

A Joint Analysis for Field Goal Attempts and Percentages of Professional Basketball Players: Bayesian Nonparametric Resource

ELIOT WONG-TOI¹, HOU-CHENG YANG², WEINING SHEN¹, AND GUANYU HU^{3,*}

¹*Department of Statistics, University of California Irvine, USA*

²*Department of Statistics, Florida State University, USA*

³*Department of Statistics, University of Missouri, USA*

Abstract

Understanding shooting patterns among different players is a fundamental problem in basketball game analyses. In this paper, we quantify the shooting pattern via the field goal attempts and percentages over twelve non-overlapping regions around the front court. A joint Bayesian nonparametric mixture model is developed to find latent clusters of players based on their shooting patterns. We apply our proposed model to learn the heterogeneity among selected players from the National Basketball Association (NBA) games over the 2018–2019 regular season and 2019–2020 bubble season. Thirteen clusters are identified for 2018–2019 regular season and seven clusters are identified for 2019–2020 bubble season. We further examine the shooting patterns of players in these clusters and discuss their relation to players' other available information. The results shed new insights on the effect of NBA COVID bubble and may provide useful guidance for player's shot selection and team's in-game and recruiting strategy planning.

Keywords *Chinese restaurant process; mixture model; shot charts data; spatial spline; sport analytics*

1 Introduction

The National Basketball Association (NBA) is undergoing a massive revolution partly due to the recent advances in data science. Data analytics has attracted a great deal of attentions in many aspects of NBA over the years, including player drafting and evaluation, in-game strategy planning, team and player match-ups, and adaption of the three-point shot. An important question of interest in basketball analytics is to find effective ways to quantify and understand individual player's shooting patterns, which often can be related to analyzing the "hot spots" of players, that is, where the most field goals have been attempted and made over the court. *Shot chart* is a widely used tool for this purpose because it provides an easy-to-understand graphical representation of shot locations and the associated field goal attempts over different locations. In the literature, several spatial models and machine learning methods have been proposed to study players' shot charts. For example, Miller et al. (2014) used spatial point processes to account for the randomness nature of shot locations. Franks et al. (2015) characterized shot attempt locations by combining spatial and spatio-temporal processes, matrix factorization methods and hierarchical models. Jiao et al. (2021) proposed a joint model for simultaneously making inference on random loca-

*Corresponding author. Email: guanyu.hu@missouri.edu.

tions and outcomes of field goal attempts. Hu et al. (2021a) proposed a distance-based clustering approach to model the heterogeneity of shot selection based on the log Gaussian Cox process. Hu et al. (2021b) considered a model-based clustering approach and used a Bayesian zero-inflated Poisson regression to handle large proportion of zero shot attempts over the court. Yin et al. (2022a) proposed a Bayesian mixture model of matrix normal distributions to study the spatial heterogeneity in the shot charts. Yin et al. (2022b) proposed a novel nonparametric Bayesian method for learning the underlying intensity surface of shot charts for selected NBA players.

Compared to fields goals attempted, the *field goal percentage*, which is defined as the ratio between goals made and attempted, seems to have received less attention in the literature, and it remains an important quantity for investigation. For example, identifying players that share similar “hot spots” and field goal percentages can provide valuable information for coaches and teams that aim to find and trade specific types of offensive players. Defensive strategies can also be planned accordingly based on field goal percentages. Reich et al. (2006) developed a spatially varying coefficients model for shot chart data analysis, where the court is divided into multiple, small, non-overlapping regions, and the field goal percentage in these regions were fitted by a multinomial logit model. However, that paper ignored the player-level variability in their analysis.

Our focus in this paper is to simultaneously model the field goal attempts and shot percentages over different shot locations while taking account for the heterogeneity among players. In other words, players are clustered based on the locations in the basketball court where they shoot most often and how accurately they shoot from there. This information could be useful for roster construction for NBA general managers or avid fantasy basketball fans. To model both the field goal attempts and shot percentages at the same time, a joint modeling framework is adopted. Despite its popularity in statistics literature including environmental health (Xu et al., 2019), and longitudinal and survival data analysis (Ibrahim et al., 2010; Long and Mills, 2018), joint modelling has not been well explored in basketball analytics. The only work that we are aware of is Jiao et al. (2021), where a joint model was first used for shot chart data analysis and then a hierarchical clustering was performed based on the obtained joint model parameter estimates. However, the inherent uncertainty in the estimation of the cluster number was ignored in that paper.

Unlike the aforementioned clustering approaches, we propose a Bayesian nonparametric mixture model under the *joint modeling* framework to cluster NBA players based on their shooting patterns and goal percentages. Bayesian models such as the Dirichlet process (DP; Ferguson (1973)) and its mixture are well known for their flexibility in providing a natural way to simultaneously estimate the number of groups and the group configurations in clustering problems. Moreover, the joint modeling framework allows us to conveniently learn the player shooting patterns by taking account for the dependence between the shot attempts and field goal percentages using the shared random effects. By combining the strengths from these two resources, our proposed method bypasses the need for pre-specifying the cluster number and provides deeper insights of players’ shooting hot spots than what the traditional position classification can offer. The utility of our method is demonstrated by applications to 2018–2019 and 2019–2020 NBA regular season data analysis.

The rest of this paper is organized as follows. We discuss a motivating example of 2018–2019 NBA regular season data in Section 2. In Section 3, we introduce the joint Bayesian nonparametric mixture model framework and provide details of Bayesian inference and posterior sampling implementation. Applications of the proposed method to NBA players data are reported in Section 4. We conclude with a discussion of future directions in Section 5. For ease of exposition, additional numerical results are given in a supplementary material.

