Is it time for an NBA expansion?*

My subtitle if needed

Yan Mezhiborsky

April 19, 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….

2 Data

Some of our data is of penguins (**?@fig-bills**), from Horst, Hill, and Gorman (2020).

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.

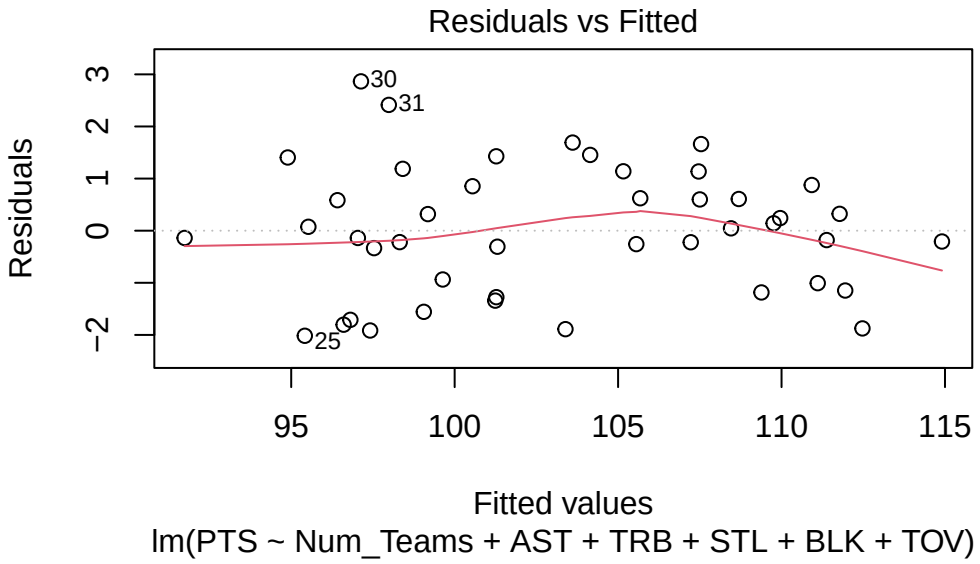
3.1 Model set-up

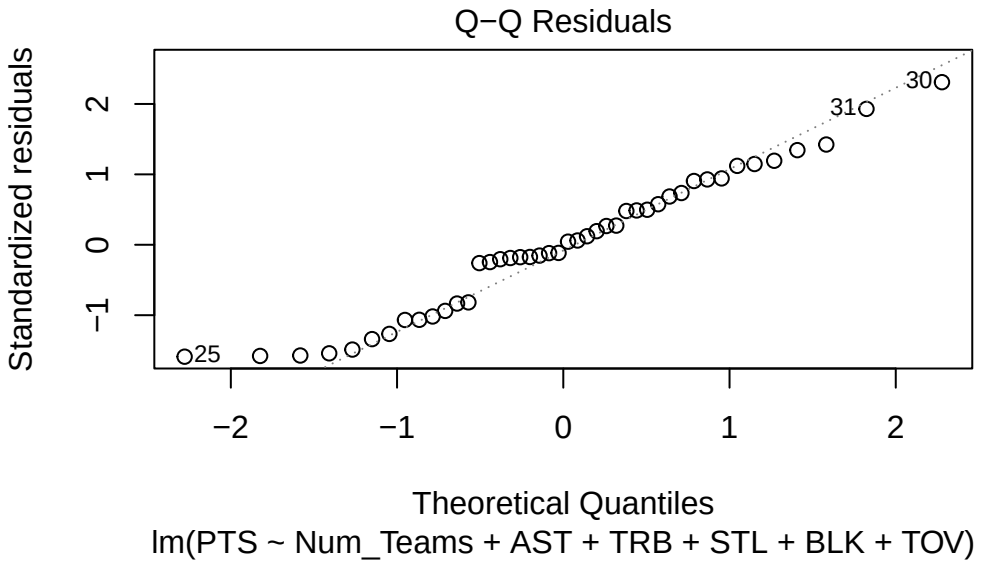
Define y_i as the average number of points per game scored by a team through out the NBA season. Then α is the average assists per game, ρ the average rebounds per game, β is blocks per game, ψ is steals per game and lastly, τ is turnovers per game.

