

Predicting NBA Longevity Through Draft and Performance Metrics

Eugene, Miles, & Andrew

1 Introduction

The NBA boasts an incredibly high concentration of talent, with approximately 500 of the world’s best players competing at the highest possible stage. Each year, sixty recruits are selected in the illustrious NBA draft, competing to make their mark on the league. Millions are invested in the draft yearly, and predicting the best talent in the pool is vital for teams to make necessary improvements in performance. Recent scholarship suggests that the draft is less effective in evaluating career longevity than previously thought. Literature suggests that short-term performance leading up to the draft might artificially inflate a player’s position in the draft (Ichniowski & Preston, 2012). Others have pointed to racial bias in media and upper management, which improves draft position for White individuals (Millian & Preston, 2024). In the NBA community, there is frequent discontent from fans based on their teams’ picks. All of this evidence suggests that there is contention with the assertion that better draft picks ultimately result in better players and subsequently longer careers. Draft day decisions shape the NBA for years to come and , and understanding the effects of pre-draft metrics is critical in determining the efficacy of previously implemented scouting techniques in the NBA.

This paper seeks to analyze the careers of retired NBA players drafted between 1990-2002. We approach this data with two guiding research questions. First, we ask: to what extent do draft positions and pre-draft metrics predict the longevity of NBA careers? Subsequently, we seek to determine if player longevity is better observed looking backwards. Thus, our second research question asks: to what extent does post-career performance metrics predict NBA players’ career longevity? We seek to compare the predictive power of pre-draft metrics and post-career metrics. Using a dataset sourced from Kaggle using sports-reference.com data, we observe the careers of NBA players from the 1990-2002 draft classes to answer our research questions.

2 Methods and Materials

As mentioned above, our dataset created by Ben Wieland allows us to observe the careers of NBA players drafted anywhere between 1990-2002. This original dataset had 1868 observation and 26 variables but we made several alterations to the data so that it could be used in models and interpreted effectively. First, we removed players with 0 years in the NBA since we want to analyze the careers of players who actually played in the NBA and to preserve the Poisson (figure with Yrs distribution including Yrs = 0 data). Secondly, we create draft-year limits to preserve draft classes that are entirely retired (1990-2002). This is important, as analyzing years that have active players would negatively alter our results, even if we were to account for retired status. Since we are using our response, Yrs, as a Poisson random variable, we created $\log Yrs$ (the natural log of Yrs) to analyze its relationship with different predictors. We wanted to investigate the relationship between a player’s career length and their age at the time they were drafted, so we created the `age_at_draft` variable. To make it, we joined the `nbaPlayers` dataset, which contained the birthday of each player. We then found the date that each NBA draft took place and subtracted the birthdays of players drafted in that year to find their age at which they were drafted. We also had a number of NAs in key predictors like 3-point percentage and free-throw percentage, and we made the decision to assign 0% to replace these NAs, as we are still interested in observing this population, who have likely not taken these types of shots. While not a perfect solution, we generalized that players to not take 3 pointers or free throws

likely struggle with them, so assigning 0% for these NAs remains the best way to include these players, and will not create too much noise. Additionally, in attempting to analyze basic performance metrics, such as points, rebounds, and assists, we standardized these metrics per 36 minutes, which is a frequently used technique in the NBA, as 36 minutes is considered a general full game played for a starter. We also create a variable to better account for position, creating two categories: backcourt players and frontcourt players. This illustrates whether the player contributes in the backcourt (typically guards and some smaller forwards) or the frontcourt, closer to the net (typically larger forwards and centers). Next, we centered several variables to make their interpretations more meaningful. We centered DraftYear at 1990 as that is the earliest draft year in our dataset. We also centered Pick at 1, Weight and Height at their averages, at age at 17, since that is the youngest level of age. There was also missing data for several variables that we wanted to use as predictors (figure with missing data patterns). When fitting models with these predictors, we remove observations lacking values for those variables. This editing leaves us with 625 observable players, with a total of 48 variables in the dataset, of which not all will be utilized in our statistical analysis.

3 Modeling

When beginning our modeling process, we first had to analyze our response variable. The response of interest, `Yrs`, is a count which is right skewed (See figure with `Yrs` distribution in appendix), leading us to model it using a Poisson regression. We began by assessing Poisson modeling conditions with our primary predictor, Round Type. We already had evidence of a Poisson response and we assume that players aren't significantly influencing each others' career lengths, position, or career statistics. To assess linearity, we plotted the natural log of Years versus draft pick (Figure `logYrs` vs. `Pk`). We observed that the relationship is linear although there is a lot of variability around the trend line. Thus, our linearity assumption is shaky but not completely violated so we continued by checking the relationship between mean and variance within the Round Type groups (See table with mean vs. var by RoundType in appendix). While variance does increase linearly as mean career length increases, we observed that variance is generally greater than the mean. This is evidence of overdispersion, and since the relationship is linear (see mean vs. var plot in appendix), we decided to utilize a quasipoisson model to address this.

In answering our research questions, we have decided to fit two separate models to assess longevity using pre-draft metrics and longevity using post-career metrics. We want to assess differences in predictive power, to ascertain the effect of pre-draft parameters as they compare to tangible statistical performance in predicting longevity. Thus, we built two quasi-Poisson models, one with purely pre-draft and draft predictors, and one with advanced performance metric predictors. We signify statistically significant evidence to reject null hypotheses with p-values less than 0.05. Including and removing variables was determined by considering Wald test coefficient outputs and drop-in deviance tests with the Anova function to determine predictive impact. Initially, we just considered how round type predicts career length. By a drop-in-deviance test against the null model, the variable was significant (p-value = nearly 0) and, unsurprisingly, estimated lottery picks to have the longest careers, at 11.3 years (model1.1 results table). Next, we wanted to test the effect of age when drafted. By a Wald-type test, this was a significant predictor in the model (p-value = nearly 0) and the model estimates that career length decreases as age at draft decreases (model1.2 output table). We then added player position into our model and, using a drop-in-deviance test, found that it significantly improved it (p-value = 0.002). Next, we checked if height improved our model, but it did not (DDT p-value = 0.6). This was surprising, since height is such a valued physical metric in competitive basketball leagues. One reason for this insignificance could be due to the fact that almost all NBA players already meet sufficiently tall standards, and those that don't typically make up for it through other physical metrics. Interestingly, weight was a significant predictor (DDT p-value = 0.03), estimating an increase in longevity as weight increases, holding round type, age at draft, and position constant (model1.5 output table). We were interested in whether the effect of weight on career length differed by position or age so we added interactions between position and weight and between age and weight. Conducting a drop-in-deviance test, we found that neither interaction made the model significantly better so we didn't include them in the final model. Finally, we wondered if the effect of draft pick order on longevity differed by round type so we pick and its interaction with round type. Conducting a drop-in-deviance test, we found that this interaction

is significant (p-value = 0.002) and kept it as our last addition to our final first model. The final model for research question one uses the variables `RoundType`, `Pk`, `Position`, `cWeight`, `age17`, and their interactions.