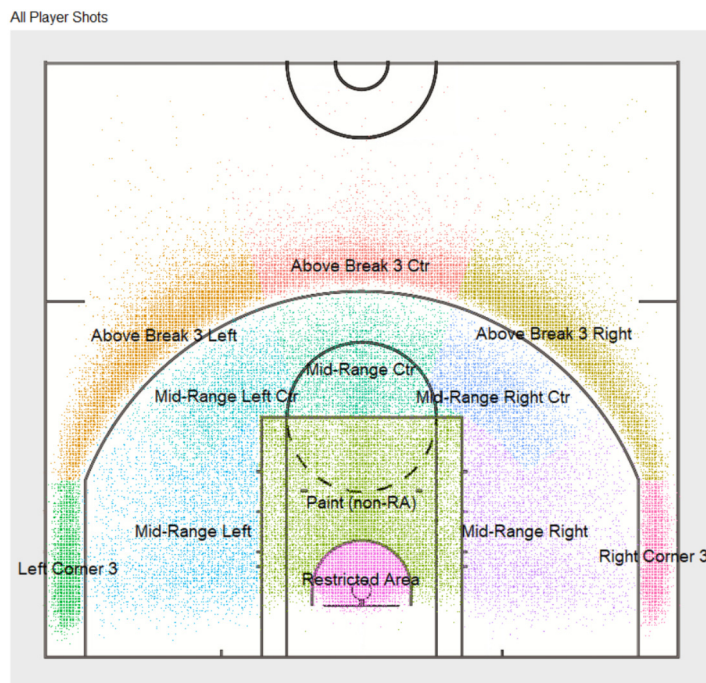


Figure 1: The twelve zones for our analysis, where each point represents a shot attempt from one of the 167 players that were analyzed.

2 Motivating Data

The shot chart information was collected via the publicly available website (https://github.com/swar/nba_api). We only discuss the data collected from the 2018–2019 NBA regular season, where records of all 219,458 shots that were taken by 526 players. Each shot record has information on the game when the shot was made (location, opponents, date), the nature of the shot (jumper, lay-up, floater, etc.), time in the game when the shot was taken, location (Cartesian coordinates), whether or not the shot was made, and who took the shot.

We only considered the shot attempts from the front court and divided the front court into twelve zones as shown in Figure 1. Shot attempts from the back court (beyond the half-court line) were omitted from this analysis because these shots were relatively rare ($n = 466$) and they did not play a major role in game strategy. Shooting patterns were then defined to be the frequency of shot attempts and field goal percentage from twelve non-overlapping regions around the front court throughout this paper. Only players who had taken at least four shots from each of the twelve front court regions were included in the clustering analysis. There were 167 players remaining which accounted for 123,944 of the total shots. Information on player height, weight, salary, and years of NBA experience was used to assess trends within the clusters. Players who were traded (changed teams mid-season) during the 2018–2019 (or 2019–2020) regular season had their statistics counted across the two teams and combined into a single player record.

To illustrate the shooting pattern heterogeneity, in Figure 2, we show the histogram of shot attempts over 12 regions for three representative players, DeMar DeRozan, Karl-Anthony Towns, and Stephen Curry, during season 2018–2019. These three players took a similar number of shot attempts in total (1313, 1314, and 1340, respectively), but exhibited quite different shooting

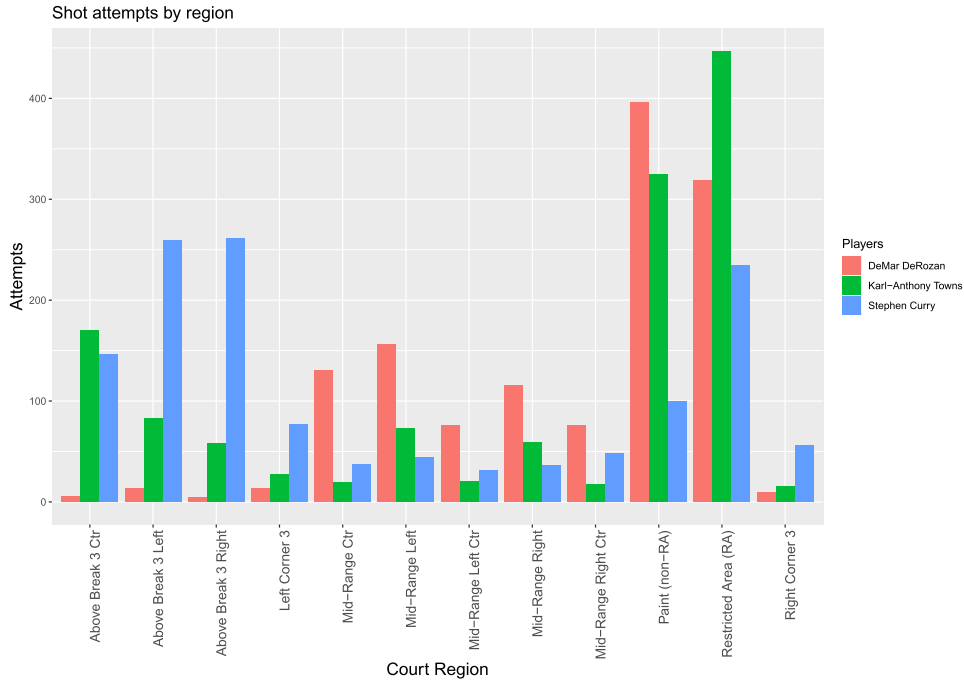


Figure 2: Number of shot attempts across 12 regions for each three selected players.

patterns. DeMar DeRozan had more mid-range shot attempts than the other two players. Most of Karl-Anthony Towns' shot attempts were in the paint or restricted area. Meanwhile Stephen Curry tended to shoot more beyond the three-point line. These findings certainly relate to and will help understand the player's game play styles and team's strategic planning.

