Date of submission: 12/9/2023

NBA Draft Picks

Alex Goodman, Statistics, California Polytechnic State University SLO, STAT-414

Michael Grandi, IME, California Polytechnic State University SLO, STAT-414

Ben Fick, IME, California Polytechnic State University SLO, STAT-414

1 Introduction

Each year in the National Basketball Association (NBA) there is a draft held for teams to select players from college, the G-league, or other professional basketball teams outside the league. The purpose of the draft is for teams to be able to bring in young, new talents who can help improve their team and potentially even out the playing field. The NBA draft order is set by how well teams do in the previous season: teams with worse records pick players earlier in the draft, while teams with better records pick players later. However, teams are allowed to trade their draft picks, so some teams may not have any draft picks in a particular year and others may have multiple. As the years have gone on, the amount of teams in the league has expanded and thus, the amount of draft picks in each draft has also increased (Bishop). There are currently 30 teams in the league and 60 total picks per draft (Appendix A).

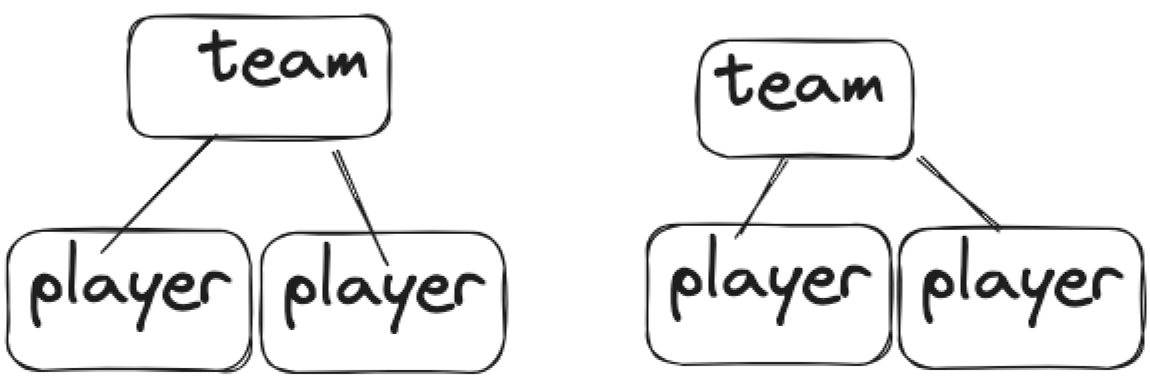
Our overall motivation for this project stems from our collective interest in sports, specifically basketball. The NBA draft happened this summer, and the season just started back up recently, with several different teams picking up generational talents from all over. With the draft being so unique this year, it led us to our overall research question: Do players picked higher in the draft get more playing time than players picked lower in the draft (picked earlier or later overall)? We are curious if this relationship varies by draft class/year, team, conference, and more. For the most part, higher draft picks are more skilled than lower draft picks, hence the reason for them getting drafted earlier. Therefore, we hypothesize that players picked later in the draft (higher draft pick #) will receive less playing time on average than players picked earlier in the draft (lower draft pick #).

2 Data source/Methods

We pulled our data from Kaggle, which is a data science company that provides various datasets. When we first looked at the data, there were several variables that seemed redundant and unnecessary, so we removed them. For example, there were several data metrics such as a player's total points, total assists, total rebounds, and box score plus/minus. We already had a player's point, assist, and rebound averages, so these didn't seem to be very beneficial. Box score plus/minus values and other value metrics are known to be very "unreliable" according to several sources so we dropped this too (Appendix B). The snippet (Appendix C) shows a brief overview of what our final dataset looked like after cleaning.

Our multilevel treatment structure (Figure 1) consists of players (level 1 units) nested in teams (level 2 units). We created this schematic of our multilevel data in Excalidraw, with some earlier drafts being done on paper. We had considered a three level structure at one point, with players nested in years nested in teams, but found that we wouldn't be able to carry out procedures with that structure, because it did not fit a traditional three level hierarchy with our data.. So the one below is the design we carried out instead.

Figure 1: Multilevel Data Schematic



Overall, most of the metrics and variables we kept were simple player statistics, such as their averages in each one of the main three categories (points, rebounds, and assists). We renamed most of these variables to "per-game" to help alleviate confusion. We also created a couple of level 2 variables, such as which conference a team was in (conference, assigned either "East" or "West"), and whether or not the team had moved or will move (change, assigned either a 1 a team that changed, and 0 for a team that did not), since our dataset did not have any.

