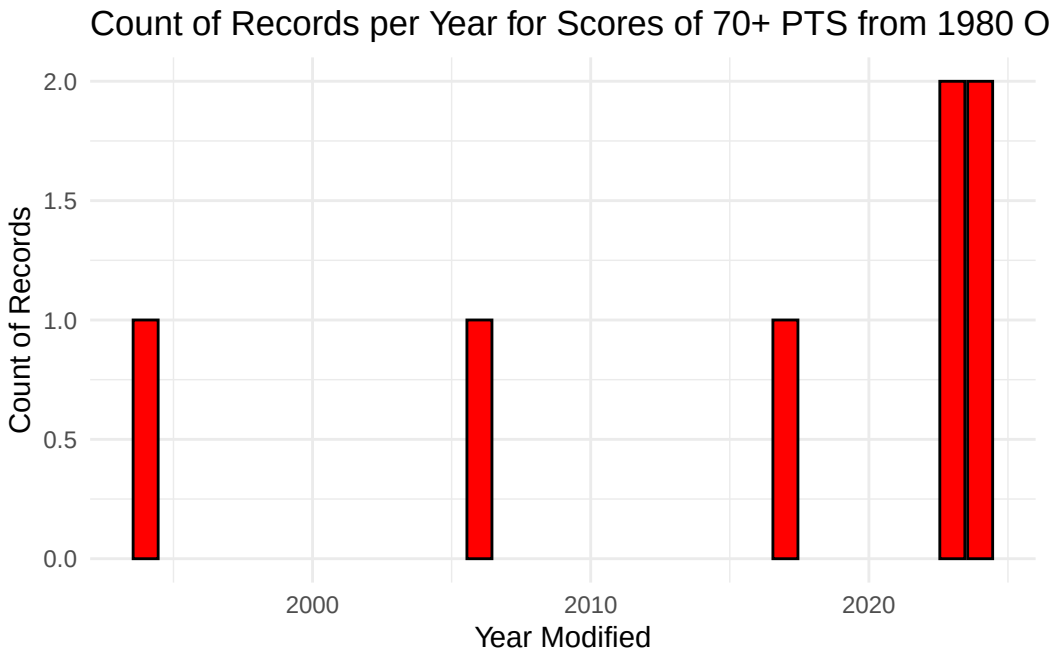
4 Results

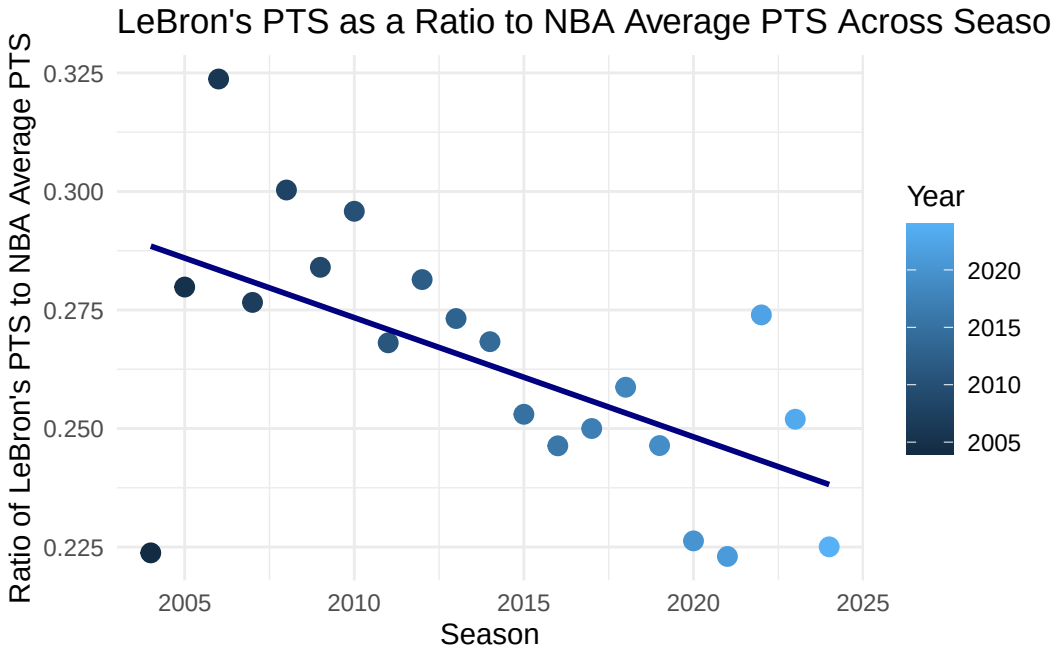
Our results are summarized in **?@tbl-modelresults**.