When investigating the relationship between draft year and career years, we see the average years played in the NBA has decreased progressively. As it appears, the average length of players career didn't change much by draft class though players drafted in 1996 have on the longest average years in the NBA of 8.6 years and in fact, Lottery Players in 1995 have an average career of 13 years. Seeing this, we explore if this is a good indicator for explaining the variability in the years in the NBA. It turns out it is not a significant predictor, shown by our Drop-In Deviance Test(df = 2, Deviance = 0.68211, F-stat = 0.1249, p-value 0.8826).

After showing that the years players were drafted is not significant in our model, we began to investigate players on-court efficiency to estimate their years in the NBA. Including players Win Share Per 48 minutes significantly reduced our model residual deviance(df = 1, Deviance = 142.02 58.198 p-value = 9.019e-14 < 0.05) and more variance is explained when the round a player is drafted interacts with their Win Share 48-minutes, (deviance = 352.95 F-stat = 90.943, p-value < 2.2e-16). Win Share Per 48-minutes measures the total win share a player contributes per 48 minutes (or one regulation-length game). This metric combines a player's offensive and defensive contributions, aggregating them into win shares, where one share is equivalent to a third of a team's wins. In doing so, players are credited with individual "wins" they provide for the team.

After establishing that on-court efficiency estimates career length better than just draft year, we investigated if the relationship between performance and longevity depends on a player's draft status. By introducing an interaction between Draft Round and Win Shares Per 48 Minutes (Model 3), we significantly reduced the model's residual deviance ($p < 0.01$), confirming that the impact of winning on career survival differs for Lottery picks versus Second Rounders. We further refined this by interacting Box Plus-Minus (BPM) with Value Over Replacement Player (VORP) to account for the effect between a player's rate of efficiency and their total accumulated value i.e. once a player has already accumulated massive career value (High VORP), their efficiency rate (BPM) becomes slightly less critical to keeping them in the league which is also an indicator that other metric are then used to assess players like Kevin Garnett (player with the highest VORP).

Finally, we developed an additional parameter to shed new light on the role of shooting specialists in seeing longer careers. By adding Three-Point Percentage (3P%) to the model, we observed a further reduction in residual deviance, allowing us to test if shooting specialists persist in the league independently of their all-around value. Our final model utilizes a quasipoisson family to correct for the observed overdispersion in the count data ($\phi \approx 1.61$), ensuring that our standard errors and p-values are robust. The final model to answer research question two uses the variables `RoundType`, `Position`, `cDraftYear`, `WS/48`, `BPM`, `VORP`, `3p%`, and their significant interactions.

4 Results

Research Question 1

$$\begin{aligned} \log(Yrs) = & \beta_0 + \beta_1 RoundTypePlayoff + \beta_2 RoundTypeTrueRound2 + \beta_3 age17 + \beta_4 positionC-F \\ & + \beta_5 positionF + \beta_6 positionG-F + \beta_7 positionG + \beta_8 cWeight + \beta_9 cPk + \beta_{10} RoundTypePlayoff \times cPk \\ & + \beta_{11} RoundTypeTrueRound2 \times cPk \end{aligned}$$

In analyzing our model results, we see numerous important coefficients worth examining. Firstly our intercept predicts that a 17 year old center player with an average weight for draft picks who was drafted as the first lottery pick will have a 15.3 year career in the NBA. Round type is a highly significant predictor, illustrating that a 17-year-old center player of average weight who was drafted first in true round 2 has an estimated career length of 6.7 years, compared to 10.6 years if drafted first in the playoffs and 15.3 years if drafted

Table 1: Final Model 1 Results

Variable	Estimate	exp(Estimate)	Standard Error	Test Stat	p-val
(Intercept)	2.7303	15.3375	0.11981	22.789	< 2e-16 ***
RoundTypePlayoff	-0.1391	0.8701	0.20912	-0.665	0.506063
RoundTypeTrue Round 2	-0.7368	0.4786	0.26086	-2.825	0.004889 **
age17	-0.0886	0.9152	0.01854	-4.781	2.19e-06 ***
position2C-F	0.3252	1.3843	0.08816	3.689	0.000245 ***
position2F	0.1103	1.1166	0.08228	1.341	0.180556
position2G-F	0.3958	1.4856	0.11378	3.479	0.000539 ***
position2G	0.2904	1.3369	0.11965	2.427	0.015516 *
cWeight	0.0030	1.0030	0.00141	2.144	0.032399 *
cPk	-0.0306	0.9699	0.00889	-3.442	0.000617 ***
RoundTypePlayoff:cPk	0.0138	1.0139	0.01285	1.075	0.282887
RoundTypeTrue Round 2:cPk	0.0275	1.0278	0.01071	2.564	0.010579 *

first in the lottery. Generally speaking, the role of age is quite interesting in our model. Every year increase in age when drafted is associated with an 8.5% decrease in career length, after accounting for draft pick number, position, and weight. We are 95% confident that each one-year increase age is associated with a decrease between 5% and 11.8% respectively. Additionally, after accounting for draft pick number, age, and weight, center forwards have 38% longer careers than centers, as opposed to 12% longer for forwards, 49% for guard forwards, and 34% for guards. Our weight variable suggests that every pound increase in weight is associated with a 0.3% increase in career length, after accounting for draft pick number, age, and position. Finally, we observe the interaction between roundtype and pick. We see that, after accounting for age at draft, position, and weight, every increase in draft pick for lottery picks is associated with a 3% decrease in career length as opposed to a 2% decrease for playoff picks and a 0.3% decrease for true round 2 picks.

Research Question 2

Now, we pivot to investigate our second question. After controlling for draft metrics, what is the relationship between on-court efficiency metrics (WS/48, VORP, BPM) and career duration?

$$\begin{aligned}
\log(\lambda) = & \beta_0 \\
& + \beta_1 I(\text{Round}_{\text{Playoff}}) + \beta_2 I(\text{Round}_{\text{Rd2}}) \\
& + \beta_3 I(\text{Pos}_4) + \beta_5 I(\text{Pos}_F) + \beta_6 I(\text{Pos}_{G-F}) + \beta_7 I(\text{Pos}_G) \\
& + \beta_8 (\text{DraftYear}_c) + \beta_9 (3P\%) \\
& + \beta_{10} (\text{WS}/48) + \beta_{11} (\text{BPM}) + \beta_{12} (\text{VORP}) \\
& + \text{Interactions terms (see below)}
\end{aligned}$$

$$\begin{aligned}
\text{Performance} \times \text{Draft} : & + \beta_{13} \cdot I(\text{Round}_{\text{Playoff}}) \cdot (\text{WS}/48) \\
& + \beta_{14} \cdot I(\text{Round}_{\text{Rd2}}) \cdot (\text{WS}/48) \\
& + \beta_{15} \cdot I(\text{Round}_{\text{Playoff}}) \cdot (\text{VORP}) \\
& + \beta_{16} \cdot I(\text{Round}_{\text{Rd2}}) \cdot (\text{VORP})
\end{aligned}$$

$$\text{Metric} : + \beta_{17} \cdot (\text{BPM} \times \text{VORP})$$

We see numerous notable findings in our analysis of post-career metrics as it relates to years in the NBA. Firstly our intercept illustrates that center players drafted in the lottery round in 1990 with 0 win shares per 48 minutes, 0 box plus minus, 0 VORP, 0 three point percent, and 0 assists per 36 minutes are expected to have a career length of 10.1 years. We also account, again, for round type in this model, finding that,