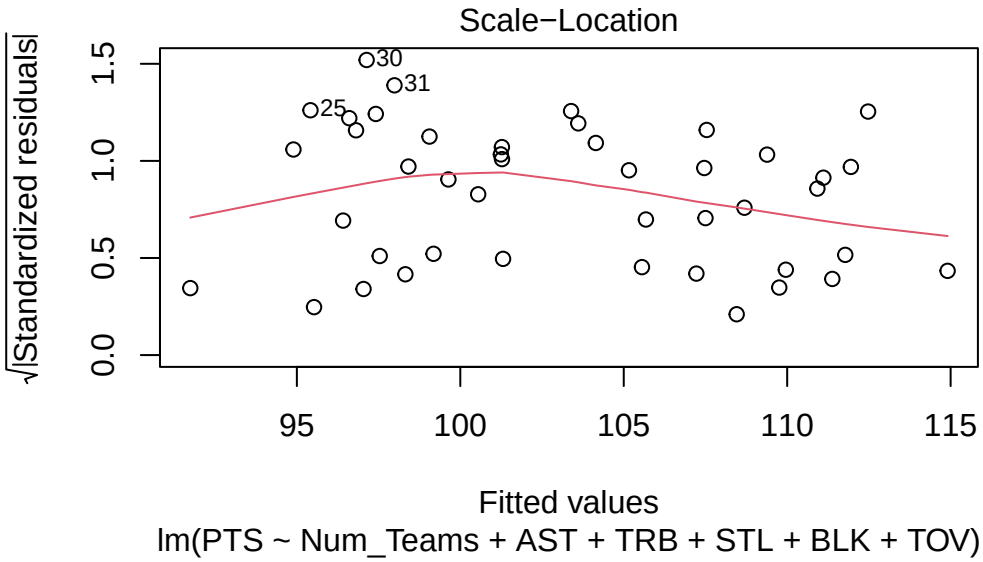
$$
\begin{aligned}
y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (1)\\
\mu_i &= \alpha + \rho_i + \beta_i + \xi_i + \tau_i && (2)\\
\alpha &\sim \text{Normal}(0, 2.5) && (3)\\
\rho &\sim \text{Normal}(0, 2.5) && (4)\\
\beta &\sim \text{Normal}(0, 2.5) && (5)\\
\psi &\sim \text{Normal}(0, 2.5)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

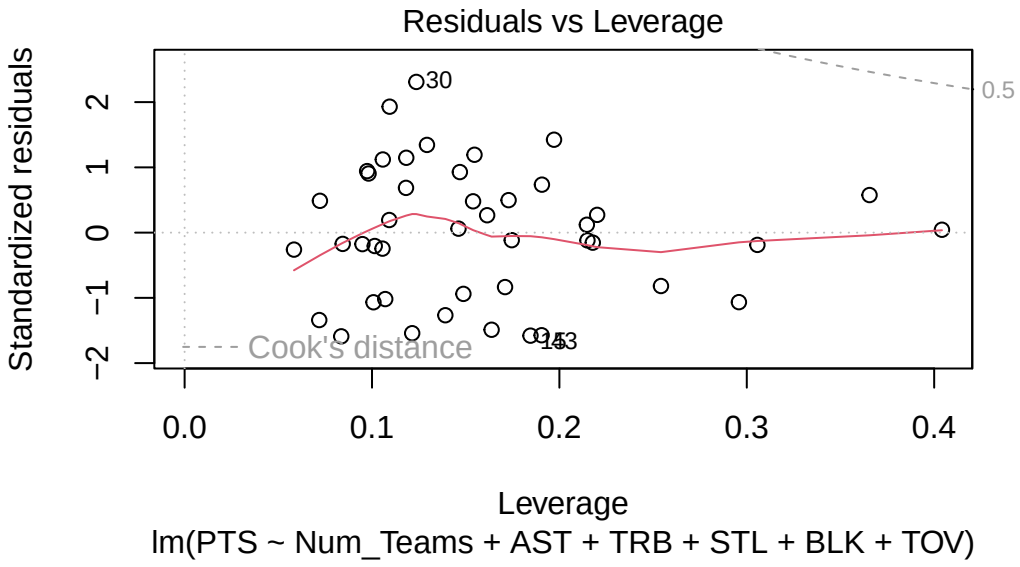
```
# Fit a multiple linear regression model
nba_model <- lm(PTS ~ Year * Num_Teams + AST + TRB + STL + BLK + TOV, data = nba_data)
nba_model_2 <- lm(PTS ~ Year + Num_Teams + AST + TRB + STL + BLK + TOV, data = nba_data)
model_ny <- lm(PTS ~ Num_Teams + AST + TRB + STL + BLK + TOV, data = nba_data)

plot(model_ny)
```




