Player/Team Shot Selection Composition Impact On NBA Offenses

Members: Shaikh Noman, Mason Reed, Kabir Chaudry, Kapil Taspa

Research Question

How has the distribution and efficiency of NBA shot selection evolved across eras, and what do player clustering patterns reveal about the most optimal team compositions in each era?

Objective

Our objective is to examine the evolution of NBA shot selection across different eras (1996–2024), focusing on how player tendencies and team strategies align or diverge. We aim to analyze how individual shot selection bias (i.e., how much a player's shot profile deviates from their team's average) affects offensive efficiency. This project uses clustering to identify distinct player types such as paint-focused slashers, midrange specialists, and three-point shooters and investigates how the prevalence of these types has changed over time. We also use linear regression to explore whether certain combinations of these player types within teams are more significant to offensive success, as measured by offensive scoring efficiency ratings.

Motivation and Relevance

The goal of this research is to explore the evolving dynamics of basketball strategy. Players bring distinct shot tendencies. Some excel at finishing near the rim, others thrive in the midrange, and many are perimeter threats. At the same time, team strategies (like the "Moreyball" model) increasingly emphasize efficiency zones: three-pointers and shots at the rim. By studying the relationship between player shot bias and team tendencies, we can better understand how fit impacts offensive performance. Additionally, this project highlights how shot values have shifted. For example, from the midrange-heavy 1990s and early 2000s to today's three-point-dominant era. Through regression and clustering techniques, we assess which player types and team compositions have been most optimal in different decades.

Data Collection/Preparation

Access: We will use the hoopR package that gets its data from the [NBA Stats API.](NBA Stats API) The package makes it easy to specify the year for the data, and that will be used to access the different data sets. We will use the 1996-2024 seasons for this analysis.

Player Shot Data

The player shot dataset contains 402,395 rows and 24 columns, covering seasons from 1996 to 2024. Key variables include shot attempts, makes, and field goal percentages from various court locations, such as the restricted area, mid-range, and multiple three-point zones (e.g., corner threes). The purpose of this data set is to analyze individual players' shot selection patterns and efficiency.

Team Shot Data

The team-level dataset includes 402,395 rows and 24 columns, spanning the 1996 to 2024 seasons. It captures shot attempts, makes, and field goal percentages from the same court locations as the player dataset, with the goal of identifying systematic shot selection biases at the team level.

Offensive Team Statistics Data

The offensive team statistics dataset contains 30 rows and 184 columns for each year analyzed (2006 and 2016). Since data for 1996 was unavailable, a custom dataframe was created using offensive scoring efficiency metrics sourced from Basketball Reference, a site that pulls from the NBA's official statistical archives. The key variable of interest is offensive scoring efficiency, used in the regression analysis.

Data Cleaning/Preprocessing

Shot location bias was calculated by joining player and team data and by subtracting team-level shot frequencies from player-level frequencies for each court area (e.g., restricted area, corner three). All variables were then standardized using z-scores before applying k-means clustering. There were no unique NA values in the data.

Exploratory Data Analysis

Summary Statistics for Shot Attempts

To understand how the NBA offenses have progressed, it is important to analyze the distribution of shot attempts over time. When we graph the summary statistics like the mean and median shot frequencies of the 1996, 2006, and 2016 seasons, we can better appreciate significant developments in where players are taking shots and how these trends relate to shifting strategies. This allows us to examine player trends as well as comprehend how team makeups have been changing over time.

The 1996 season shot attempt data show a tendency towards shots near the rim. The Restricted Area averaged 108 attempts, and a higher average of 160, since there were continuous attempts in close range to the rim. Mid-range shots also saw high attempts, with an average of 171.1 and median of 108. Three-pointers were quite rare, with the Left and Right Corner 3 averaging around 6 and the Above Break 3 around 8 and 32.6. Overall, shots closer to the basket were in favor, and few three-pointers were tried.

Shot Attempts	**Median**	**Mean**
Restricted Area	108	160
Paint (Non-Restricted Area)	26	50.74
Mid-Range	108	171.1
Left Corner 3	1	7.152
Right Corner 3	1	6.312
Above Break 3	8	32.6

In 2006, also, players favored shots in the area around the basket, with the Restricted Area averaging 109.5 and a mean of 140.8 attempts. Mid-range shots were still common (median 86, mean 141.7), but down a bit from 1996. Most notably, three-point attempts increased across zones: the Above Break 3 increased to a median of 17.5 and mean of 64.46, and corner threes had medians of 3–3.5 and averages of about 13. This shows drift toward more perimeter shooting than a decade earlier.