Figure 1: Relationship between Years in the NBA and Player Characteristics Statistics

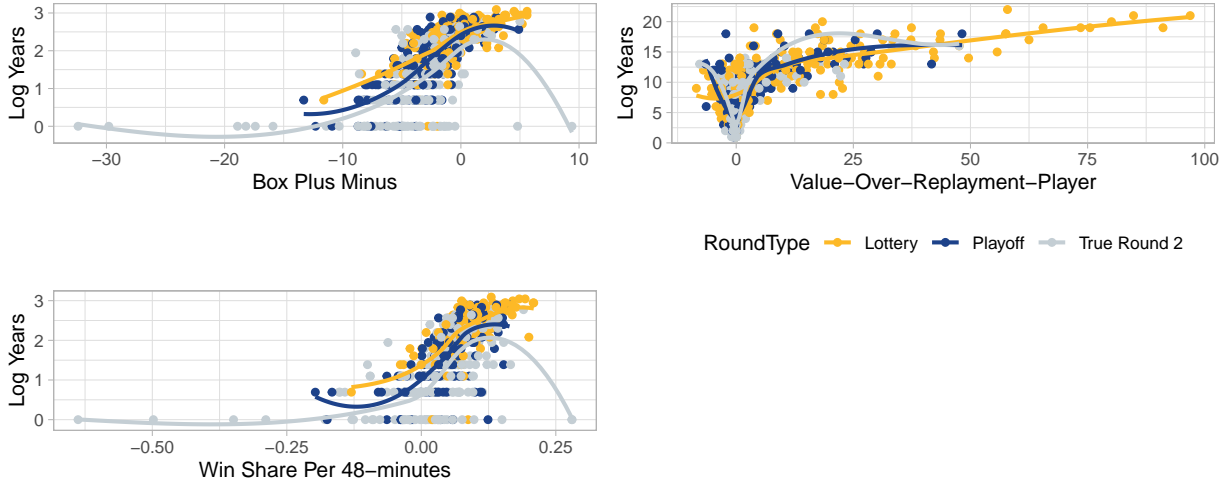


Figure 1: These plots illustrate relationships between log years and advanced statistical metrics such as BPM, WS/48, and VORP. We see that generally, increasing averages in these metrics results in increases in log years. Notably, some BPM and WS outliers create a downward slope for Round 2 players

Table 2: Final Model 2 Results

Variable	Estimate	exp(Estimate)	Standard Error	Test Stat	p-val
(Intercept)	2.3159	10.1341	0.15825	14.634	< 2e-16 ***
RoundTypePlayoff	-0.4097	0.6639	0.10978	-3.732	0.000208 ***
RoundTypeTrue Round 2	-0.3322	0.7173	0.09413	-3.529	0.000448 ***
position2C-F	0.0118	1.0118	0.07103	0.165	0.868620
position2F	-0.1666	0.8465	0.06482	-2.571	0.010381 *
position2G-F	-0.0841	0.9193	0.08747	-0.962	0.336489
position2G	-0.1813	0.8342	0.08214	-2.207	0.027684 *
cDraft Year	-0.0032	0.9968	0.00517	-0.620	0.535344
‘WS/48’	0.8359	2.3069	1.24424	0.672	0.501953
BPM	0.0694	1.0718	0.01827	3.796	0.000162 ***
VORP	0.0176	1.0178	0.00388	4.536	6.92e-06 ***
‘3P%’	0.3377	1.4017	0.17027	1.983	0.047770 *
AstPer36	-0.0266	0.9738	0.01553	-1.713	0.087301 .
RoundTypePlayoff:‘WS/48’	3.7219	41.3449	1.35295	2.751	0.006119 **
RoundTypeTrue Round 2:‘WS/48’	-2.6348	0.0717	1.09319	-2.410	0.016239 *
BPM:VORP	-0.0031	0.9969	0.00066	-4.785	2.15e-06 ***
RoundTypePlayoff:VORP	-0.0003	0.9997	0.00466	-0.057	0.954946
RoundTypeTrue Round 2:VORP	0.0260	1.0264	0.00488	5.332	1.37e-07 ***

Figure 2: Relationship Between NBA Player Assists/36 Minutes, Three-Point Percentages and Log Years In NBA

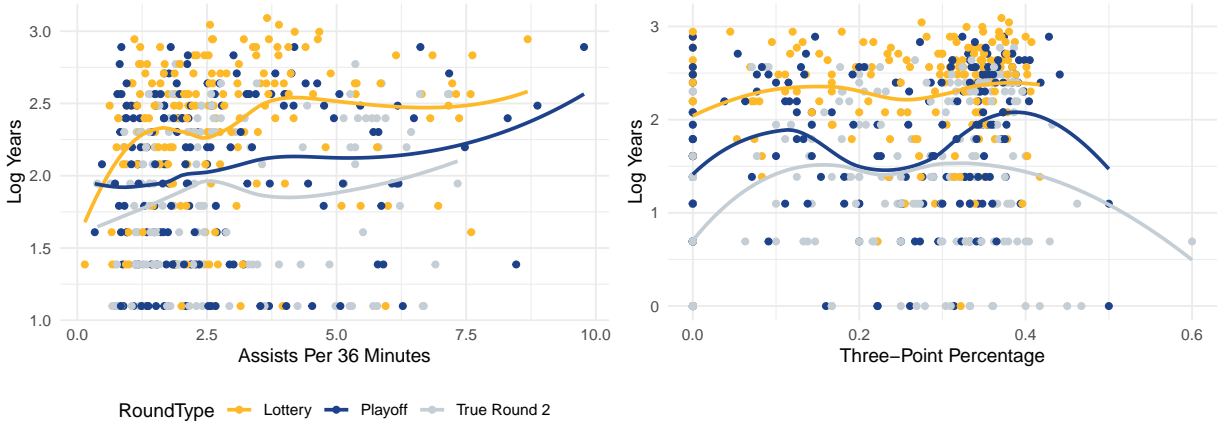


Figure 2: Relationship Between NBA Career Length and Key Performance Metrics (3P% and Assists per 36 Minutes). Illustrates how general increases in 3P% and Assists per 36 result in average increases in log years. Important to note that inflated 3 point percentages due to small sample sizes for players creates downwards trend as 3P% reaches above 0.4