3 Statistical Methodology

After dividing the front court into twelve zones, the number of shots that each player took from each zone is tallied and the percentage of their shots that they successfully made in each zone is computed. Let $\mathbf{Y}_i = (Y_{i,1}, \dots, Y_{i,12})$ be the vector of attempted shots from the twelve regions and $\mathbf{Z}_i = (Z_{i,1}, \dots, Z_{i,12})$ be the vector of field goal percentages from the twelve regions for player i , respectively. Our Bayesian model is defined as:

$$\begin{aligned}
 Y_{ij}|\theta_i &\sim \text{Poisson}(\exp(\mu_{ij} + \omega_{ij})), & \boldsymbol{\mu}_i|\theta_i &= \mathbf{X}\boldsymbol{\beta}_{\theta_i}, & \boldsymbol{\beta}_{\theta_i} &\sim N(0, \mathbf{I}), \\
 \text{logit } Z_{ij}|\theta_i &\sim N(\eta_{ij} + \omega_{ij}\delta_i, \sigma_j^2), & \boldsymbol{\eta}_i|\theta_i &= \mathbf{X}\boldsymbol{\alpha}_{\theta_i}, & \boldsymbol{\alpha}_{\theta_i} &\sim N(0, \mathbf{I}), \\
 \boldsymbol{\omega}_i &\sim N(0, \boldsymbol{\Sigma}), \boldsymbol{\Sigma} \sim \text{IW}(\mathbf{I}, 15), & \sigma_j &\sim \text{IG}(0.25, 0.25) & \delta_i &\sim N(0, 10^2),
 \end{aligned}$$

where i is the player index taking values from 1 to 167 (number of players) and j is the shooting region index ranging from 1 to 12 (number of regions). Here "IW" represents the inverse Wishart distribution, "IG" represents the inverse gamma distribution, θ_i is the cluster assignment for player i and $\theta_i \in \{1, \dots, K\}$, where K denotes the number of clusters. We define $\boldsymbol{\omega}_i = (\omega_{i,1}, \dots, \omega_{i,12})$ as the collection of random effects for player i and $\boldsymbol{\mu}_i, \boldsymbol{\eta}_i$ can be defined in a similar way. Note that $\boldsymbol{\omega}_i$ is the shared random effect that links the outcomes Y_{ij} and Z_{ij} together. In our model, δ_i is a scale factor for the random effect on the logit scale of the field

goal percentage for player i , Σ is the common covariance matrix for each ω_i , $\mathbf{X} \in \mathbf{R}^{12 \times 5}$ is the basis function and it can belong to any arbitrary class of spatial basis functions (e.g., thin plates basis function (Yang and Bradley, 2022), radial basis function, and wavelet (Lim, 2021), etc.). We choose the Moran's I basis function (Moran, 1950; Li et al., 2007; Bradley et al., 2015) in our analysis, and this shared spatial basis function expansion can capture dependence over basketball court between these two different types of responses (Xu et al., 2019). And it is possible to be sensitive to the choice of the number of basis functions for the spatial pattern (Bradley et al., 2011). The α_{θ_i} and β_{θ_i} are the regression coefficients that determine the rate and success probabilities for cluster θ_i .

Moran's I basis is a transformation based on the adjacency matrix of spatial locations and it has been commonly used in spatial statistics. It is a class of functions used to model areal spatial processes in a reduced dimensional space. In our case, it is a transformation of the adjacency matrix for the twelve regions over the half court. More specifically, any two regions on the court that share the same border are considered neighbors. For example, the neighbors of left corner three are mid-range left, mid-range left center, and above the break three left. Meanwhile the only neighbor for restricted area is paint (non-restricted area). Considering other options of adjacency matrix (e.g., marking all of the three point regions as neighbors or connecting the corner regions due to the symmetry of the court) is an interesting direction for future investigation. We choose not to pursue this direction because players have tendencies to drive to one side of the court, generally related to their dominant hand. By only connecting the adjacent regions, the asymmetry of player behavior is still captured.

In our model for $\text{logit}(Z_{ij})$, we assume a constant variance σ_j^2 for a fixed shooting region j over different players. This assumption can be relaxed by considering a more general variance term, i.e., replacing σ_j^2 with σ_j^2/Y_{ij} or $\sigma_j^2/f(Y_{ij})$ for some function $f(\cdot)$. In our data analysis, we have checked this assumption using residual plot and found the assumption to hold for the data analysis. Therefore we keep the current variance specification and leave its extension to a more complex form as a future work direction.

Since the number of clusters, K , is unknown, one popular way to model the joint distribution of $\theta_1, \dots, \theta_k$ is the Dirichlet process mixture model (DPMM) (Antoniak, 1974), which can be written as follows:

$$\mathbf{Y}_i \sim F(\cdot, \Theta_i), \quad \Theta_i \sim G(\cdot), \quad G \sim DP(\gamma G_0), \quad (1)$$

where $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ij})$, $\Theta_i = (\Theta_{i1}, \Theta_{i2}, \dots, \Theta_{ij})$ and $\Theta_{ij} = (\mu_{ij}, \omega_{ij})$. The process G is parameterized by a base measure G_0 and a concentration parameter $\gamma > 0$. With Θ_i for $i = 1, \dots, n$ drawn from G , a conditional prior distribution for a newly drawn Θ_{n+1} can be obtained via integration (Blackwell et al., 1973):

$$p(\Theta_{n+1} \mid \Theta_1, \dots, \Theta_n) = \frac{1}{n + \gamma} \sum_{i=1}^n \delta_{\Theta_i}(\Theta_{n+1}) + \frac{\gamma}{n + \gamma} G_0(\Theta_{n+1}), \quad (2)$$

with $\delta_{\Theta_i}(\Theta_{i'}) = I(\Theta_{i'} = \Theta_i)$ being the point mass at Θ_i . The model can be equivalently represented by introducing group membership θ_i 's and having K , the number of groups, approach infinity (Neal, 2000):

$$\begin{aligned} \mathbf{Y}_i \mid \theta_i, \Theta^* &\sim F(\Theta_{\theta_i}^*), \\ \theta_i \mid \boldsymbol{\pi} &\sim \text{Discrete}(\pi_1, \dots, \pi_K), \\ \Theta_c^* &\sim G_0, \\ \boldsymbol{\pi} &\sim \text{Dirichlet}(\gamma/K, \dots, \gamma/K), \end{aligned} \quad (3)$$

where $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)$. It can be seen that under this construction, the group-specific distribution $F(\cdot | \Theta_c^*)$ solely depends on the vector of parameters Θ_c^* .

In Equation 3, the prior distribution of $(\theta_1, \dots, \theta_n)$, which would allow for automatic inference on the number of groups K , can be obtained by integrating out $\boldsymbol{\pi}$, the mixing proportions. This is also known as the Chinese restaurant process (CRP; Aldous, 1985; Pitman, 1995; Neal, 2000). The conditional distribution for θ_i is defined through the metaphor of a Chinese restaurant (Blackwell et al., 1973):

$$P(\theta_i = c | \theta_1, \dots, \theta_{i-1}) \propto \begin{cases} |c|, & \text{at an existing table labeled } c \\ \gamma, & \text{if } c \text{ is a new table} \end{cases}, \quad (4)$$

where $|c|$ denotes the size of group c .