Q−Q Residuals
Standardized residuals
−1
0
1
2
25
30
31
−2
−1
0
1
2
Theoretical Quantiles
lm(PTS ~ Num_Teams + AST + TRB + STL + BLK + TOV)


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
30
31
25
95
100
105
110
115
Fitted values
lm(PTS ~ Num_Teams + AST + TRB + STL + BLK + TOV)




```
# Summary of the linear model
msummary(
  list(
    "times Year" = nba_model,
    "Without Year" = model_ny,
    "plus year" = nba_model_2
  ),
  fmt = 2
  )
```

```
summary(model_ny)
```

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

Residuals:
     Min       1Q   Median       3Q      Max
-2.01666 -1.04222 -0.04709  0.85908  2.86532

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 38.99412   20.62114   1.891 0.066477 .
Num_Teams   -0.11553    0.30279  -0.382 0.704972
AST          3.51264    0.27285  12.874 3.09e-15 ***
```

	times Year	Without Year	plus year
(Intercept)	3648.76	38.99	−68.40
	(1438.71)	(20.62)	(141.22)
Year	−1.81		0.06
	(0.72)		(0.07)
Num_Teams	−121.55	−0.12	−0.23
	(46.76)	(0.30)	(0.34)
AST	3.82	3.51	3.50
	(0.28)	(0.27)	(0.28)
TRB	0.27	0.98	0.76
	(0.43)	(0.31)	(0.42)
STL	−2.01	−3.42	−3.08
	(0.95)	(0.80)	(0.92)
BLK	−3.75	−6.02	−5.51
	(1.64)	(1.45)	(1.61)
TOV	−0.80	0.09	0.21
	(0.64)	(0.52)	(0.54)
Year × Num_Teams	0.06		
	(0.02)		
Num.Obs.	44	44	44
R2	0.967	0.960	0.961
R2 Adj.	0.960	0.954	0.953
AIC	153.6	158.0	159.3
BIC	171.4	172.3	175.4
Log.Lik.	−66.783	−71.011	−70.653
RMSE	1.10	1.22	1.21

```
TRB          0.97978    0.30797   3.181 0.002965 **
STL         -3.41709    0.79800  -4.282 0.000126 ***
BLK         -6.02443    1.45399  -4.143 0.000191 ***
TOV          0.08844    0.51968   0.170 0.865802
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.325 on 37 degrees of freedom
Multiple R-squared:  0.9602,    Adjusted R-squared:  0.9537
F-statistic: 148.8 on 6 and 37 DF,  p-value: < 2.2e-16
```

```
summary(nba_model_2)
```

```

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

Residuals:
     Min       1Q   Median       3Q      Max
-2.10159 -1.05765 -0.04215  0.78005  2.89171