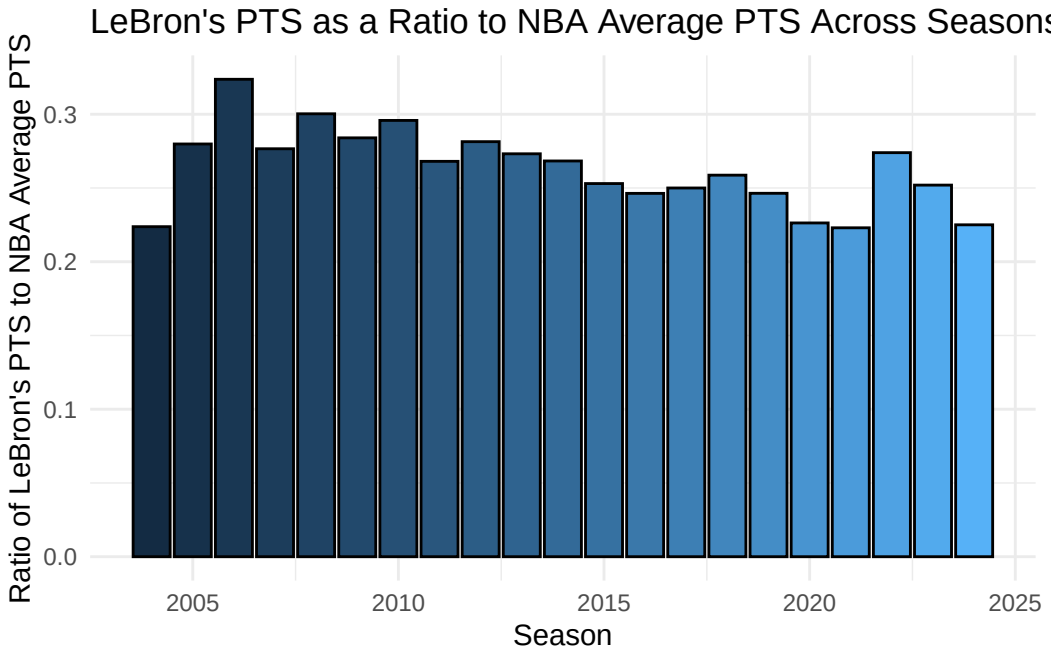

NBA Stats Over Years
AST
BLK
PTS
STL
TOV
TRB
Value
Year
NBA Stats Over Years (Post−2004)
AST
BLK
PTS
STL
TOV
TRB
Value
Year


Frequency of Records by Year
Count of Records
15
10
5
0
1960
1980
2000
2020
Year


Count of Records per Year from 1980 Onwards
Count of Records
6
4
2
0
1990
2000
2010
2020
Year Modified


Count of Records per Year for Scores of 70+ PTS from 1980 O
Count of Records
2.0
1.5
1.0
0.5
0.0
2000
2010
2020
Year Modified


LeBron's PTS as a Ratio to NBA Average PTS Across Seaso
Ratio of LeBron's PTS to NBA Average PTS
0.325
0.300
0.275
0.250
0.225
2005
2010
2015
2020
2025
Season
Year
2020
2015
2010
2005



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

```
Call:
lm(formula = PTS ~ Year + AST + TRB + STL + BLK + TOV + Num_Teams,
    data = nba_data)

Residuals:
     Min       1Q   Median       3Q      Max
-2.20354 -1.04038  0.08029  1.00630  2.85776

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.82377  136.99833   0.035 0.972101
Year          0.01899    0.07137   0.266 0.791682
AST           3.26299    0.23993  13.600 5.73e-16 ***
TRB           1.08913    0.37587   2.898 0.006283 **
STL          -3.16206    0.93494  -3.382 0.001711 **
BLK          -6.47018    1.53284  -4.221 0.000151 ***
TOV           0.17886    0.55620   0.322 0.749587
Num_Teams    -0.25656    0.34706  -0.739 0.464413
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.363 on 37 degrees of freedom
Multiple R-squared:  0.9607,    Adjusted R-squared:  0.9533
F-statistic: 129.3 on 7 and 37 DF,  p-value: < 2.2e-16
```

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

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.