By adapting DPMM to our model setting for clustering, our model can be written as:

$$\begin{aligned} Y_{ij} | \theta_i &\sim \text{Poisson}(\exp(\mu_{ij} + \omega_{ij})), & \boldsymbol{\mu}_i | \theta_i &= \mathbf{X} \boldsymbol{\beta}_{\theta_i}, & \boldsymbol{\beta}_{\theta_i} &\sim N(0, \mathbf{I}), \\ \text{logit } Z_{ij} | \theta_i &\sim N(\eta_{ij} + \omega_{ij} \delta_i, \sigma_j^2), & \boldsymbol{\eta}_i | \theta_i &= \mathbf{X} \boldsymbol{\alpha}_{\theta_i}, & \boldsymbol{\alpha}_{\theta_i} &\sim N(0, \mathbf{I}), \\ \boldsymbol{\omega}_i &\sim N(0, \boldsymbol{\Sigma}), \boldsymbol{\Sigma} \sim \text{IW}(\mathbf{I}, 15), & \sigma_j^2 &\sim \text{IG}(0.25, 0.25) & \delta_i &\sim N(0, 10^2), \\ \boldsymbol{\theta} &\sim \text{CRP}(\gamma), & \gamma &\sim \text{Gamma}(0.25, 1), \end{aligned}$$

where ‘‘CRP’’ represents the Chinese restaurant process in Equation (3).

Our model can be conveniently fitted using R-package **nimble** (de Valpine et al., 2017, 2021a,b) in which a Gibbs sampler is implemented. The implementation code is given in the supplementary materials, which will be available online. Two chains were run for 20,000 iterations each with the first 5,000 discarded as burn-in. Adequacy of the mixing was assessed by viewing trace plots for individual parameters and checking autocorrelation between samples. We also checked the trace plots for convergence and found the convergence results to be satisfactory.

Another important step in posterior inference is to obtain clustering labels. Note that the Chinese Restaurant Process prior on the cluster assignments allows for different numbers of clusters between samples and the labels on the clusters do not carry any meaning. Therefore posterior inference for the group configurations needs to be carried out based on posterior samples of $\{\theta_1, \theta_2, \dots, \theta_n\}$. Further complicating matters, taking the average assignment may not lead to a valid configuration of the individuals. In our case, we conduct posterior inference on the clustering labels based on Dahl’s method (Dahl and Vannucci, 2006), which proceeds as follows. Define a membership matrix $\mathcal{A}^{(\ell)}$ as:

$$\mathcal{A}^{(\ell)} = (\mathcal{A}^{(\ell)}(i, i'))_{i, j' \in \{1:n\}} = (\theta_i^{(\ell)} = \theta_{i'}^{(\ell)})_{n \times n}, \quad (5)$$

where $\ell = 1, \dots, L$ indexes the number of post-burnin MCMC iterations, $\theta_i^{(\ell)}$ and $\theta_{i'}^{(\ell)}$ denote the memberships for players i and i' , respectively. An entry $\mathcal{A}^{(\ell)}(i, i')$ equals 1 if $\theta_i^{(\ell)} = \theta_{i'}^{(\ell)}$ which means players i and i' are in the same cluster, and 0 otherwise. An element-wise mean of the membership matrices can be obtained as

$$\bar{\mathcal{A}} = \frac{1}{L} \sum_{\ell=1}^L \mathcal{A}^{(\ell)},$$

where the summation is also element-wise. The posterior iteration with the smallest squared distance to \mathcal{A}^* is obtained by

$$\mathcal{A}^* = \underset{(\ell)}{\text{argmin}} \|\mathcal{A}^{(\ell)} - \bar{\mathcal{A}}\|_F. \quad (6)$$

where ‘‘ F ’’ is the Frobenius norm.

Table 1: Position Demographics: Mean and standard deviation for height measured in inches, weight measured in pounds, salary in millions of dollars, and NBA experience measured in seasons played.

| Cluster | | Height (in.) | | Weight (lbs.) | | Salary (million USD) | | Exp (yrs played) | |
|--------------|------|--------------|------|---------------|-------|-------------------------|-------|---------------------|------|
| Position | Size | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| In sample | | | | | | | | | |
| PG | 51 | 74.31 | 1.73 | 193.40 | 12.45 | 10.80 | 10.76 | 5.35 | 3.96 |
| SG | 61 | 76.70 | 1.58 | 205.82 | 14.18 | 8.85 | 7.50 | 4.85 | 3.65 |
| SF | 23 | 79.48 | 1.20 | 221.30 | 15.86 | 13.31 | 10.86 | 5.09 | 3.67 |
| PF | 26 | 80.73 | 1.59 | 233.73 | 12.06 | 10.49 | 8.69 | 6.65 | 5.21 |
| C | 6 | 82.17 | 0.75 | 245.67 | 6.77 | 14.76 | 12.00 | 5.5 | 3.89 |
| All | 167 | 77.18 | 2.93 | 209.93 | 20.47 | 10.52 | 9.45 | 5.34 | 4.03 |
| Whole League | | | | | | | | | |
| PG | 94 | 74.60 | 2.13 | 194.69 | 12.88 | 7.07 | 9.23 | 4.24 | 4.23 |
| SG | 121 | 76.63 | 1.64 | 204.88 | 13.63 | 5.61 | 6.65 | 3.43 | 3.59 |
| SF | 68 | 79.16 | 1.23 | 218.53 | 13.55 | 6.73 | 8.70 | 3.87 | 4.17 |
| PF | 97 | 80.82 | 1.63 | 233.35 | 13.21 | 6.28 | 7.60 | 4.79 | 4.49 |
| C | 90 | 82.68 | 1.42 | 250.39 | 16.26 | 8.89 | 8.43 | 5.41 | 4.30 |
| All | 470 | 78.61 | 3.35 | 219.41 | 24.37 | 6.83 | 8.10 | 4.32 | 4.18 |

4 Real Data Application

4.1 2018–2019 NBA Regular Season

4.1.1 Exploratory Data Analysis

The most natural way to cluster players a priori is by their position. In basketball, each team fields five players and the typical positions are point guard (PG), shooting guard (SG), small forward (SF), power forward (PF), and center (C). At any given time there is typically one player from each of these position groups on the floor. Table 1 shows the height, weight, salary, and years of experience of the players who we have data on for these categories (470 out of 526). Stratifying by position in both the subsample of players who took enough shots to be included and across the entire league shows that there are physical similarities within each position group. On average, point guards are the smallest, followed in an increasing order by shooting guards, small forwards, power forwards, and centers. The physical characteristics within positions between the sample and the whole league are similar, while the mean salary and years of playing experience are greater in the sample than the whole league, but with wide standard deviations.