after accounting for position, draft year, box plus minus, 3 point percent, and assists per 36 minutes, players with 0 win shares per 48 minutes and 0 VORP drafted in the playoffs are expected to have a 34% shorter career than lottery picks, compared to 28% shorter for round 2 picks. In regards to position, and after accounting for round type, draft year, win shares per 48 minutes, box plus minus, VORP, 3 point percent, and assists per 36 minutes, center forwards have 1.2% longer careers than centers, compared to 15% shorter careers for forwards, 8% shorter for guard forwards, and 17% shorter for guards. Now, interpreting our advanced metrics, and beginning with win shares per 48 minutes, we see that every additional win share per 48 minutes a lottery pick player has is associated with a 131% longer career, after accounting for position, draft year, box plus minus, VORP, 3 point percent, and assists per 36 minutes. For box-plus minus, we find that every additional box plus minus point a player with 0 VORP has is associated with a 7.2% longer career, after accounting for round type, position, draft year, win shares per 48 minutes, 3 point percent, and assists per 36 minutes. In observing VORP (Value-Over-Replacement-Player), every additional VORP point a player drafted in the lottery round with 0 box plus minus has is associated with a 1.78% longer career, after accounting for position, draft year, win shares per 48 minutes, 3 point percent, and assists per 36 minutes. Three point percentage is also a relatively important metric to analyze, and we see that every additional percent of 3 point shots made by a player is associated with a 40% longer career, after accounting for round type, position, draft year, win shares per 48 minutes, box plus minus, VORP, and assists per 36 minutes. Finally, when investigating the interaction between round type and VORP, we see that, after accounting for position, draft year, win shares per 48 minutes, box plus minus, 3 point percent, and assists per 36 minutes, every additional VORP point a player drafted in the playoff round with 0 box plus minus has is associated with a 1.75% longer career, compared to 4.4% longer for round 2 picks and 1.78% for lottery picks. Every single one of these interpretations is significant at a 95% confidence level.

5 Discussion

We see impressive predictive power in significance in both models addressing their respective research question. Of course, model 2 which examines post-career statistical metrics and their relationship with longevity seems to be more predictive, generally. This is to be expected, as there is more predictors, and it is much easier to predict longevity looking “backwards” compared to “forwards”. That being said, we still see relatively impressive predictive power in model 1, which analyzes pre-draft metrics and players’ physical characteristics

to predict longevity. We see very significant implications with most of our predictors, and, as hypothesized, players drafted in earlier rounds like lottery and playoffs see much more longevity than those drafted in the second round, and the same is true for each increase in pick in relation to round type. The following discussion highlights key findings, the implications of those findings. We also note the generalizability of these models, identifying strengths, weaknesses, and limitations, while also outlining how future studies might improve our research.

In model 1, we see numerous factors are involved in significantly predicting an NBA player’s career length. The variables that are significant are Round Type, Pick, Age At Draft, Weight, Position, and the interaction between Pick and Round Type. One really interesting finding was the significance of age in our model, yet this makes sense when accounting the nature of the draft. Players who are drafted young, particularly those that are 17-18, were so talented they skipped college to pursue the profession basketball league. These players seem to have the highest odds of longevity, and we see this when examining the league’s “legends”: players like Kobe Bryant and LeBron James were national stars at 17 and 18 respectively, and boast extremely lengthy careers. We were surprised to see that ‘Height’ was an insignificant variable, but weight remained significant. This likely suggests that stronger players higher in weight AND height might see slightly more longevity based on their size rather than their height. Finally, it is especially interesting to see the role of position, namely that centers see the shortest average careers in the NBA. This makes sense, as centers typically rely on their physical metrics to achieve success, with less emphasis on technical skill compared to guards and forwards, and suggests that physicality matters less than pure skill in the NBA, and some of our findings in model 2 suggest this.

It is interesting that our two models predict different positions to have the longest careers. It highlights the importance of what models account for. When accounting for just pre-career variables, we predict that guard forwards have the longest careers but when accounting for career-long metrics (and assuming 0 win shares), we predict center forwards to have the longest careers. It is also compelling to see how significant our advanced metrics, like VORP, WS/48, and BPM ended up being. We saw that VORP had extremely significant relationships both alone and in interactions, which makes sense, as this variable articulates the contributions a player makes compared to a baseline average player, or a “replacement player”. This finding bolsters the efficacy of using advanced aggregated metrics to determine a player’s skill and, in tandem, longevity.

We asked our research questions to evaluate the extent to which draft metrics are ineffective in predicting longevity, and we see that they are highly significant in predicting NBA career lengths. There are numerous implications of this: firstly, we now know that investing in players through the draft, especially when a team has a lottery pick, is quite successful in producing talent that remains in the league. It is not perfectly clear the extent to which bias contributes to these draft decisions (bias suggested by Ichniowski & Preston 2012, Millian & Preston, 2024), yet evident that choosing young prospective stars gleans success, especially when a team has a early pick (with a impressive rate of longevity for the first pick). Unfortunately, we believe there are too many intangibles that occur during scouting processes, and likely even in player-by-player differences in unmeasurable metrics that hinder the ability to use our model findings to predict future draft results, but our model of pre-draft performance is still quite eye-opening. Our model with post-career performance metrics illustrates the impressive capability of backwards looking statistical predictors in analyzing career length. We see that these advanced metrics like VORP, WS/48, and BPM significantly predict career lengths. This is an interesting finding that suggests these underused, more granular statistical metrics can play a great role in determining players’ success. This might be a useful implication, which suggests that players with above average results in these metrics early in their career might see more success in the long run than others.

Despite many significant results in our statistical analysis, there are numerous problems with the generalizability of this study, as well as notable confounding variables that inhibit our true ability to draw conclusions. We are disappointed to only be able to analyze the years 1990-2002, as this significantly inhibits our generalizability. The NBA faces extreme structural changes each year, and over the course of a decade, the league alters greatly through rule developments and cultural differences (Nourayi, 2019). This is the challenge in analyzing the NBA. There is so much variation from year-to-year and from decade-to-decade. Take, for example, the G-league, which is the pre-professional league where talents develop to become NBA-ready.