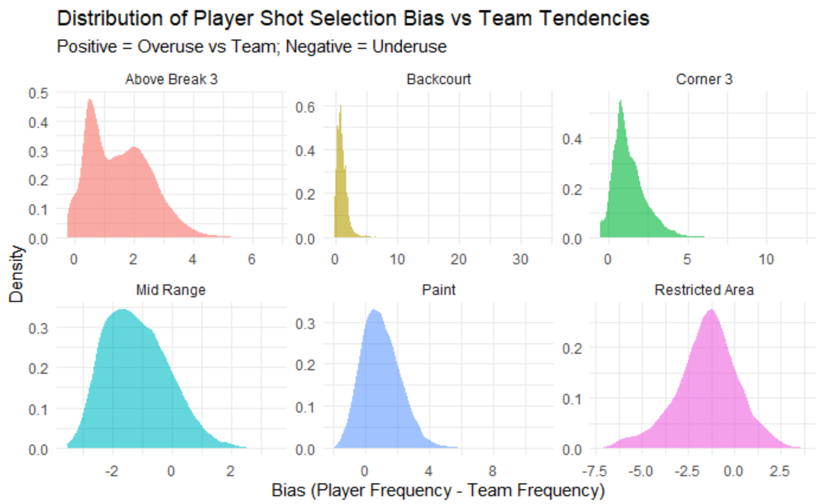
Shot Attempts	Median	Mean
Restricted Area	109.5	140.8
Paint (Non-Restricted Area)	12	22.78
Mid-Range	86	141.7
Left Corner 3	3	13.93
Right Corner 3	3.50	12.03
Above Break 3	17.50	64.46

By 2016, shot location had shifted markedly towards the three-point line. While the Restricted Area still suffered frequent attempts (median 100, mean 138.7), mid-range attempts fell (median 55, mean 96.03). Faintly higher Paint (Non-RA) attempts were attempted. Most spectacularly, three-point shooting rose. Over Break 3 attempts grew to median 59 and average 103, while corner threes both sported medians of 10 and averages around 16. This is a stark league-wide emphasis on perimeter shooting.

Shot Attempts	Median	Mean
Restricted Area	100	138.7
Paint (Non-Restricted Area)	38	61.49
Mid-Range	55	96.03
Left Corner 3	10	16.77
Right Corner 3	10	15.92
Above Break 3	59	103

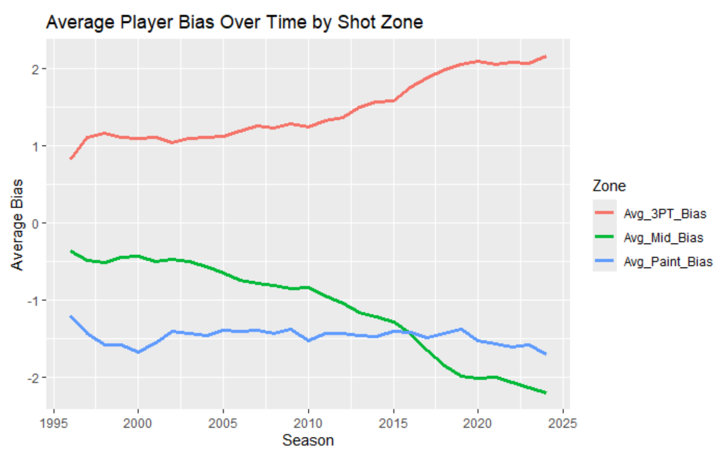
Shot Selection Bias vs. Team Tendency

The plot shows player shot selection bias compared to team tendencies from 1996 through 2024. Players consistently show negative bias in the Restricted Area and Mid-Range, reflecting underuse compared to team averages, perhaps because of a league-wide decline in these types of shots. Three-point shots show a positive trend, with players using them slightly more than their teams. Shots in the Paint show slight positive bias, reflecting that players follow closely with team tendencies in this region.



Player Bias Over TIme

The average player bias line plot, 1996 to 2024, testifies to the widely documented trends. Mid-range shot bias reduced steadily, especially post-2016, showing the modern abandonment of such shot attempts. Three-point bias increased steadily, corresponding with the rising league role of perimeter offense. Paint shot bias fluctuated somewhat similarly, representing steady usage vis-a-vis team preferences.