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -68.40156  141.22015  -0.484  0.63106
Year          0.05646    0.07344   0.769  0.44702
Num_Teams    -0.23141    0.33974  -0.681  0.50014
AST           3.49823    0.27501  12.720 6.96e-15 ***
TRB           0.76455    0.41748   1.831  0.07533 .
STL          -3.07788    0.91575  -3.361  0.00185 **
BLK          -5.50569    1.61027  -3.419  0.00158 **
TOV           0.20509    0.54416   0.377  0.70847
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.333 on 36 degrees of freedom
Multiple R-squared:  0.9608,    Adjusted R-squared:  0.9532
F-statistic: 126.2 on 7 and 36 DF,  p-value: < 2.2e-16
```

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`.

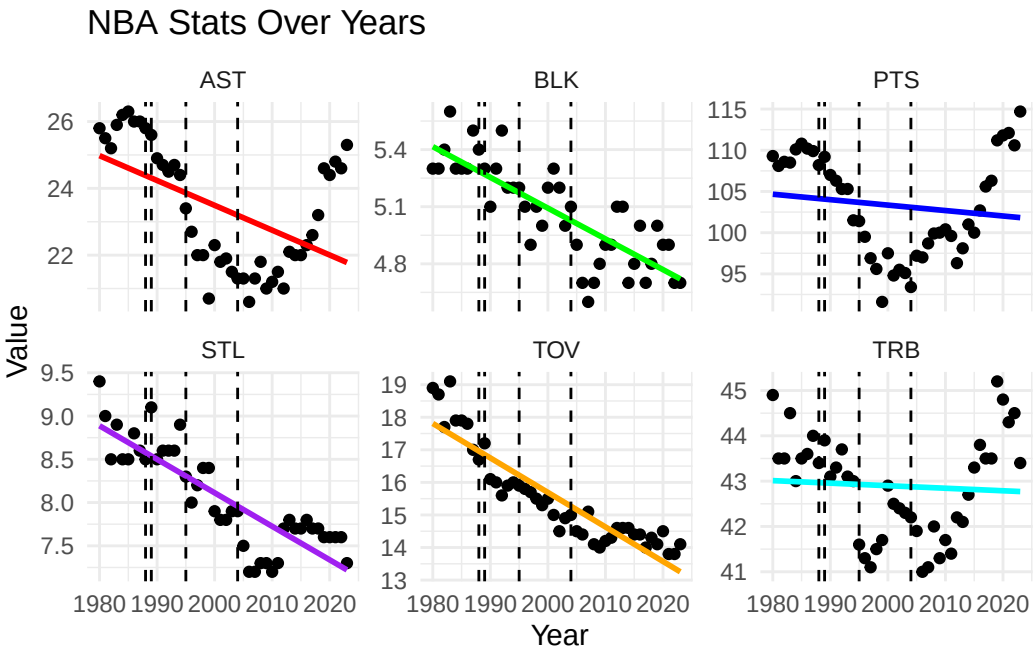
3.1.1 Model justification

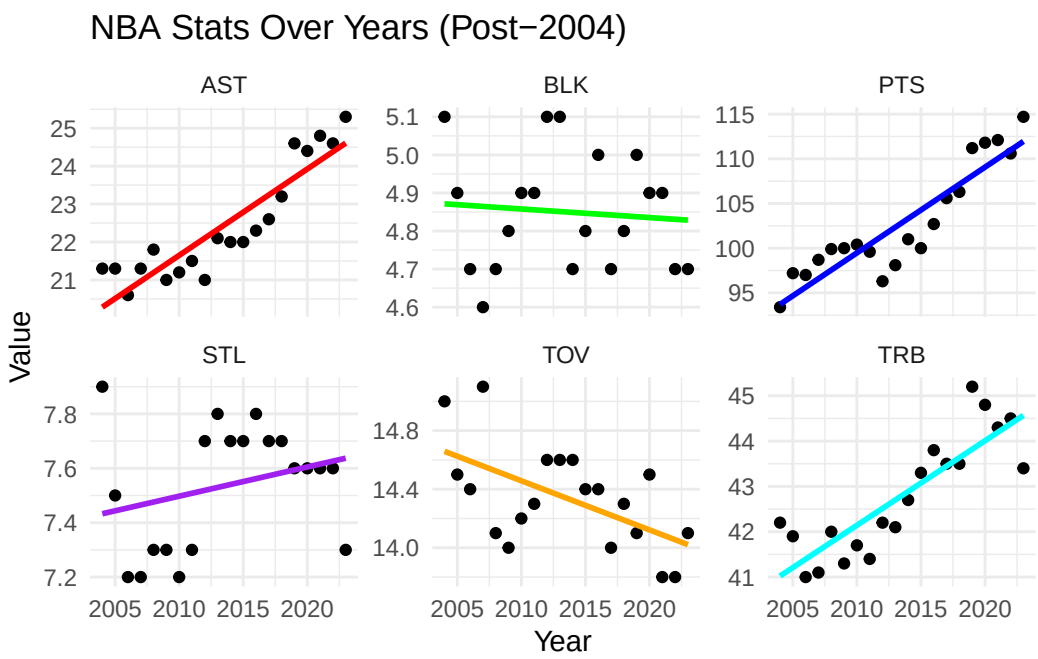
We expect a positive relationship between the size of the wings and time spent aloft. In particular...

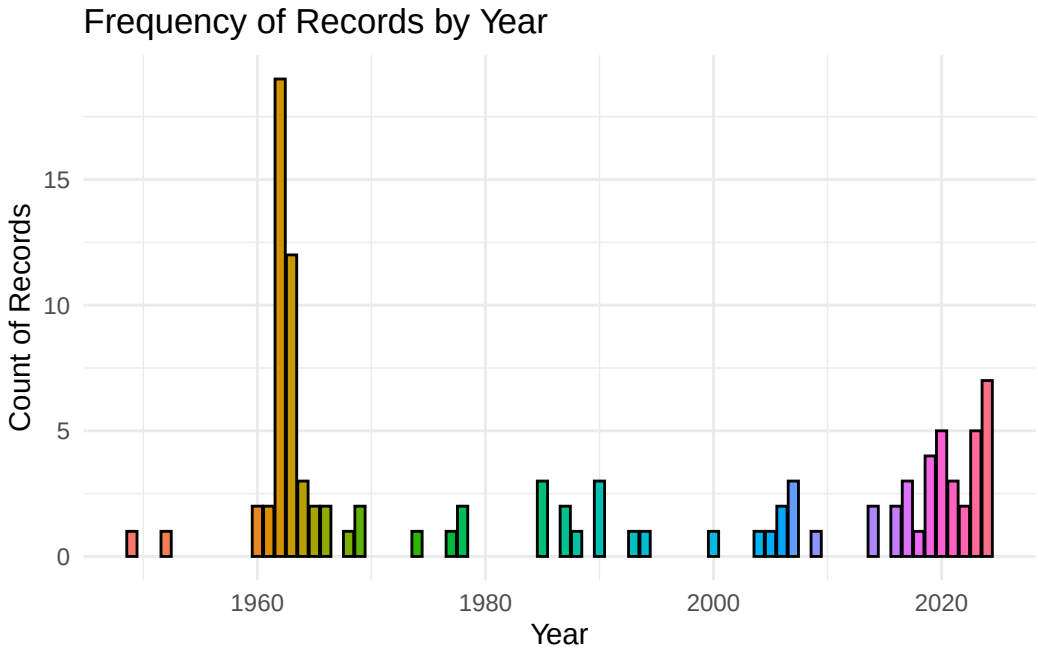
We can use maths by including latex between dollar signs, for instance θ.

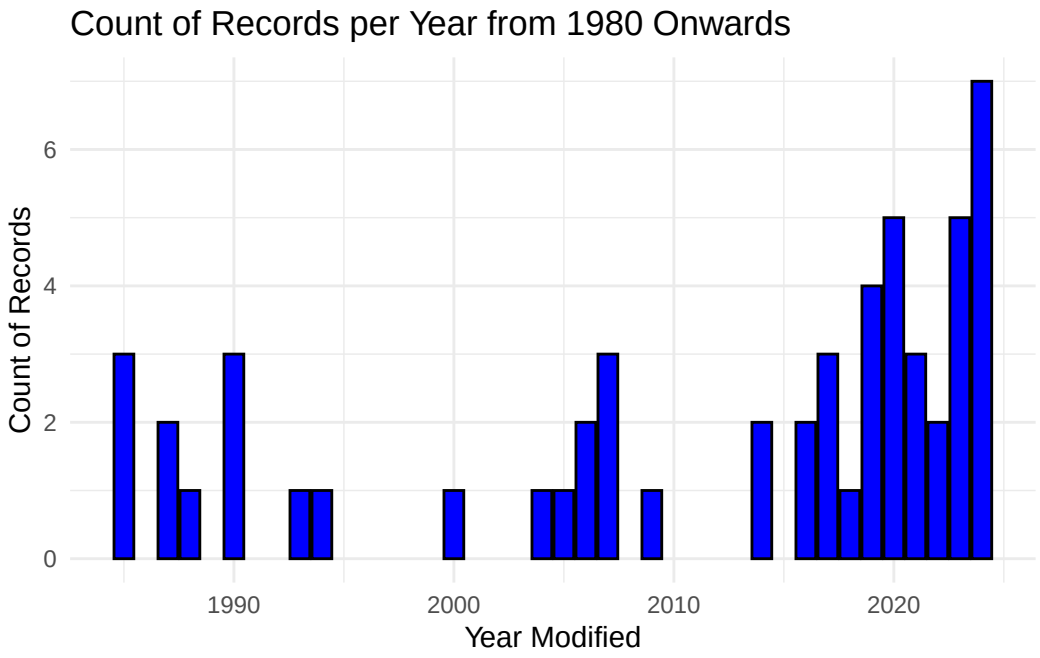
4 Results

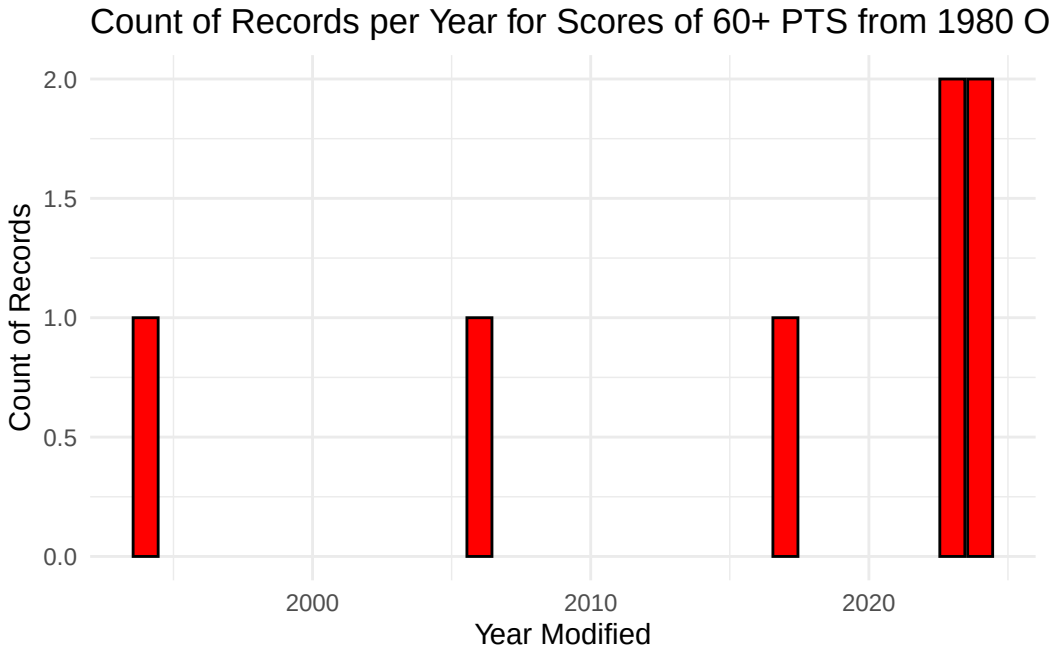
Our results are summarized in **?@tbl-modelresults**.




NBA Stats Over Years (Post−2004)
AST
BLK
PTS
STL
TOV
TRB
Value
Year
25
24
23
22
21
5.1
5.0
4.9
4.8
4.7
4.6
115
110
105
100
95
7.8
7.6
7.4
7.2
14.8
14.4
14.0
45
44
43
42
41
2005 2010 2015 2020


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