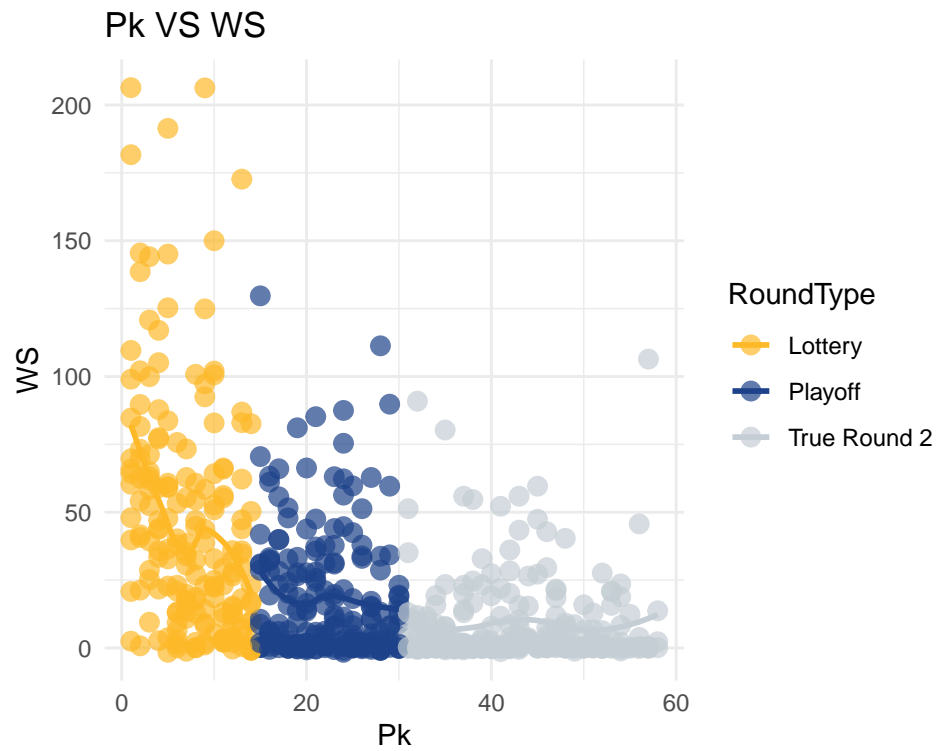
This league was established in 2001, and now gives underdeveloped players an opportunity to develop their skills, and many make it to the professional league after time in the G-league. This likely alters the relationships between metrics and years, and harms our studies generalizability, as we only discuss 1990-2002 careers. There are also a number of confounding variables we do not have access to but would be important in predicting. Firstly, it would be really helpful to have injury data, as injuries shape the longevity of players greatly, and even the highest-performing stars fall victim to season and career-ending injuries, and these might significantly alter our results. Additionally, we have no way to account for player role, which would be interesting to analyze, as some types of players fit better into the league's dynamics than others, and some roles see lesser performance metrics than others. We would also like to observe the effects of coaching, the role of certain teams in producing long-lasting talent, and other variables related to team that would adjust pre-draft and post-career metrics, as well as career length.

We also have several limitations to our model to acknowledge. Since quasi-Poisson regression is an adhoc model, it doesn't use full likelihoods, only quasi-likelihoods. Therefore, we can't calculate AIC or BIC to compare our non-nested models and we can't conduct goodness of fit or Vuong tests, which would boost the credibility of our results. We also had to reduce our sample size several times to deal with missing or incomplete data, which could weaken the strength of our models. We also didn't consider the possible correlation between career lengths of players drafted in the same year or to the same team. This could be accounted for by adding a random effect for draft year or team.

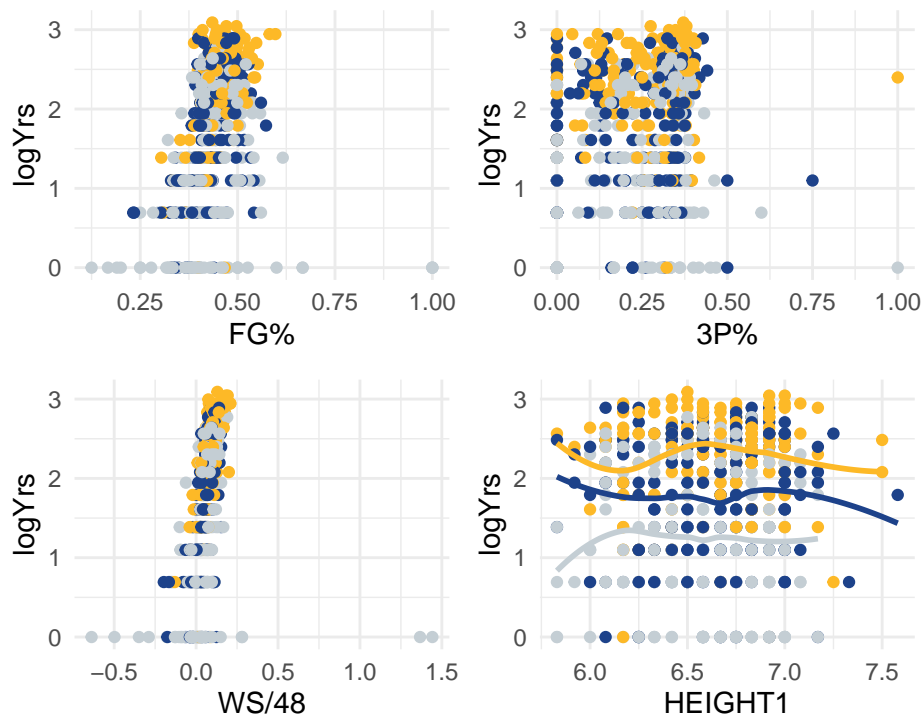
There is a great deal of potential for future research investigating similar questions. First, of course, it would be quite interesting to observe the modern league today, and re-analyze certain variables like three-point percentage and height, which seem to have become increasingly important metrics in today's league. Additionally, we find that a multilevel structure to analyzing careers, particularly one that illustrates the role of teams as a level 2 unit, would produce great insight that could be better applied and generalized for future NBA picks. We would also be interested in pursuing a zero-inflated model, potentially a hurdle model, that investigates what causes players to play versus the players that never touch the court. All of the above would create great future studies, and we hope to investigate the extent to which these future plans can better predict longevity.

6 Appendix & References

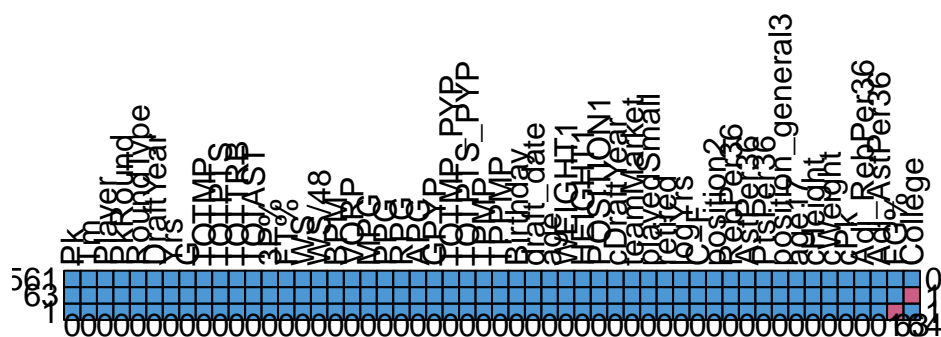
6.1 Figures and tables



This graph exploring the relationship between pick and winshare illustrates that lottery players tended to have far more win shares throughout careers than playoff and round 2 picks, who both finished with similar averaged win shares.



These graphs illustrate the relationships between years and field goal percentage, 3 point percentage, win shares per 48, and height, based on round type.