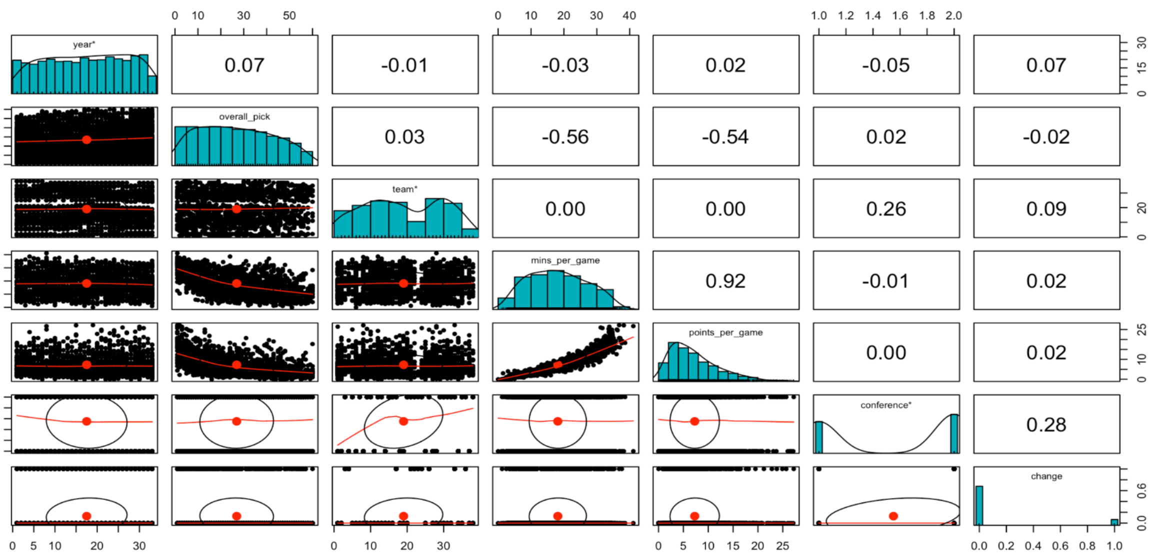
There were also players who were drafted but ended up having no results. This is most likely due to them getting injured in the pre-season and being dropped from the team. As a result of this, we dropped any players who did not have results from our dataset.

We wanted to include the Year variable in our dataset, possibly in a time series plot to see how players' minutes per game changed over time. There is not much year to year variation in playing time, so we did not include this as a fixed effect (likely due to there being different draft picks every year, so it is hard to chart each individual player).

3 Results

Our analysis conveys a few important pieces of information about the player's relationship to the amount of minutes played based on a number of factors. First and foremost, we created a scatterplot matrix (Figure 2) to see if there were any important correlations between variables and what each variable's distribution looked like when compared to other variables.

Figure 2: Scatterplot Matrix of Variables

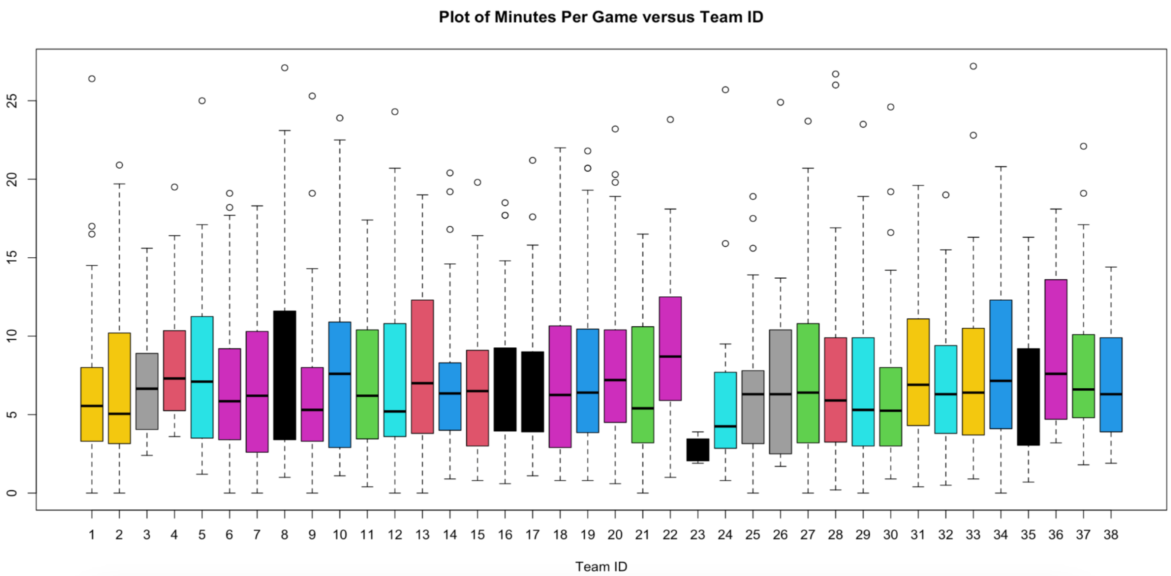


From Figure 2 above, for level 1 variables, minutes per game and points per game have high correlation (0.92), which is trivial, but it's also important to note that points per game and overall pick both have moderately high correlations with minutes per game (-0.54 and -0.56 respectively). For level 2 variables, conference seems to have some correlation with change and team, but not much else (which later proves rather unpromising, unfortunately).

Our null model included random intercepts for the team variable. Figure 3 below shows the boxplots and variabilities of the teams in minutes per game (averaged by team). To be able to plot the minutes per game played by team on average (to investigate the variability in game time by team), we had to convert the team variable into numeric values, so we encoded the 38 teams, in alphabetical order, starting with (1=ATL, 2 = …, 38 = WSB).