This EDA provides context for understanding the manner in which NBA shot selection has evolved over time and directly informs the development of the analysis plan. When we examine shifts in distributions of shot attempts across the period between 1996 and 2016, we observe a clear shift away from mid-range and paint-based offenses towards perimeter-based styles, with the most pronounced being the proliferation of three-point shooting. The analysis of shot selection bias (comparing individual player tendencies to team averages) also illustrates how player behavior conforms to or deviates from emerging team strategies. These results not only tie back to the research question of how shot distribution and efficiency have changed across eras and what clustering tendencies reveal about optimal team composition, but also imply the importance of player clustering and efficiency modeling as a means of evaluating the strategic composition of teams over time.

Methodology

Preprocessing

The first step after loading in the required data is to create a data frame that includes the 1996 NBA Season team-level data from Basketball Reference because HoopR did not have this data available. Next, we calculate the standardized field goal attempt frequencies across different players to ensure that the data could be compared on an equal footing. Field goal attempts (FGA) were categorized by shot locations, including the Restricted Area, Mid-Range, and Three-Point zones, among others. Finally, we merged the team and player data and calculated the player biases to measure how each player's shot selection aligned with their team's overall shot preferences. For K-Means clustering, the data was scaled using the Z-Score method.

Clustering

Using K-Means clustering, we grouped players according to the frequency with which they take shots from the key areas on the court. K-Means works by partitioning players into groups with similar shot profiles where one cluster might consist of players who predominantly shoot near the basket, while another cluster could represent players who are more focused on 3-point shots.

Optimal Number of Clusters (Elbow Method)

When looking at the elbow graph the x-axis represents the number of clusters, and the y-axis shows the Total Within Sum of Squares (WSS), which measures the variance within each cluster. The "elbow" point in the graph is where the curve flattens, and it suggests the optimal number of clusters. This comes into play when we do a multi season analysis because the clusters change over the years.

Regression Analysis

The next step in our analysis was to examine the relationship between the composition of player types on a team and their offensive efficiency, essentially computing how the player contributes to the overall team success. To do this, we used regression analysis to model the offensive efficiency of each team as a function of the number of players in each cluster. We aimed to identify whether the proportion of specific player types within a team could predict its success in terms of offensive efficiency scores. In the end, we did a multi-season analysis by running these regressions on 3 different seasons: 1996, 2006, and 2016. We looked for differences in the clustering and found insights through these.

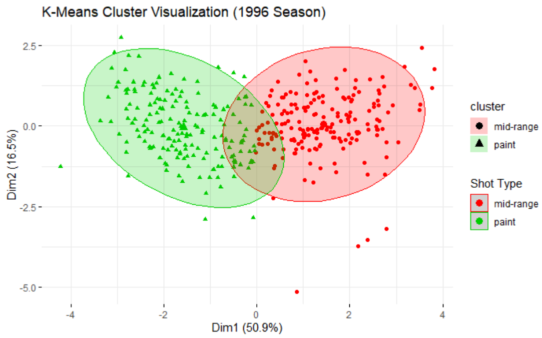
Estimating Best Team Composition

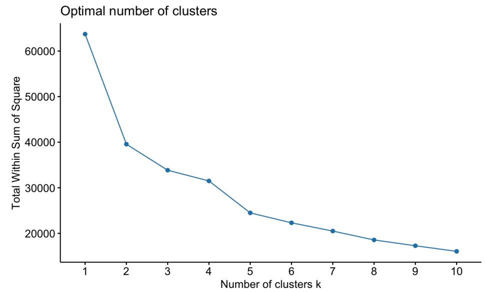
Lastly, we estimated the optimal team composition by minimizing a loss function that accounts for the proportions of shooters and inside players relative to each team's offensive efficiency. The loss function calculates the squared differences between actual player proportions and the desired proportions, weighted by offensive efficiency. This method allowed us to determine the ideal balance of shooters, inside players, and versatile players, with the 2023-24 season

Results

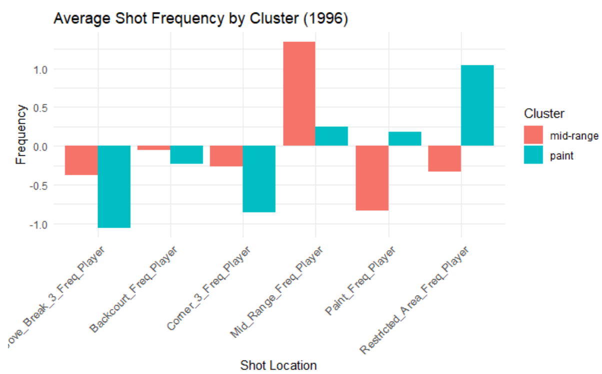
1996

The elbow plot tells us that the two clusters provide the best fit for shot choice by players in this era, classifying paint and mid-range players as the dominant clusters. Both clustering and elbow plot establish the rare incidence of three-point specialists, thus giving us our first sense of the era's low reliance on 3-point shooting.