The position group shooting patterns are shown in Figure 3 for players in the sample and numerically summarized in Table 2 for all players. Figure 3 shows the average location of shots made from each region, the field goal percentage from each region, and a contour density of shooting locations for each position group. Across different positions, it is clear that the proportion of 3-pointers (among all shots) varies. Within each position, shot tendencies also differ between the whole league and the sample. In general the players in the sample took more

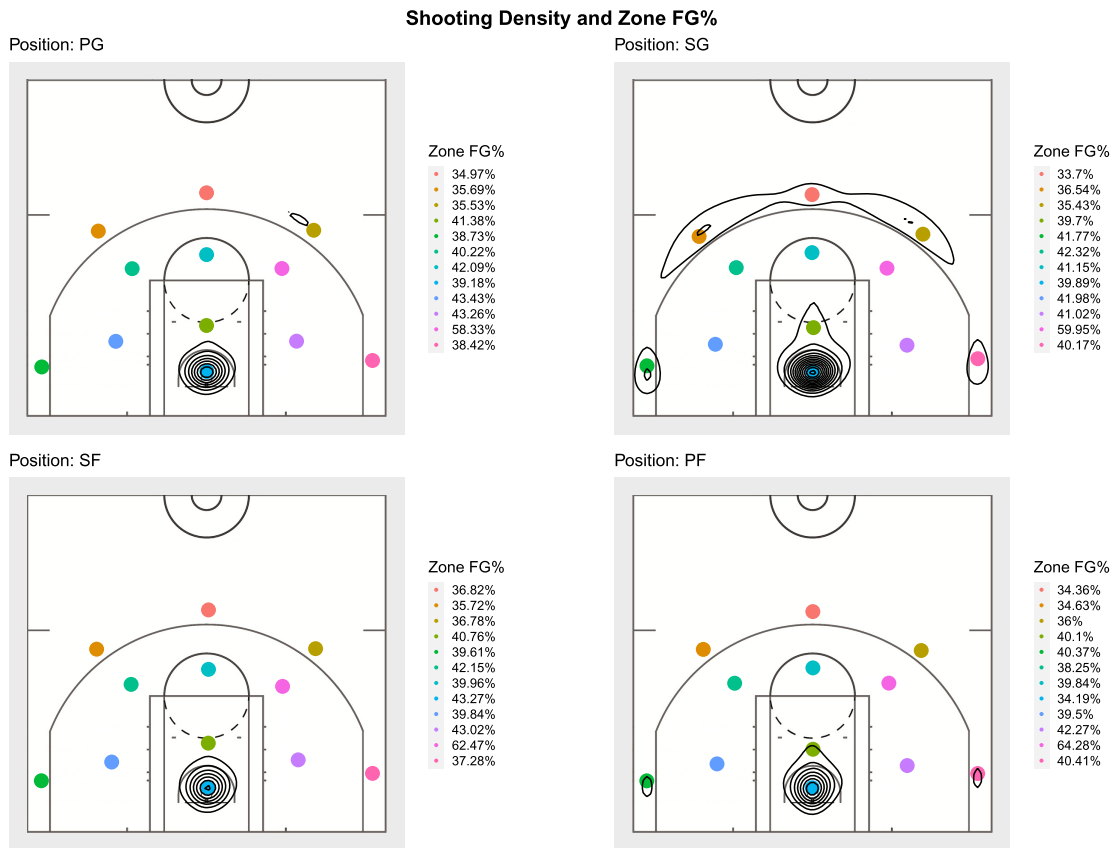


Figure 3: Shooting charts by position. Centers were excluded because there are only six centers in the dataset. The colored dots are the average position that shots were taken from each region. Each position’s collective field goal percentage in each region is shown on the right. The contour lines show where the players tend to shoot more often.

shots, which is as expected because we only choose players that have made at least the minimum number of shots in each region. In general, the goal percentages are similar between the players in the whole league and the ones we consider in the sample.

4.1.2 Cluster Analysis

We apply the proposed clustering method in Section 3 to the shot chart data of selected 167 NBA players in 2018–2019 regular season. Model diagnostics showed signs of adequate mixing. The final clustering results (e.g., \mathcal{A}^*) contain thirteen clusters. Complete cluster assignments are summarized in Table 1 of the supplementary file. When splitting the data by position, there are clear physical differences in the players in each group. Interestingly, in our obtained clusters, the height and weight differences between groups is less pronounced. More specifically, the mean heights/weights are more similar and the standard deviations of height and weight are larger indicating overlaps in height and weight among clusters. Though the physical characteristics are similar, the shooting patterns differ noticeably as shown in Figure 4.