Figure 3: Null Model / Boxplot of Minutes Per Game versus Team

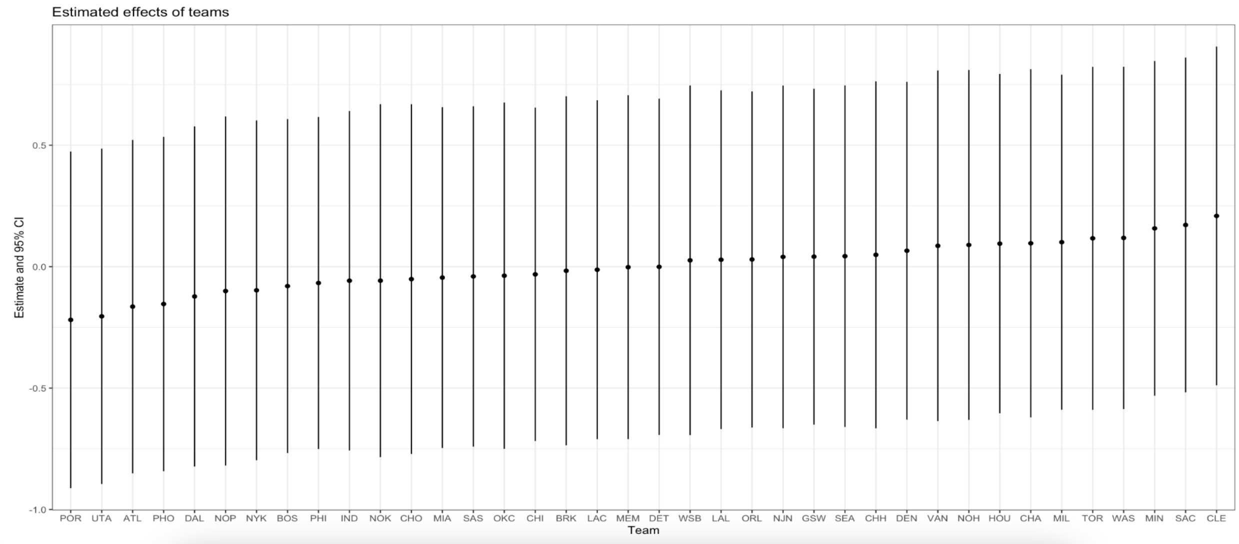
```
model0 <- lmer(mins_per_game ~ 1 + (1|team), data=nba, REML=FALSE)
```



According to the plot, there seems to be a moderate amount of variability in team game time, with lots of outliers. We also found an ICC of 0.002 for our null model, which means there is a low amount of within-team variation. This is somewhat of a good thing, since teams seem to have either a lot of equally matched players in each team, and there are not any “really unfair” picks in a team. Further, we aimed to explain these team-to-team differences in game time.

We investigated a “caterpillar plot” of the effects of each team from the null model (Figure 4). What we see here is that practically every team has about the same magnitude of their effect (with a few differing slightly), but each team has a slightly different point estimate for the center of its confidence interval, meaning that each team has an estimated mean minutes per game played that is different from each other (which makes sense, different teams will play for longer/shorter on average based on the factors we will later include in our models).

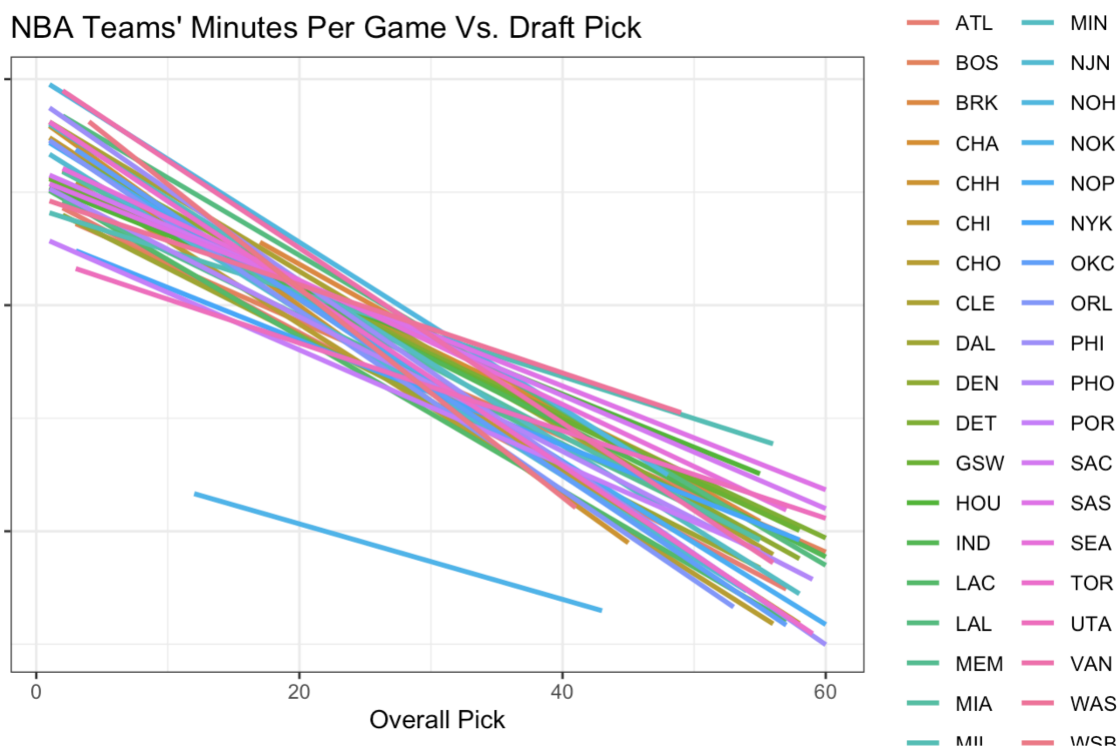
Figure 4: Caterpillar Plot of team effects



After trial and error, we settled on a model that includes random slopes for the draft pick (Figure 5). To see our model building process in more depth between the null model and our final model, see Appendix D. This model, listed below also includes: random intercepts for team, overall draft pick as a level 1 quantitative predictor and random slopes, points per game as a level 1 quantitative predictor, and an interaction between overall pick and points per game.