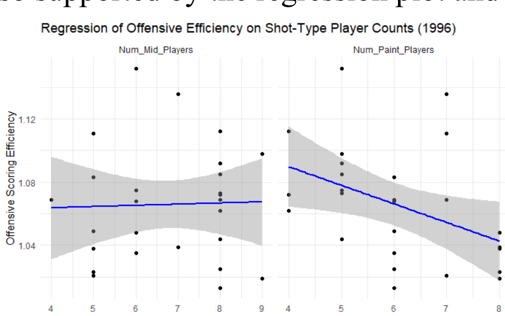




The bar plot reveals average shot rate by cluster for 1996, showing differences in preference by shot between paint and mid-range players. Surprisingly, all three-point shot categories like above the break and corner threes also have negative frequency values in both clusters, indicating low usage relative to other places. This indicates that teams and players strongly favored mid-range and paint areas at that time. Mid-range players positively dominate mid-range areas strongly, while paint players tend to operate closer to the Restricted Area and Paint (Non-RA). The narrative can quite readily depict a pre-modern style of NBA play in which mid-range and interior scoring ruled and three-point shooting was reduced.



Regression analysis of efficiency in scoring demonstrates that there exists a statistically significant negative correlation between roster paint player presence and efficiency on offense, or that heavy reliance on players focused on the paint may have had low efficacy at scoring over this time period. This finding aligns with the popularity of mid-range-oriented players like Michael Jordan in 1996. Indeed, while mid-range shots are trendy, there is no strong correlation between the existence of mid-range players and offense efficiency, so while mid-range shots were trendy, they may not have been the reasons for team scoring being greater. These findings are also supported by the regression plot and summary table below.

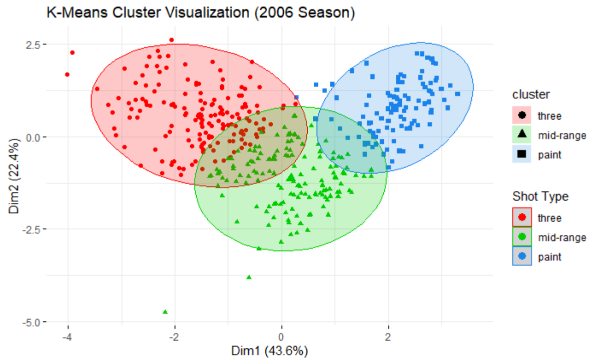


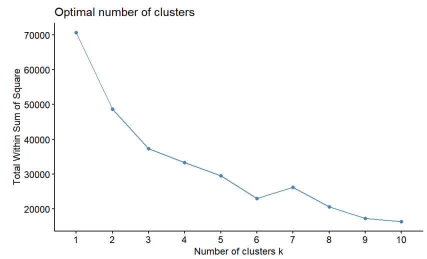

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.046796 -0.026016 -0.001939  0.010251  0.085467