Pk Tm Player PkRound RoundType DraftYear Yrs G TOTMP TOTPTS TOTTRB TOTAST

```

## 561 1 1 1 1 1 1 1 1 1 1 1 1
## 63 1 1 1 1 1 1 1 1 1 1 1 1
## 1 1 1 1 1 1 1 1 1 1 1 1
## 0 0 0 0 0 0 0 0 0 0 0 0
## 3P% FT% WS WS/48 BPM VORP MPG PPG RPG APG GPYP TOTMP_PYP TOTPTS_PYP TPMP
## 561 1 1 1 1 1 1 1 1 1 1 1 1
## 63 1 1 1 1 1 1 1 1 1 1 1 1
## 1 1 1 1 1 1 1 1 1 1 1 1
## 0 0 0 0 0 0 0 0 0 0 0 0
## TPPMP Birthday draft_date age WEIGHT1 HEIGHT1 POSITION1 cDraftYear
## 561 1 1 1 1 1 1 1
## 63 1 1 1 1 1 1 1
## 1 1 1 1 1 1 1
## 0 0 0 0 0 0 0
## teamMarket playedSmall retired logYrs C-F position2 RebPer36 AstPer36
## 561 1 1 1 1 1 1 1
## 63 1 1 1 1 1 1 1
## 1 1 1 1 1 1 1
## 0 0 0 0 0 0 0
## PtsPer36 position_general3 age17 cHeight cWeight cPk Adj_RebPer36
## 561 1 1 1 1 1 1
## 63 1 1 1 1 1 1
## 1 1 1 1 1 1
## 0 0 0 0 0 0
## Adj_AstPer36 FG% College
## 561 1 1 1 0
## 63 1 1 0 1
## 1 1 0 1 1
## 0 1 63 64

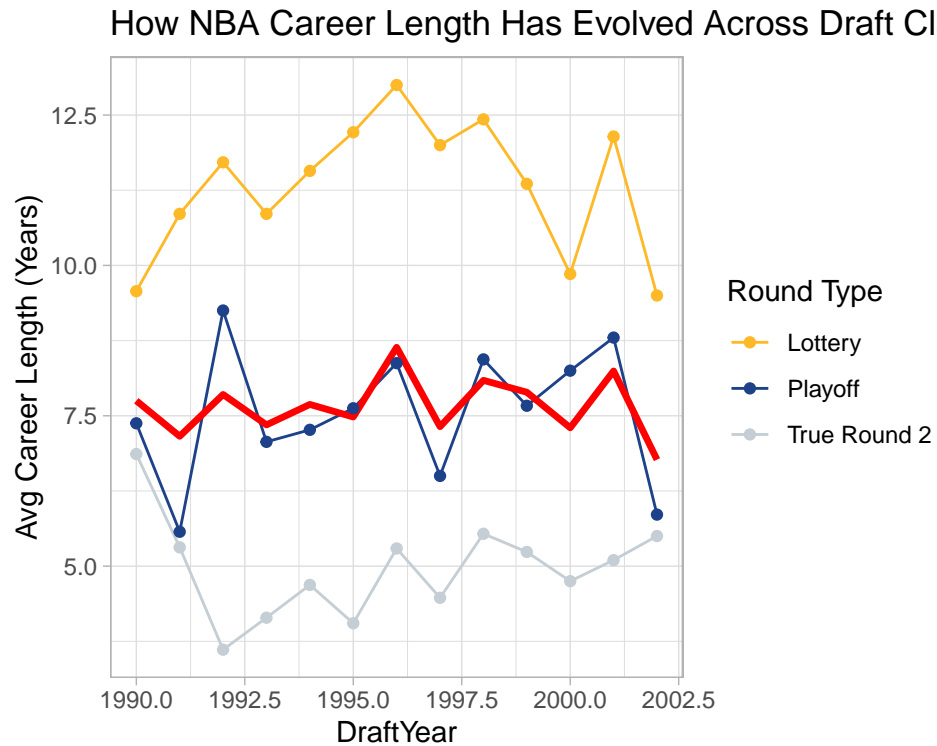
```

Plot of missing data patterns.

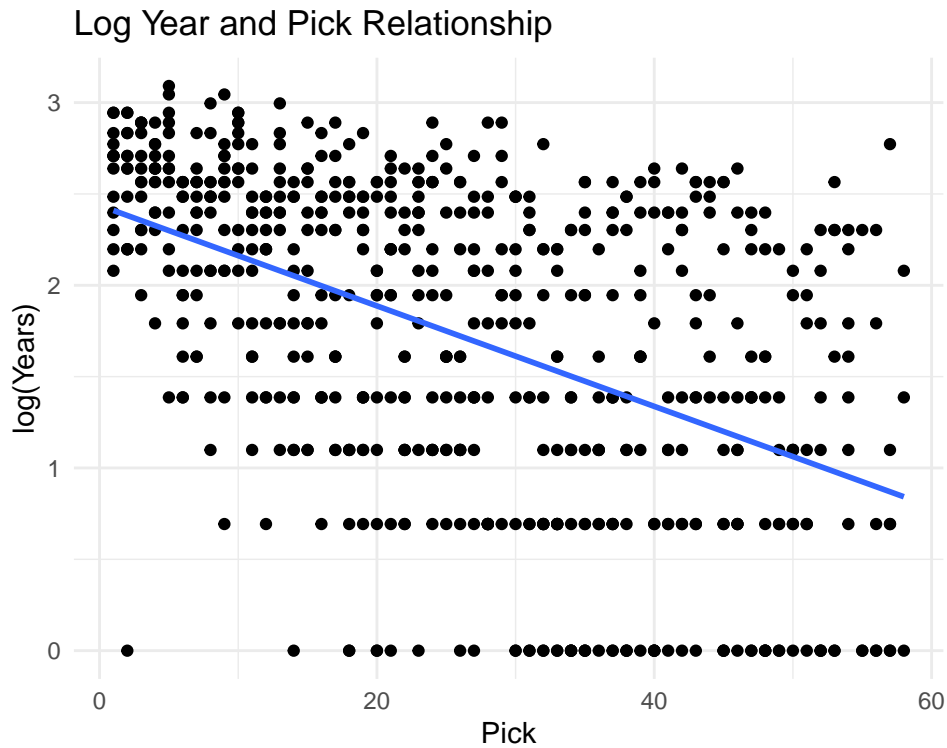
Yrs	n	mean_TOTMP	sd_TOTMP	num_games	sd_games	mean_RPG	sd_RPG	mean_PPG	sd_PPG
1	65	267.9538	418.9553	23.56923	20.01325	1.390769	0.8523356	2.838462	2.055989
2	65	716.0923	732.9722	58.46154	36.32625	1.530769	0.7079731	3.400000	2.116823
3	54	1324.1111	984.7369	101.12963	47.47752	1.731481	0.8590965	4.020370	2.166589
4	55	2063.4727	1349.0231	152.38182	60.68452	2.185455	1.0069055	4.370909	2.043569
5	30	3114.3667	1899.1053	195.10000	65.73109	2.880000	1.4237881	5.146667	2.630423
6	26	5306.3462	2384.0711	282.61538	68.82795	3.034615	1.7090213	6.650000	2.525272
7	28	6709.5000	3218.6845	356.67857	88.73851	2.964286	1.7016798	6.196429	2.584782
8	26	8064.4231	3871.4662	391.19231	98.03266	3.746154	1.5448575	7.811539	3.878590
9	34	10443.9118	4241.0979	488.00000	81.41774	3.844118	1.5585913	7.732353	3.993382
10	35	12982.6571	4697.7225	574.25714	83.31490	4.091429	1.6670459	8.294286	3.405009
11	40	14823.4000	5412.4680	611.90000	118.22616	3.907500	1.6788866	9.450000	4.555020
12	37	16807.1351	5640.5379	725.45946	92.97120	4.040540	1.8200213	8.656757	4.102610
13	44	21948.2045	5874.0873	824.15909	74.43678	4.504545	1.9781310	10.138636	4.223920
14	24	25027.5417	6648.5704	864.16667	91.00772	4.529167	1.8841396	12.491667	4.880834
15	16	28016.0625	4053.6718	945.18750	106.91693	5.943750	2.1651694	13.443750	3.704226
16	11	30053.3636	6078.0697	1042.72727	94.19245	6.281818	1.5071949	11.845455	3.959132
17	11	32785.2727	6802.8139	1085.00000	135.50793	5.345454	2.6234952	12.927273	2.804315
18	12	34404.3333	8432.2589	1156.58333	124.24204	5.258333	2.6314387	13.375000	4.297806
19	7	42226.0000	6399.1523	1301.57143	105.97933	7.285714	3.1205921	15.714286	5.280918
20	2	43815.5000	6818.6307	1336.50000	13.43503	3.700000	2.1213203	19.800000	7.353910