Figure 5: Final Model / Graph of Random Slopes

```
model7 <- lmer(mins_per_game ~ 1 + overall_pick + points_per_game + overall_pick:points_per_game +
(1 + overall_pick | team), data=nba, REML=TRUE)
```



From this graph, we can see that on average, for an average team, a lower-ranking draft pick (higher number pick) will receive a lower amount of minutes per game on average for that team. For the most part, this is to be expected, and makes sense in context, and almost every team has roughly the same rate of change between overall picks and game time, (with the exception of the New Orleans Pelicans, as seen in the graph). To learn more about this difference, see the Discussion section.

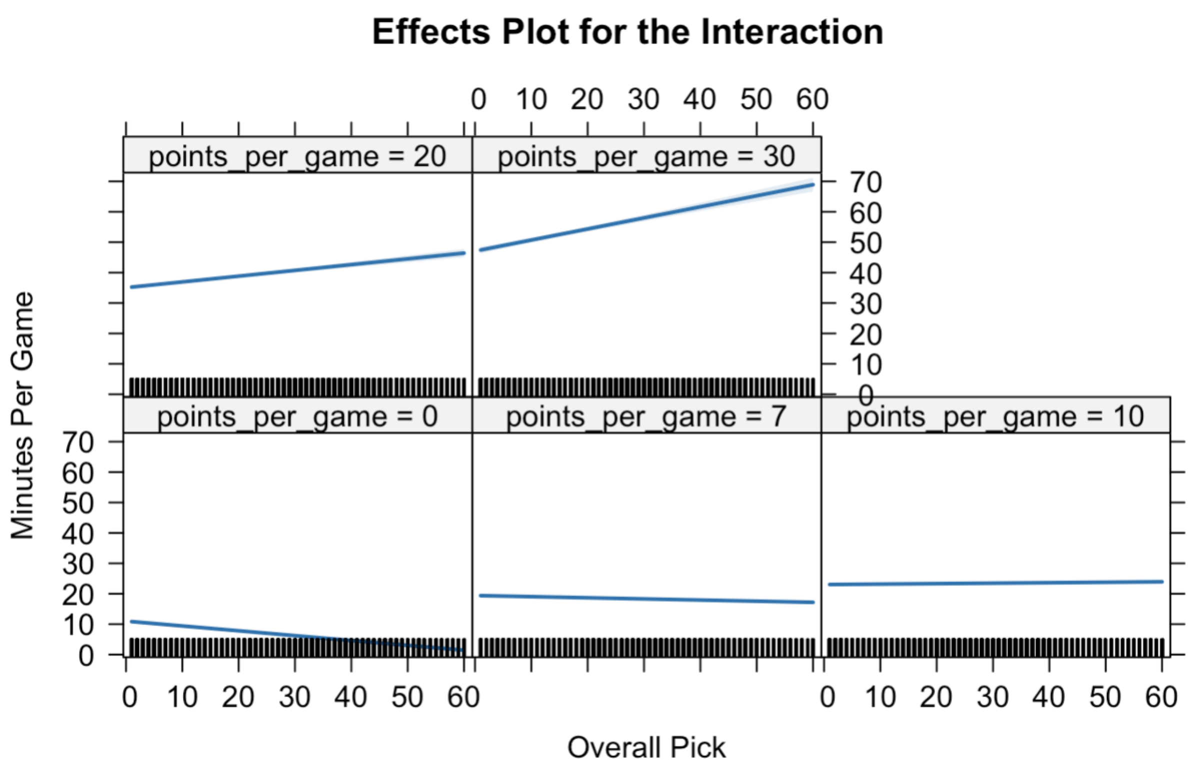
The parameters in our final model can be interpreted in the following figure (Figure 6) with 95% confidence. For the intercept, we are 95% confident that when an average team has a number 1 draft pick, their predicted average minutes per game played for that draft pick is between about 10.39 and 11.61 minutes per game. For overall pick, we are 95% confident that when a team's draft pick number for a player decreases by 1 (e.g. gets larger in value, closer to 60), their predicted average game time tends to fall between about 0.14 and 0.17 minutes per game on average. For points per game, we are 95% confident that each decrease of one in a player's points per game for an average draft pick number on an average team is associated with an increase of between about 1.15 and 1.25 minutes per game on average for a particular team. For our interaction term: we are 95% confident that each time a team's player's draft pick number gets worse by 1 spot changes the amount of minutes per game played for that team's player depending on the amount of points per game that they score by between about 0.015 and 0.019 minutes per game on average. For sigma, we are 95% confident that the amount of unexplained variability in minutes per game played for all teams on average is between about 2.97 and 3.19 minutes per game. For sigma01 (intercept variability, between team variability), we are 95% confident that the average minutes per game played for all teams on average with a #1 draft pick varies by between about 0 and 0.95 minutes per game on average. For sigma02 (overall pick) we are 95% confident that the variability in overall draft picks for an average team is between -1 and 1 (doesn't tell us much information about how accurate the random slopes are, unfortunately). For sigma03 (within team variability), we are 95% confident that the within team variability for an average team is between about 0 and 0.025 minutes per game on average.