We first discuss clusters 3, 6, 7, 12, and 13. Players from these clusters take relatively high frequency of shots from beyond the three-point arc. Interestingly, the shooting accuracy of

Table 2: Position Shooting Tendencies: The percentage of made and attempted 3-pointers and averaged all field goals (FGs). The percentage made (Made/Att) is the percentage across all shots taken by players in that position. Attempts (Att) is the average number of attempts per player in that position.

| Position | 3-pointers | | All FGs | |
|--------------|------------|--------|----------|--------|
| | Made/Att | Att | Made/Att | Att |
| In Sample | | | | |
| PG | 0.37 | 298.00 | 0.46 | 775.48 |
| SG | 0.36 | 193.83 | 0.53 | 869.83 |
| SF | 0.36 | 286.88 | 0.46 | 706.69 |
| PF | 0.37 | 314.51 | 0.44 | 741.34 |
| C | 0.36 | 276.06 | 0.44 | 725.51 |
| All | 0.36 | 291.86 | 0.45 | 740.43 |
| Whole League | | | | |
| PG | 0.36 | 167.34 | 0.45 | 408.81 |
| SG | 0.33 | 54.73 | 0.55 | 407.78 |
| SF | 0.36 | 139.95 | 0.47 | 357.54 |
| PF | 0.36 | 199.73 | 0.43 | 453.37 |
| C | 0.36 | 178.80 | 0.44 | 482.70 |
| All | 0.36 | 150.75 | 0.46 | 424.28 |

players in these cluster is not necessarily is higher than the rest. Clusters 3 and 6 have among the highest accuracy for three-point attempts combined with a high volume of attempts from all around the arc. Cluster 3 features Joe Ingles, Kyle Korver, and Stephen Curry while cluster 6 includes Buddy Hield and Seth Curry. All of these players are known for their shooting abilities from distance. The shot locations of cluster 13 are similar to those of clusters 3 and 6, but the efficiency is lower among this cluster. Cluster 13 has the lowest average height (75.54 in.) and weight (189.54 lbs.). Despite this, Dirk Nowitzki (84 in., 245lbs.) is in this cluster and 2018–2019 is his last season in his career. Clusters 7 and 12 take fewer three-point shots and they are generally located in the corners or on the wings, regions that are considered easier to make shots from.

Clusters 2, 8, 9, and 11 focus their shot attempts much more in the interior, closer to the basket. Several players in these clusters are former NBA Slam Dunk Contest winners (cluster 8: Aaron Gordon, Zach LaVine, Blake Griffin; 9: Donovan Mitchell; 11: John Wall). This is consistent with these clusters housing players who tend to take their shots near the basket. Of clusters with more than six players, clusters 8 and 9 on average have the highest salary (however, the standard deviations are quite big) and also take the most shots.

Clusters 1, 4, 5, and 10 all have six or fewer players. Clusters 1 and 10 have high salaried players who on average take many shot attempts over the course of the season. Some notable players in these two clusters are James Harden, Bradley Beal and Damian Lillard. Clusters 4 and 5 are primarily made up of bench players, which is reflected in their lower salaries and fewer shot attempts.

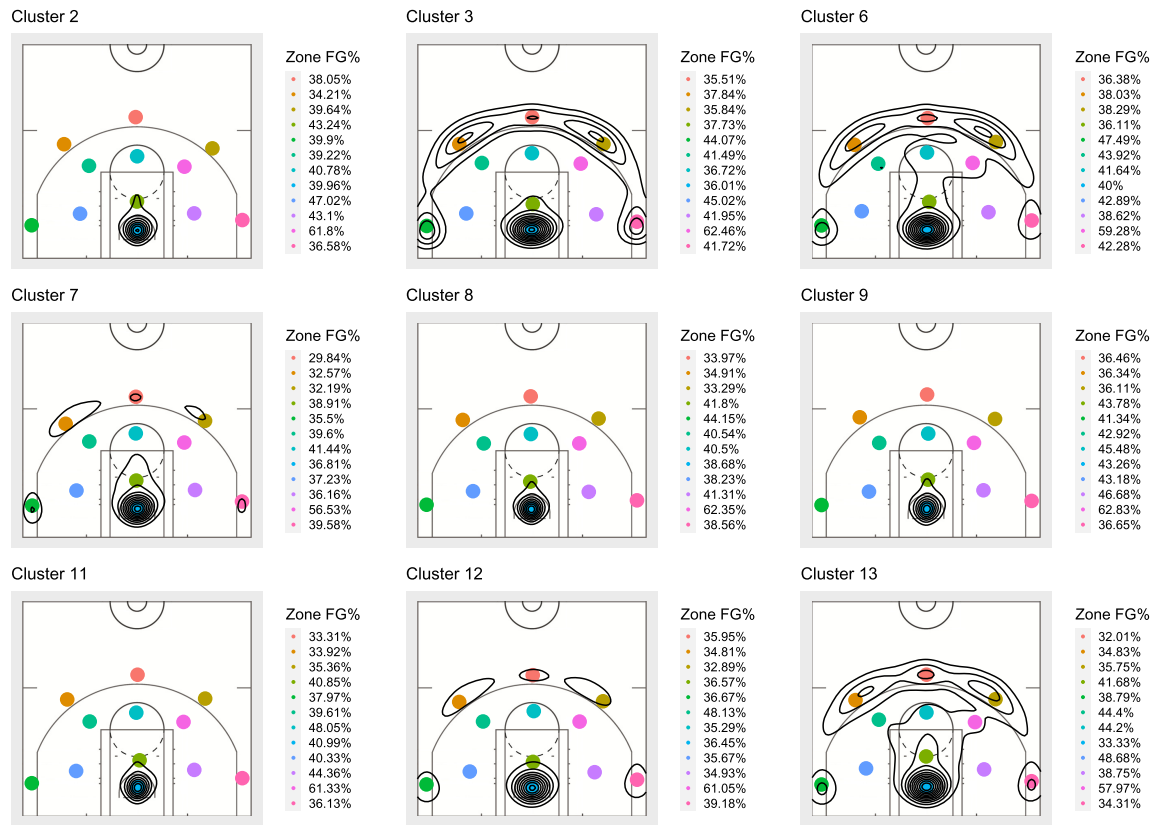


Figure 4: The shooting charts for the players in the nine clusters with at least ten players. The colored dots are the centers for each of the zones. Each cluster’s collective field goal percentage in each zone is shown on the right. The contour lines show where the players in each cluster tended to shoot from more frequently.

4.2 2019–2020 NBA Bubble Season

Another dataset of interest is collected from the 2019–2020 NBA season. The 2019–2020 bubble season is a unique season since it was interrupted by the coronavirus pandemic approximately three quarters of the way into the season. After four months suspension, the game returned with all of the remaining games hosted at a single site in Florida with no fans in attendance. Another interesting feature is the “seeding” games scheme, where twenty-two out of thirty NBA teams were invited to participate in the game in the Florida bubble and each team played eight additional regular season games to determine playoff seeding. As a result, the total number of games for 2019–2020 bubble season is less than that of a regular season. Moreover, players and coaches deemed “high-risk individuals” by their team, or players who had already suffered season-ending injuries prior to season suspension, were not permitted to play. Players that are medically cleared could also decline to participate in the bubble season. With these restrictions, there were only 138 players who met the shot requirement for each shooting zone.