Yrs	n	mean_TOTPTS	sd_TOTPTS	num_games	sd_games	mean_RPG	sd_RPG	mean_PPG	sd_PPG
21	2	50893.0000	671.7514	1492.00000	42.42641	8.750000	1.7677670	19.250000	2.050610
22	1	46367.0000	NA	1541.00000	NA	4.300000	NA	16.700000	NA

Summary statistics for each level of Yrs.



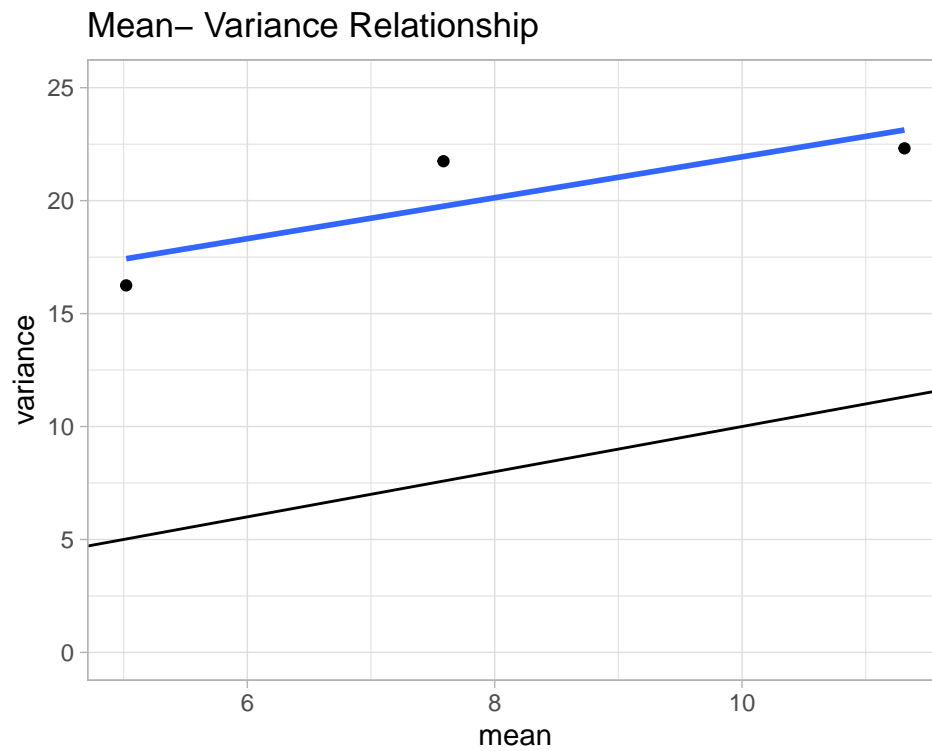
This figure shows how the average number of years played in the NBA varies by draft year, separated by draft round type. Colored lines represent mean career length for each round type, while the red line indicates the overall league-wide average career length for each draft year.



Graph of pick vs. log(years)

shows a linear trend with high variability around the trend line.

Table of mean vs. variance by round type.



Graph of mean vs. vari-

ance by round type compared to a one to one ($y = x$) relationship (black line).

Table 4: Variables Table, Model 1

Variable	Role	Type	Units_or_Levels	Description
Years	Response	Numeric (count)	Years (1–22)	Total number of seasons played in NBA
RoundType	Explanatory	Factor	Lottery / Playoff / Round 2	Categorical draft round classification
Pick	Explanatory	Numeric (ordinal)	1–60	Overall draft pick number
Position	Explanatory	Factor	G / G-F / F-G / F / F-C / C-F / C	Generalized player position
cWeight	Explanatory	Numeric	Pounds (centered)	Centered player body weight
DraftAge	Explanatory	Numeric	Years	Player age at draft

Table 5: Variables Table, Model 2

Variable	Role	Type	Units_or_Levels	Description
Yrs	Response	Numeric (count)	Years (1–22)	Total number of seasons played in NBA
RoundType	Explanatory	Factor	Lottery / Playoff / Round 2	Categorical draft round classification
Position	Explanatory	Factor	G / G-F / F-G / F / F-C / C-F / C	Generalized player position
cDraftYear	Explanatory	Numeric	Years since 1990	Centered draft year
WinShare/48	Explanatory	Numeric	Win Shares per 48 minutes	Player contribution to team wins
BoxPlusMinus	Explanatory	Numeric	BPM units	Box Plus/Minus
ValueOverReplacementPlayer	Explanatory	Numeric	VORP units	Estimated provided value over replacement player
3P%	Explanatory	Numeric	0–1	Three-point shooting percentage