Figure 6: Confidence Intervals for Final Model Parameters

```
2.5 %      97.5 %
.sig01                        0.00000000  0.95378300
.sig02                       -1.00000000  1.00000000
.sig03                        0.00000000  0.02528729
.sigma                        2.97836078  3.19013408
(Intercept)                  10.38788986 11.61497970
overall_pick                 -0.17639131 -0.14226690
points_per_game               1.15159125  1.25362730
overall_pick:points_per_game  0.01547222  0.01938520
```

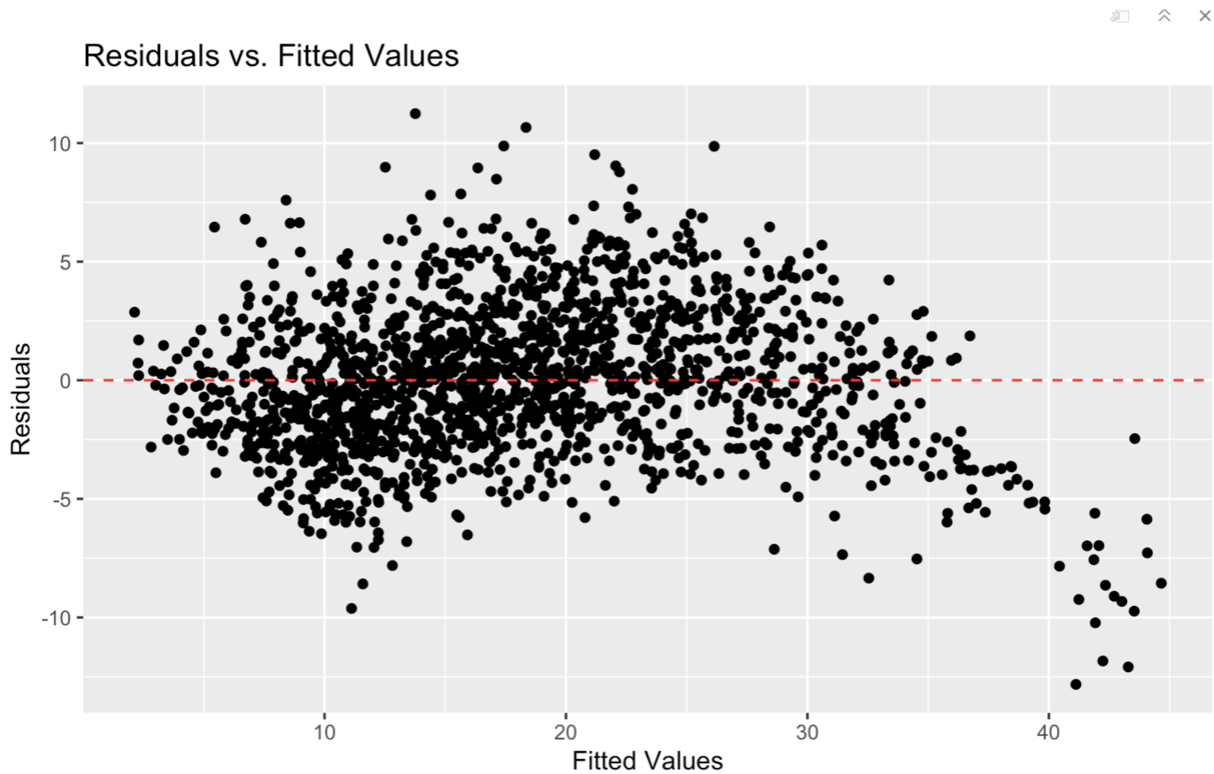
We also decided to plot an effects plot for this final model (Figure 7). We see that for our interaction, the higher points per game a team's draft pick player scores, the more minutes per game they play on average, and if a team's player scores lots of points, they will usually be a lower draft pick more on average. If a team is not doing well, they are more likely to play their higher draft pick players more, and this is evident in that the increase in points per game trend in the effects allows the slope to become more positive, as lower draft pick players will get more game time as the points increases, but higher pick players get more gametime on average with lower amount of points scored.

Figure 7: Effects Plot of Overall Pick * Points Per Game



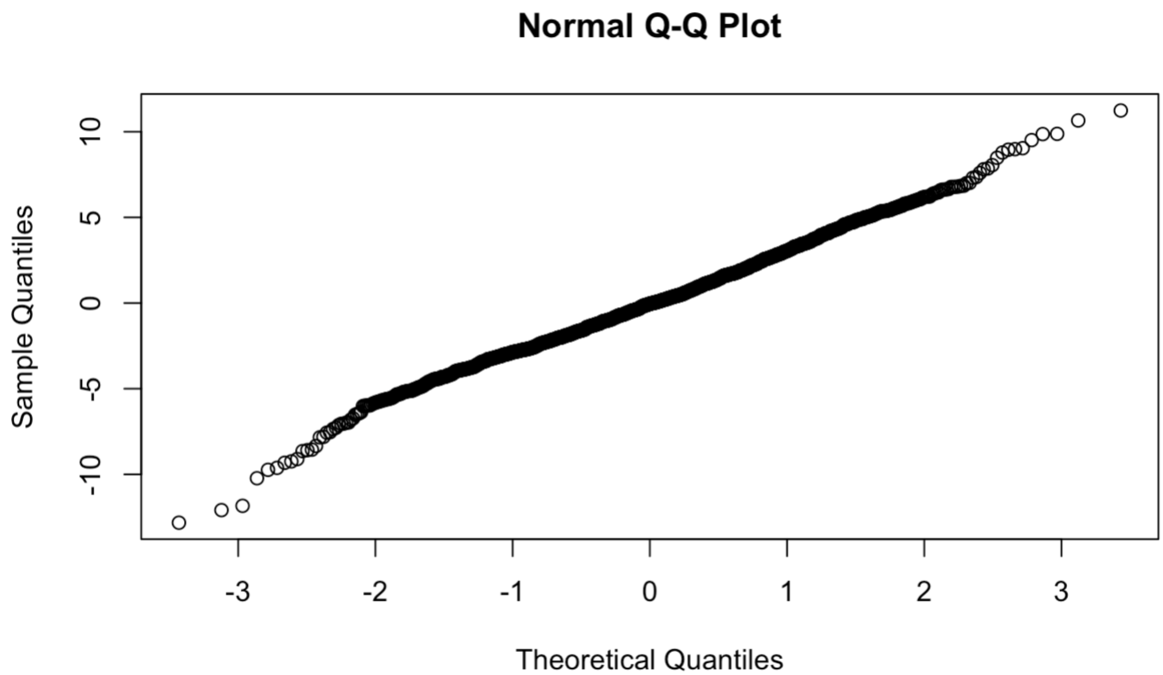
The residual plots for our final model have some interesting implications, particularly our Residuals versus Fitted plot (Figure 8). Our points here follow more of a quadratic than linear trend, which means our linearity assumption is violated. To correct this in the future, we can possibly transform our predictor (points per game, perhaps) into a quadratic term, or change Y (minutes per game) into a logarithmic term.