```
nba_avg_pts <- nba_data %>%
  group_by(Year) %>%
  summarize(Avg_PTS = mean(PTS, na.rm = TRUE))

comparison_data <- lebron_data %>%
```

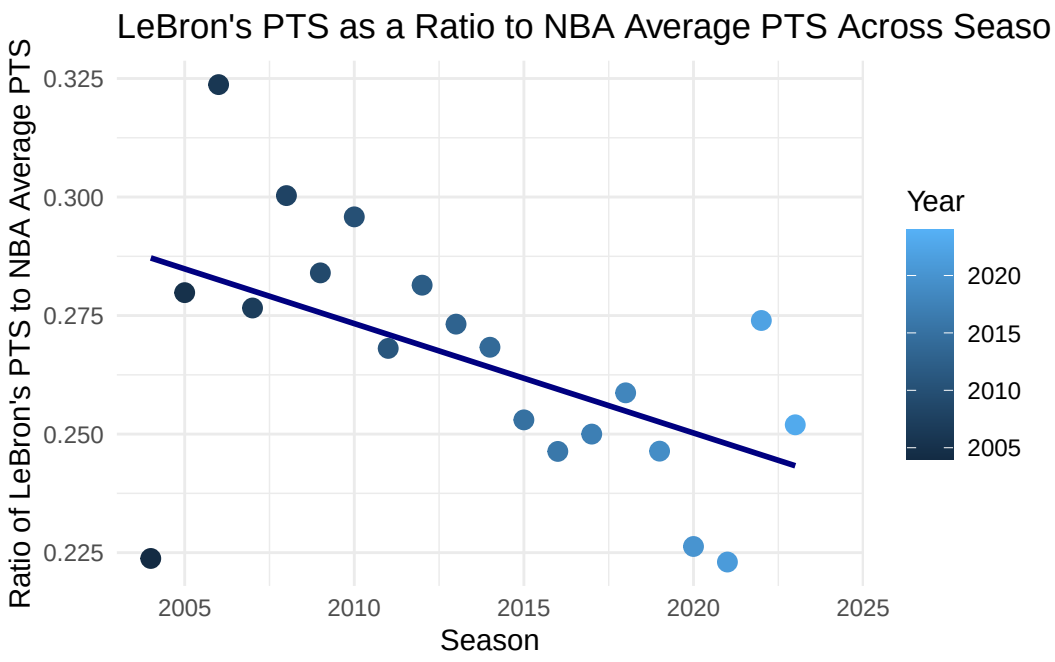
```
  rename(Year = Season) %>%  # Rename Season to Year for merging
  left_join(nba_avg_pts, by = "Year") %>%
  mutate(Ratio = PTS / Avg_PTS)