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       1.190144   0.060322  19.730  6.5e-16 ***
Num_Paint_Players -0.014604   0.005808  -2.515   0.0194 *
Num_Mid_Players   -0.005341   0.005151  -1.037   0.3106
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0336 on 23 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.2164,    Adjusted R-squared:  0.1482
F-statistic: 3.175 on 2 and 23 DF,  p-value: 0.06059
```

2006

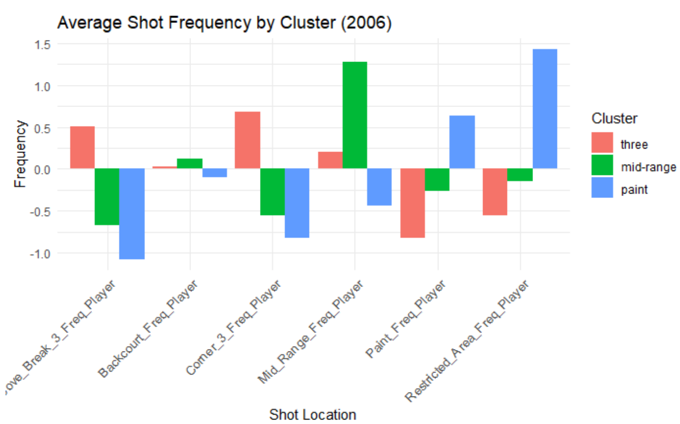
New patterns emerged as the cluster analysis currently favors three groups: paint, mid-range, and 3PT players. This is different from 1996, with more frequent three-point attempts highly evident.

The more positive trend towards 3PT shooting translates that teams and players began to value the three-point shot, representing the early phases of the modern perimeter brand of play. This shift in shot choice reflects an intensification of emphasis on spacing and efficiency.

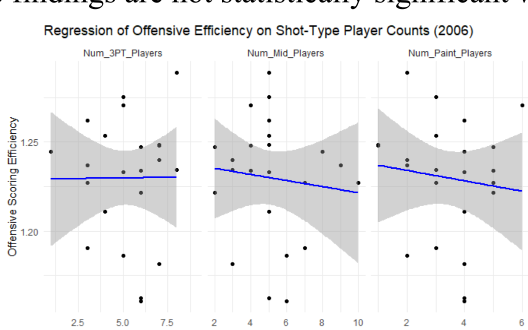




The 2006 bar chart shows a strong reversal of shot choice patterns from 1996. All three-point areas now have positive frequencies, with corner threes, especially by three-point specialists having the highest relative frequency of usage. This reflects the growing strategic value placed on the three-point shot during this period. While mid-range shots still enjoy a positive frequency, implying that they were still a staple of the offense, their relative supremacy has started to fade. Paint players, in particular, are showing a shift in preference in the paint itself, shooting more from the non-restricted area rather than directly at the rim. This may be a sign of more focus on floaters, hook shots, or short pull-ups rather than traditional post-ups. In general, this tale prioritizes a more diverse and dynamic attacking style as the NBA begins to enter a more perimeter-based era.



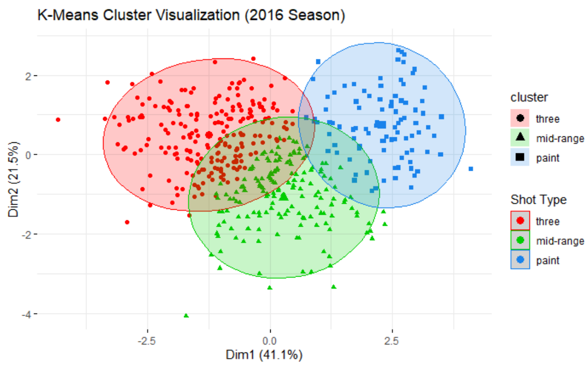
The regression analysis shows a weak positive trend for 3-point players and a weak downward trend for mid-range and paint players. However, as can be observed from the summary plot, these findings are not statistically significant with p-values being weak evidence of these trends.

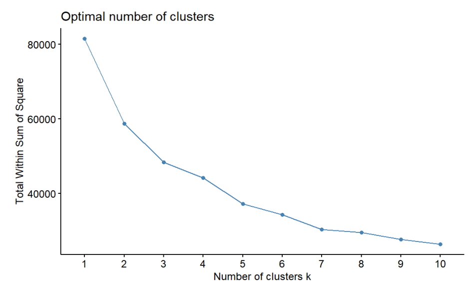



```
Residuals:
     Min       1Q   Median       3Q      Max
-0.05770 -0.01838  0.00840  0.01325  0.06737