Figure 8: Residuals vs. Fitted Values Plot for the Final Model



In terms of heteroscedasticity, we don't appear to have unequal variances, since our residuals versus fitted plot does not tend to fan outward or inward strongly. The normality assumption for our final model also appears to be met (Figure 9), with a roughly linear trend of the quantiles.

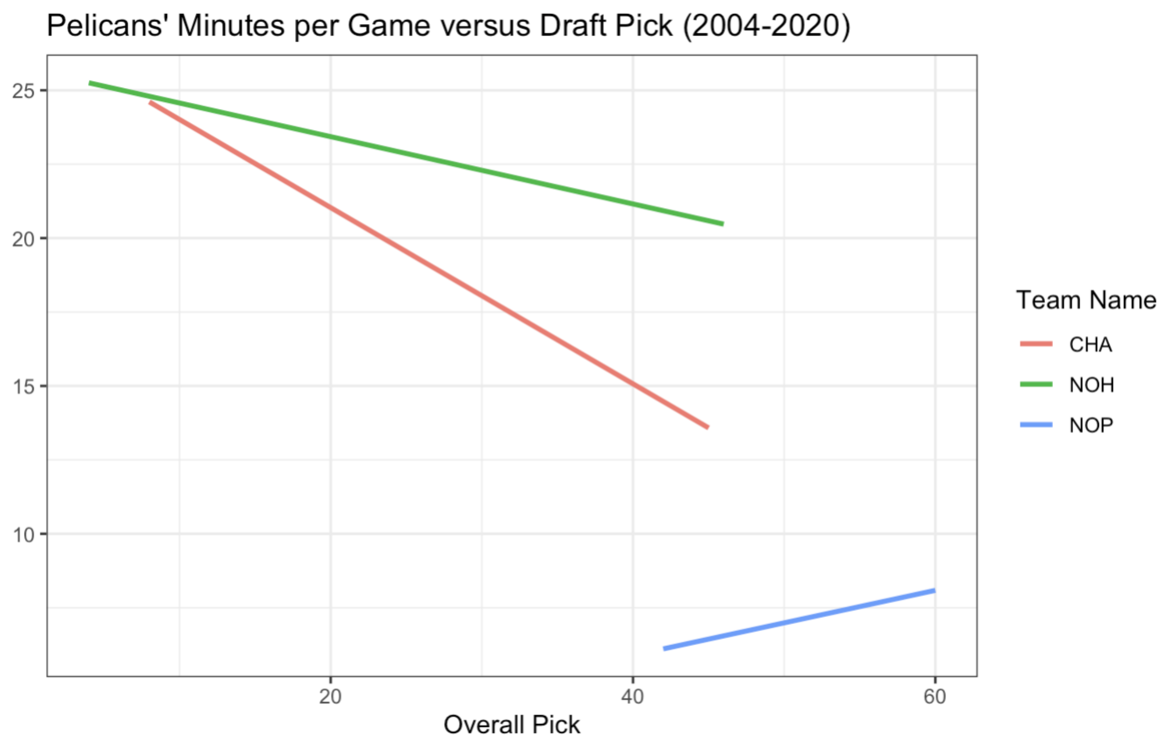
Figure 9: Q-Q Plot for the Final Model



4 Discussion

In this report we have explored the differences in game time that a draftee gets based upon the overall pick a player was in their respective draft and what team drafted them, and as some may have expected, the earlier that a player was drafted (e.g. a lower draft pick # like #1), the more playing time we would expect that draftee to get. When looking deeper into the teams that have moved from one location to another and changed their name, we discovered that the only team to give their later overall pick draftees more playing time is the New Orleans Pelicans (Figure 10).

Figure 10: Interaction plot of the Pelicans Franchise Minutes Per Game vs. Draft Pick



The limitations to our study include the removal of certain data points from our dataset as well as the exclusion of some potential confounding variables. We removed any draftees that didn't end up playing any game minutes, however this means we didn't fully assess the entire target population. With this in mind, we sampled the players of the target population that got game time thoroughly by sampling each data point within the population. Regarding confounding variables, we didn't include how long a player ended up playing in the NBA. By including this variable, we may have been able to see if certain players that were drafted earlier played for a longer period of time which would tend to result in scoring more points. To see more of the teams that got changed over time, see Appendix E.

The strengths of our analysis include being able to greatly reduce the variability within the residuals as well as identifying why certain teams are potentially outliers. We were able to reduce our residual variability from 75.64 to 9.51 (Appendix F). Our analysis is also able to attribute teams having a smaller amount of data points within the dataset to being a team that had its name changed/moved to a new location. One of the weaknesses of our analysis that was touched on earlier is the fact that we removed players that didn't end up playing any game minutes. Future research could build on our work by implementing this variable to evaluate if certain teams "waste" some of their draft picks on players that don't end up playing. This could even be taken one step further by including a new categorical variable that looks to explain why a certain draftee didn't end up playing. By applying these variables to the teams in the NBA, the new research could identify the weaknesses in a team's drafting technique.