Distribution of Years Played in NBA, Players Drafted 1990–1994

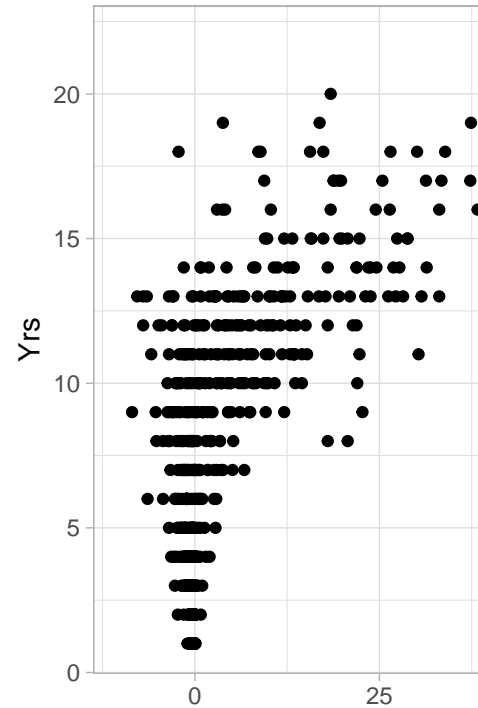
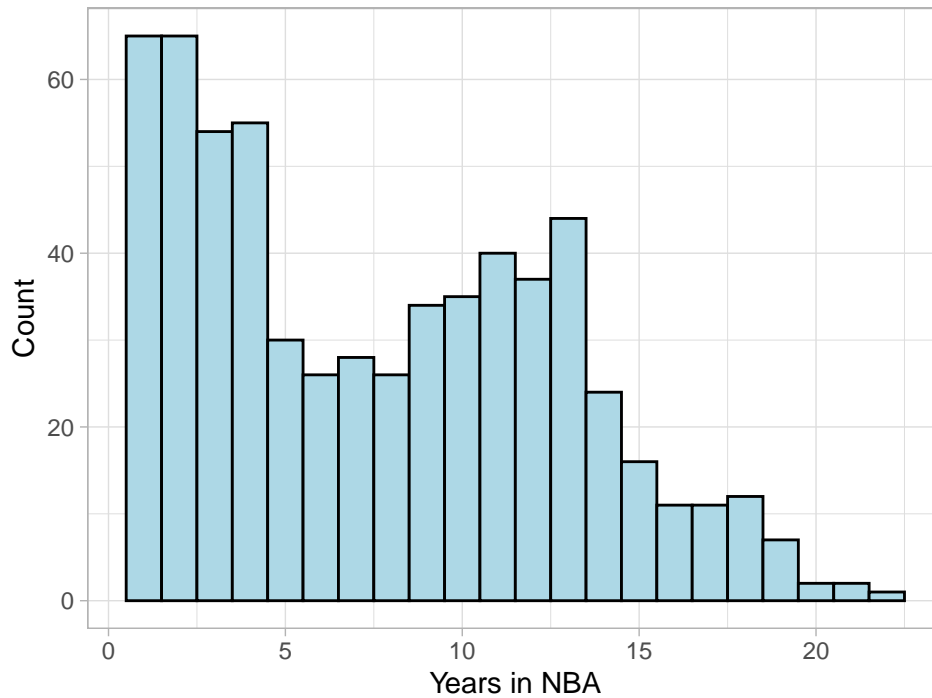


Table 6: Mean and Variance of Years Played by Draft Round Type

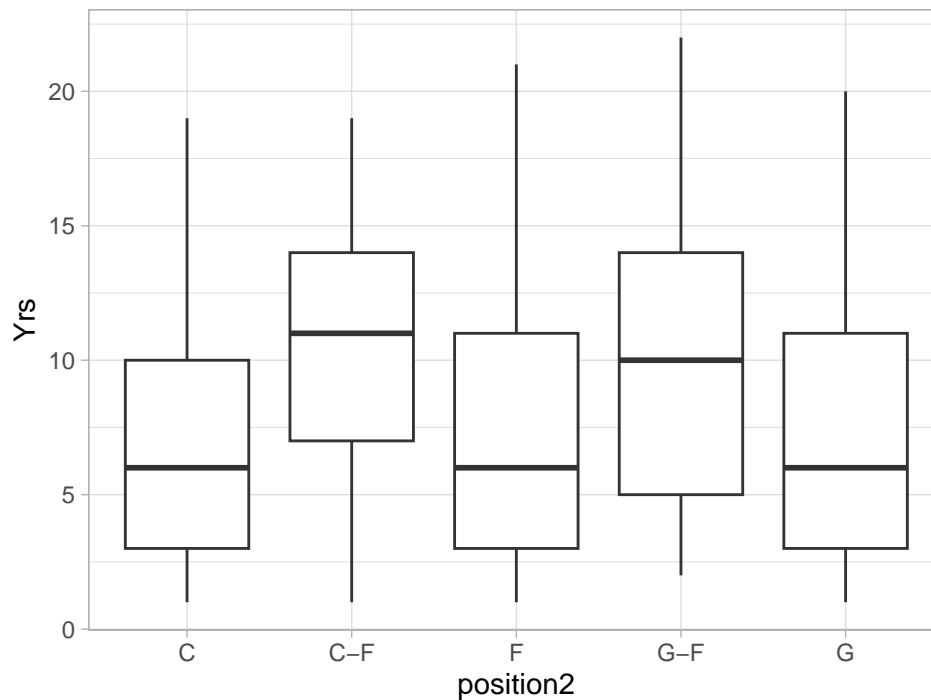
Round Type	Mean Years Played	Variance
Lottery	11.30	22.3
Playoff	7.59	21.7
True Round 2	5.02	16.2

6.2 Dataset code

Creation of our final dataset, including new and centered variables.

This code creates the age variable. It corrects syntax to fit with our base dataset and fills in some missing data before joining and filling in more missing data. We then record the draft date of each year and subtract each players' birthday from it to obtain their age when they were drafted. We also filter out duplicate rows caused by players with the same names in the `nbaPlayers` dataset, which contains more than just drafted players and from more years than our base dataset.

Years in the NBA and Player Position



6.3 Calculations for interpretations

```
## [1] 10.55098
```

```
## [1] 6.682209
```

```
## [1] 0.0848291
```

```
## [1] 0.0301263
```

```
## [1] 0.01664139
```

[1] 0.003129459

Calculations made while interpreting coefficients in model 1.

[1] 0.3361382

[1] 0.2826652

[1] 0.1535006

[1] 0.080697

[1] 0.1658065

[1] 0.0031983

[1] 0.0262497

[1] 1.01751

[1] 1.044599

Calculations made while interpreting coefficients in model 2.

7 References

Ichniowski, C., & Preston, A. E. (2012, March). Does March Madness Lead to Irrational Exuberance in the NBA Draft? High-Value Employee Selection Decisions and Decision-Making Bias. National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w17928/w17928.pdf

Nourayi, M. (2019). Strategically driven rule changes in NBA: Causes and consequences. *The sport journal*, 22, 1-11. https://www.researchgate.net/profile/Mahmoud-Nourayi/publication/332849257_Strategically_Driven_Rule_Changes_in_NBA_Causes_and_Consequences/links/5ccc9028a6fdccc9dd8b371c/Strategically-Driven-Rule-Changes-in-NBA-Causes-and-Consequences.pdf

Pizarro Milian, R. & Wijesingha, R. (2024) White men can't jump, but do they still get picked first? Race and player selection in the NBA draft, 1980–2021. *Canadian Review of Sociology/Revue canadienne de sociologie*, 61, 172–192. <https://doi.org/10.1111/cars.12471>