ggplot(comparison_data, aes(x = Year, y = Ratio)) +
  geom_point(aes(color = Year), size = 3) +
  geom_smooth(method = "lm", se = FALSE, color = "navy") +
  labs(title = "LeBron's PTS as a Ratio to NBA Average PTS Across Seasons",
       x = "Season",
       y = "Ratio of LeBron's PTS to NBA Average PTS") +
  theme_minimal()
```

```
`geom_smooth()` using formula = 'y ~ x'
```

```
Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
```

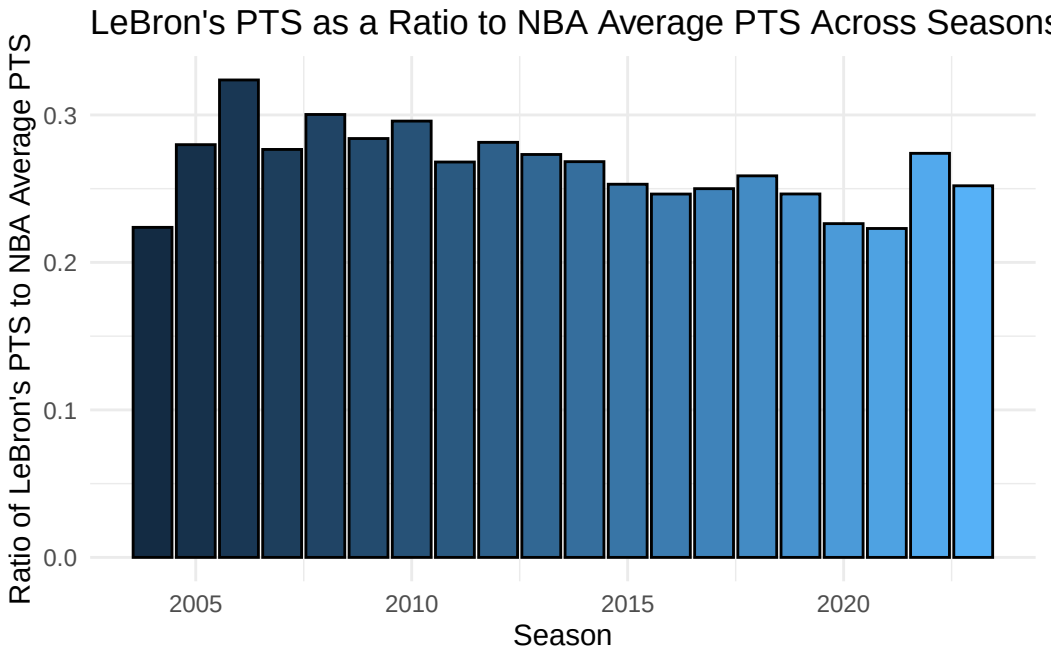
```
Warning: Removed 2 rows containing missing values (`geom_point()`).
```




```
ggplot(comparison_data, aes(x = Year, y = Ratio, fill = Year)) +
  geom_bar(stat = "identity", color = "black") +
  labs(title = "LeBron's PTS as a Ratio to NBA Average PTS Across Seasons (Bar Graph)",
       x = "Season",
```

```
      y = "Ratio of LeBron's PTS to NBA Average PTS") +
  theme_minimal() +
  theme(legend.position = "none")
```

```
Warning: Removed 2 rows containing missing values (`position_stack()`).
```



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

B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows…

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows…

B.2 Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows… This suggests…

?@fig-stanareyouokay-2 is a Rhat plot. It shows… This suggests…

References

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

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *Palmerpenguins: Palmer Archipelago (Antarctica) Penguin Data*. https://doi.org/10.5281/zenodo.3960218.

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.