Appendix

A: Bishop, Ethan. "A history of expansion teams and their impact on the NBA." *Sportskeeda*, 9 May 2023, https://www.sportskeeda.com/basketball/a-history-expansion-teams-impact-nba. Accessed 3 December 2023.

B: "Learn a Stat: Box Plus Minus and VORP." *Hack a Stat*, 29 February 2020, https://hackastat.eu/en/learn-a-stat-box-plus-minus-and-vorp/. Accessed 8 December 2023.

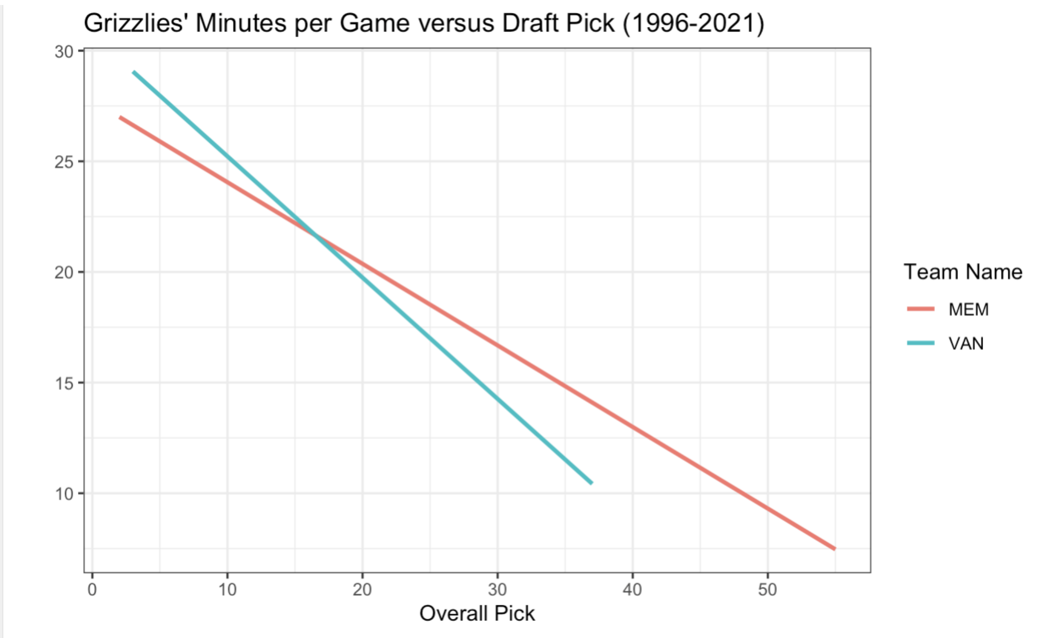
C: Here is the head of our dataset.

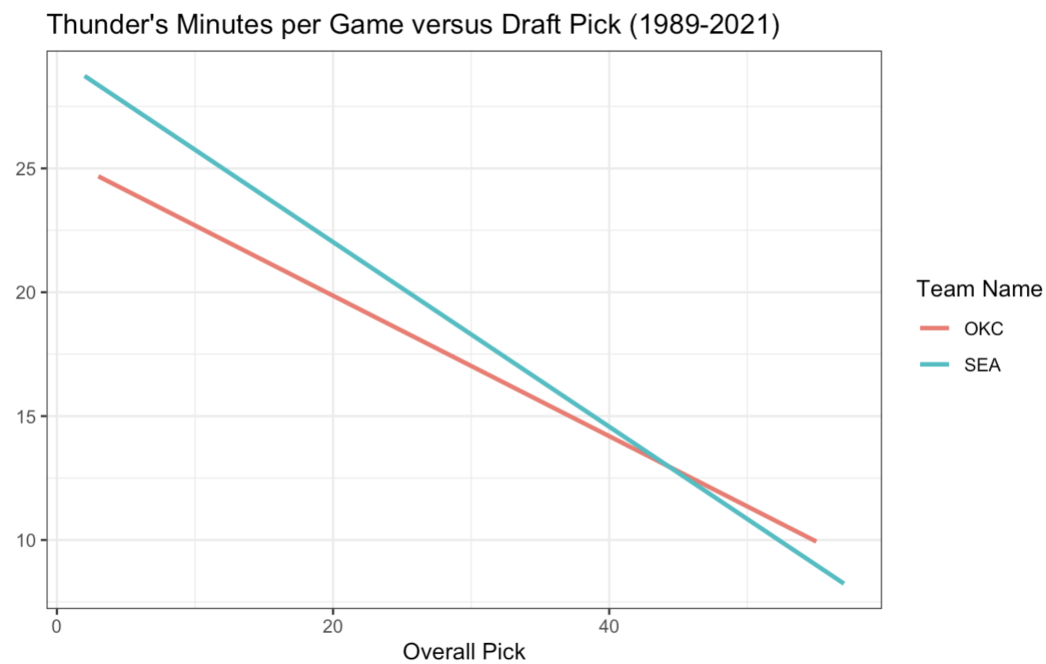
year	overall_pick	team	player	total_games	fg_percentage	3pt_percentage	ft_percentage	mins_per_game	points_per_game	rebounds_per_game	assists_per_game	conference	change
1989	1	SAC	Pervis Ellison	474	0.51	0.05	0.689	24.5	9.5	6.7	1.5	West	0
1989	2	LAC	Danny Ferry	917	0.446	0.393	0.84	19.8	7	2.8	1.3	West	0
1989	3	SAS	Sean Elliott	742	0.465	0.375	0.799	33	14.2	4.3	2.6	West	0
1989	4	MIA	Glen Rice	1000	0.456	0.4	0.846	35	18.3	4.4	2.1	East	0
1989	5	CHH	J.R. Reid	672	0.472	0.135	0.716	22.9	8.5	5	1	East	1
1989	6	CHI	Stacey King	438	0.478	0.235	0.707	16.9	6.4	3.3	0.9	East	0