We conduct the clustering analysis for this new data set and find eight clusters among 138 players. In Table 5, we provide the cluster demographics for eight clusters, respectively. Similarly with the findings for the 2018–2019 season, the average height/weight differences between groups

Table 3: Cluster Demographics: Mean and standard deviation for height measured in inches, weight measured in pounds, salary in millions of dollars, and NBA experience measured in seasons played.

| Cluster | | Height (in.) | | Weight (lbs.) | | Salary (million USD) | | Exp (yrs played) | |
|---------|------|--------------|------|---------------|-------|-------------------------|-------|---------------------|------|
| Number | Size | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 1 | 6 | 76.33 | 3.14 | 212.17 | 14.06 | 19.92 | 8.61 | 8.00 | 0.89 |
| 2 | 13 | 77.92 | 2.06 | 209.38 | 20.13 | 8.56 | 7.20 | 4.46 | 3.73 |
| 3 | 21 | 77.10 | 2.59 | 209.33 | 17.37 | 10.45 | 9.65 | 6.52 | 4.76 |
| 4 | 5 | 78.80 | 1.79 | 219.40 | 17.30 | 8.96 | 5.50 | 4.20 | 2.68 |
| 5 | 5 | 76.40 | 2.97 | 205.60 | 17.73 | 5.86 | 10.22 | 5.20 | 5.22 |
| 6 | 13 | 76.00 | 2.80 | 205.23 | 15.02 | 5.93 | 5.12 | 4.00 | 3.08 |
| 7 | 14 | 76.07 | 2.59 | 201.57 | 16.40 | 6.42 | 5.31 | 4.43 | 3.34 |
| 8 | 25 | 77.80 | 2.92 | 215.08 | 22.56 | 15.12 | 10.99 | 5.48 | 3.86 |
| 9 | 12 | 77.50 | 4.17 | 217.58 | 26.44 | 15.38 | 10.25 | 5.58 | 3.94 |
| 10 | 4 | 77.00 | 3.56 | 216.00 | 18.07 | 19.14 | 12.02 | 5.75 | 2.87 |
| 11 | 22 | 77.23 | 3.22 | 208.73 | 21.72 | 12.07 | 10.61 | 5.00 | 3.74 |
| 12 | 14 | 79.00 | 1.71 | 224.64 | 13.15 | 5.54 | 5.71 | 4.93 | 3.67 |
| 13 | 13 | 75.54 | 3.18 | 189.54 | 18.54 | 6.47 | 6.76 | 6.31 | 6.51 |

Table 4: Cluster Shooting Tendencies: The proportion of made and attempted 3-pointers and averaged across all shots (2-pointers and 3-pointers). The goal percentage (Made/Att) is the percent across all shots taken by players in that cluster. Attempts (Att) is the average number of attempts per player in that cluster.

| Cluster | 3-pointers | | All FG | |
|---------|------------|--------|----------|---------|
| | Made/Att | Att | Made/Att | Att |
| 1 | 0.37 | 614.67 | 0.45 | 1370.00 |
| 2 | 0.37 | 182.54 | 0.47 | 744.31 |
| 3 | 0.38 | 390.95 | 0.43 | 621.24 |
| 4 | 0.36 | 206.40 | 0.46 | 536.80 |
| 5 | 0.33 | 87.80 | 0.37 | 170.40 |
| 6 | 0.39 | 359.00 | 0.43 | 691.00 |
| 7 | 0.33 | 170.21 | 0.41 | 496.71 |
| 8 | 0.35 | 361.76 | 0.45 | 1099.28 |
| 9 | 0.37 | 331.00 | 0.47 | 1146.08 |
| 10 | 0.38 | 462.25 | 0.47 | 1192.00 |
| 11 | 0.35 | 236.82 | 0.45 | 682.36 |
| 12 | 0.36 | 249.21 | 0.43 | 516.36 |
| 13 | 0.35 | 183.38 | 0.42 | 384.08 |

are not obvious. The shooting patterns, on the other hand, show a high level of heterogeneity in Figure 5. And the estimation results shown in Appendix C.

Table 5: Cluster Demographics for NBA Bubble Season: Mean and standard deviation for height measured in inches, weight measured in pounds, salary in millions of dollars, and NBA experience measured in seasons played.

| Cluster | | Height (in.) | | Weight (lbs.) | | Salary (million USD) | | Exp (yrs played) | |
|---------|------|--------------|------|---------------|-------|----------------------|-------|------------------|------|
| Number | Size | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 1 | 3 | 78.67 | 2.31 | 214.33 | 28.75 | 8.66 | 1.91 | 7.00 | 3.46 |
| 2 | 7 | 75.57 | 1.90 | 197.00 | 9.81 | 6.25 | 6.42 | 3.71 | 2.93 |
| 3 | 31 | 77.48 | 3.31 | 211.19 | 20.39 | 14.13 | 10.86 | 4.10 | 3.29 |
| 4 | 30 | 77.77 | 3.40 | 212.73 | 28.95 | 17.96 | 11.27 | 6.13 | 3.58 |
| 5 | 25 | 76.48 | 3.14 | 204.68 | 22.55 | 8.17 | 8.79 | 4.04 | 4.21 |
| 6 | 6 | 77.67 | 3.01 | 215.83 | 23.95 | 8.08 | 4.83 | 3.67 | 2.34 |
| 7 | 24 | 76.71 | 2.42 | 205.33 | 19.14 | 5.83 | 7.02 | 5.29 | 4.82 |
| 8 | 12 | 77.17 | 3.74 | 216.08 | 25.64 | 20.02 | 11.83 | 7.42 | 4.62 |