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.331344   0.084036  15.843  2.1e-12 ***
Num_Paint_Players -0.007806   0.007100  -1.099    0.285
Num_Mid_Players   -0.007466   0.006467  -1.155    0.263
Num_3PT_Players   -0.007108   0.007148  -0.994    0.333
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03533 on 19 degrees of freedom
  (7 observations deleted due to missingness)
Multiple R-squared:  0.07821,	Adjusted R-squared:  -0.06734
F-statistic: 0.5373 on 3 and 19 DF,  p-value: 0.6624
```

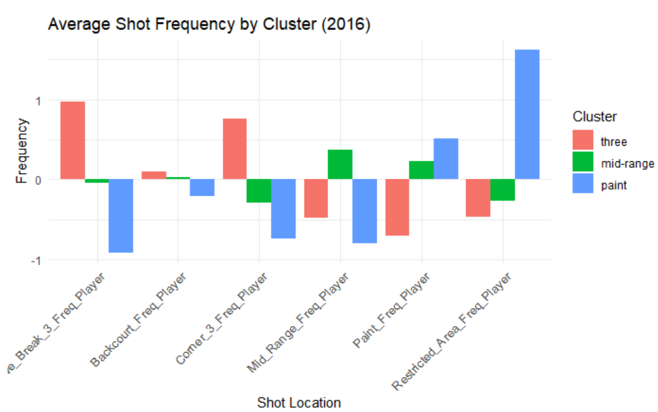
2016

Although there are still three clusters, there is a greater representation of 3-point players. As years increase in 20 years, the trend might revert to two clusters, which would most likely be 3-point and paint players.

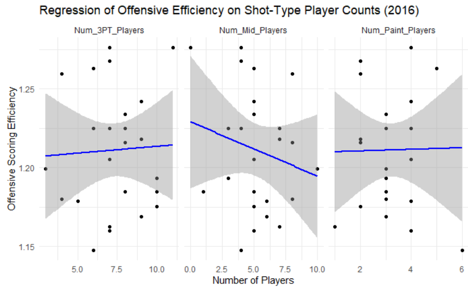




The next bar plot builds on the 2006 trends, demonstrating the continued evolution in shot type by player cluster for 2016. Perhaps most noticeable is the considerable increase in three-point and paint shot frequencies. Three-point shooters are recording higher frequencies in areas like the corner and outside the break, while paint players are once again recording high frequencies in the restricted area, foreshadowing a return to high-efficiency interior scoring. Mid-range players, on the other hand, have experienced a sharp decline in shot frequency in nearly all areas, most significantly in the mid-range and baseline areas. This reflects the league-wide pattern of optimizing shot selection for efficiency, with an emphasis on three-pointers and shots at the rim.



According to the previous regression analysis, there is no statistically significant trend in types of players. However, there is an apparent downward trend among mid-range players when one considers offensive scoring efficiency. What this means is that, with the modern age, teams gain an edge when they cut on their reliance on mid-range specialists because their shooting profile is not as productive to efficient scoring compared to three-point or paint players.




```
Residuals:
      Min        1Q    Median        3Q       Max
-0.063994 -0.029137 -0.000105  0.028429  0.058647

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.307544   0.094956  13.770 1.15e-11 ***
Num_Paint_Players -0.005664   0.009230  -0.614    0.546
Num_Mid_Players   -0.007606   0.006138  -1.239    0.230
Num_3PT_Players   -0.005274   0.006468  -0.815    0.424
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04016 on 20 degrees of freedom
  (6 observations deleted due to missingness)
Multiple R-squared:  0.07367,   Adjusted R-squared:  -0.06528
F-statistic: 0.5302 on 3 and 20 DF,  p-value: 0.6667
```

Estimating Team Composition

To estimate the optimal team structure for the 2023-24 season, we minimize a loss function based on the ratios of shooters and inside players. The optimal team structure is estimated to be 33% shooters, 22% inside players, and 44% versatile players. The loss function, $L(P_{shooters}, P_{inside}) = \Sigma E_i(P_i - P_{shooter})^2 + \Sigma E_i(P_{ii} - P_{inside})^2$, reduces the differences between actual player ratios and desired ratios, weighted by offense efficiency. The results are that with good inside players, teams need to emphasize outside shooters. The ratios are close to average but with a slight tilt toward more shooters and do-it-all players to optimize offense efficiency.

Discussion and Conclusion

Overall, while our regression estimates are not statistically significant due to the small sample size (30 teams per season), the estimates persistently validate that offensively efficient teams excel at scoring from all areas of the court, most notably from outside. Negative trends for mid-range and paint men, and positive trends for shooters even as far back as 1996 validate that shooting ability is a core component of offense. These trends reflect growing league-wide worries regarding spacing and perimeter threats for disrupting defenses.

To enhance the strength of the results, subsequent research could include incorporating playing time into the sample, aggregating data by season, using more advanced clustering statistics, and standardizing shot sample cut-offs. Additional viewpoints could involve examining team homogeneity in calling shots and evaluating the effect of roster construction on offensive and defensive productivity with more comprehensive team statistics.

In general, since teams have gotten more savvy to the value of the 3-point shot, solid offenses are now characterized by great perimeters and the ability to score from a variety of locations. The best team architectures appear to favor the multi-skilled shooter over the paint resident, confirming modern coaching methods based on spacing and offense versatility.