D: Here was our model building process (summarized in Excalidraw)

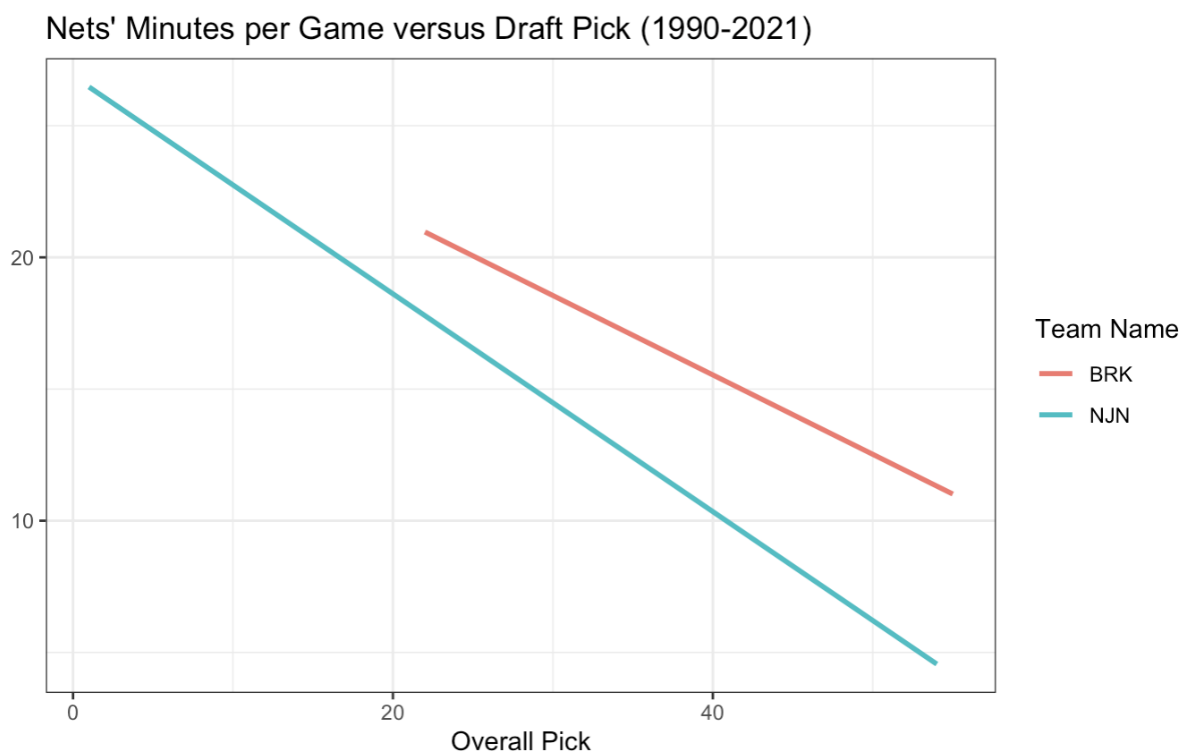
Model 0	random intercepts for teams
Model 1	adding overall pick (fixed)
Model 2	adding conference (fixed)
Model 3	random slopes of overall pick
Model 4	adding points per game (fixed, and interaction (overall_pick:points_per_game)
Model 4b	dropped conference variable, dropped interaction
Model 4c	re-added interaction
Model 5	added change (fixed)
Model 6	added points_per_game as random slopes
Model 7 (final)	decided that model4c was the best one we had from likelihood ratio tests
Model 8 (1 last try)	added interaction (conference:overall_pick) (almost significant, but not quite)

E: These are three other teams/franchises we looked at across the years to chart their changes. For context, the Vancouver Grizzlies (VAN) moved to Memphis (MEM) Grizzlies, Charlotte Hornets (CHA) moved to New Orleans Hornets (NOH) moved to Oklahoma City (NOK) Hornets moved back to New Orleans (NOH) Hornets (again) moved to New Orleans Pelicans (NOP), Seattle (SEA) Supersonics moved to Oklahoma City (OKC) Thunder, andNew Jersey (NJN) Nets moved to Brooklyn (BRK) Nets.





Thunder's Minutes per Game versus Draft Pick (1989-2021)
25
20
15
10
0
20
40
Overall Pick
Team Name
OKC
SEA



Nets' Minutes per Game versus Draft Pick (1990-2021)
20
10
0
20
40
Overall Pick
Team Name
BRK
NJN

F: Below are our random effects for the null model (model 0) and our final model (model7). We now see that the residual error went down from 75.6 to about 9.5 from the null model to the final model, so we did a good job of explaining variation in teams.

```
Random effects:
 Groups   Name        Variance Std.De
 team     (Intercept)  0.1384  0.372
 Residual             75.6395  8.697
Number of obs: 1669, groups:  team, 3
```

```
Random effects:
 Groups   Name         Variance  Std.Dev. Corr
 team     (Intercept)  0.2257739 0.47516
          overall_pick 0.0001171 0.01082  -1.00
 Residual              9.5098946 3.08381
Number of obs: 1669, groups:  team, 38
```