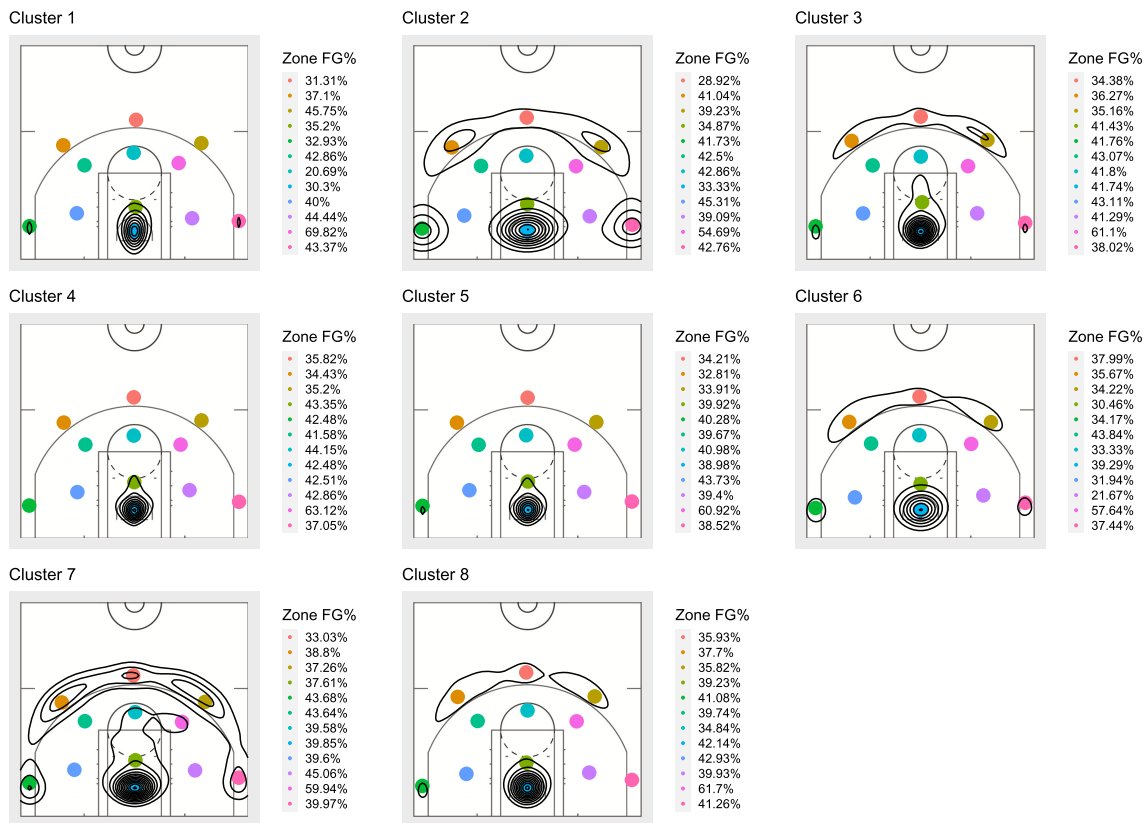


Figure 5: The shooting charts for the players in the eight clusters during NBA bubble season. The colored dots are the centers for each of the zones. Each cluster's collective field goal percentage in each zone is shown on the right. The contour lines show the most frequent shooting locations for each cluster.

Table 6: Cluster Shooting Tendencies for NBA Bubble Season: The proportion of made and attempted 3-pointers and averaged across all shots (2-pointers and 3-pointers). The percent made is the percent across all shots taken by players in that cluster. Attempts is the average number of attempts per player in that cluster.

| Cluster | 3-pointers | | All FG | |
|---------|------------|--------|----------|--------|
| | Made/Att | Att | Made/Att | Att |
| 1 | 0.39 | 180.33 | 0.50 | 478.33 |
| 2 | 0.40 | 201.71 | 0.43 | 369.14 |
| 3 | 0.36 | 326.32 | 0.44 | 816.32 |
| 4 | 0.36 | 271.80 | 0.47 | 910.43 |
| 5 | 0.35 | 156.44 | 0.45 | 517.92 |
| 6 | 0.36 | 329.50 | 0.41 | 604.00 |
| 7 | 0.38 | 258.04 | 0.43 | 497.79 |
| 8 | 0.37 | 393.25 | 0.45 | 868.33 |

To better understand the shooting pattern heterogeneity, we further examine these obtained clusters. For clusters 2, 3, 6, 7, and 8, their players take more shots from beyond the three-point arc, although the shooting accuracy does not seem higher than the average. In the order, cluster 7 has the highest frequency of shots beyond the three-point arc, followed by the clusters 2, 3, 6 and 8.

Cluster 7 can shoot the ball around all over the three-point arc than the other four clusters. Clusters 2 (e.g., Kentavious Caldwell-Pope and Kent Bazemore), 3 (e.g., Kyrie Irving and Kemba Walker), and 6 (e.g., Lonzo Ball and Fred VanVleet) share similar patterns from left 45 angle to right 45 angle. The main difference is in the corner region beyond three-point arc, i.e., cluster 2 has more shots from the corner than cluster 6 and subsequently cluster 3. For the cluster 8, it only has a limited angle three-point shooting compared to other four clusters.

Cluster 4 and Cluster 5 are very similar. The main difference is the Cluster 4 higher FG% in perimeter area than Cluster 5. Clusters 4 (e.g., LaMarcus Aldridge and Anthony Davis) and 4 (e.g., Carmelo Anthony and Russell Westbrook) are different compared to the rest of clusters since the shots are mainly made from the restricted and are closer to the basket.

Between them, the shooting tendency is still quite different as shown in Table 6, where cluster 4 has almost 50% more shooting attempts than cluster 5. In terms of player demographics, players in cluster 4 has almost double salaries compared to cluster 5. To summarize, compared to the clustering results obtained from the previous season of 2018–2019, the bubble season has a less number of clusters probably due to the less number of games (and players considered) but the clustering patterns remain similar in general. In other words, the COVID bubble does not seem to have a noticeable effect on players' shooting patterns.

4.3 Expanded Analysis for 2018–2019 Regular Season with 359 Players

In our previous analysis in Section 4.1, we have excluded some players that are primarily centers and forwards who did not take enough three-point shots. These players tend to play closer to the basket and take fewer shots from long distance (which is why they were omitted). Among them is Giannis Antetokounmpo, the 2018–2019 NBA Most Valuable Player. Here we provide an

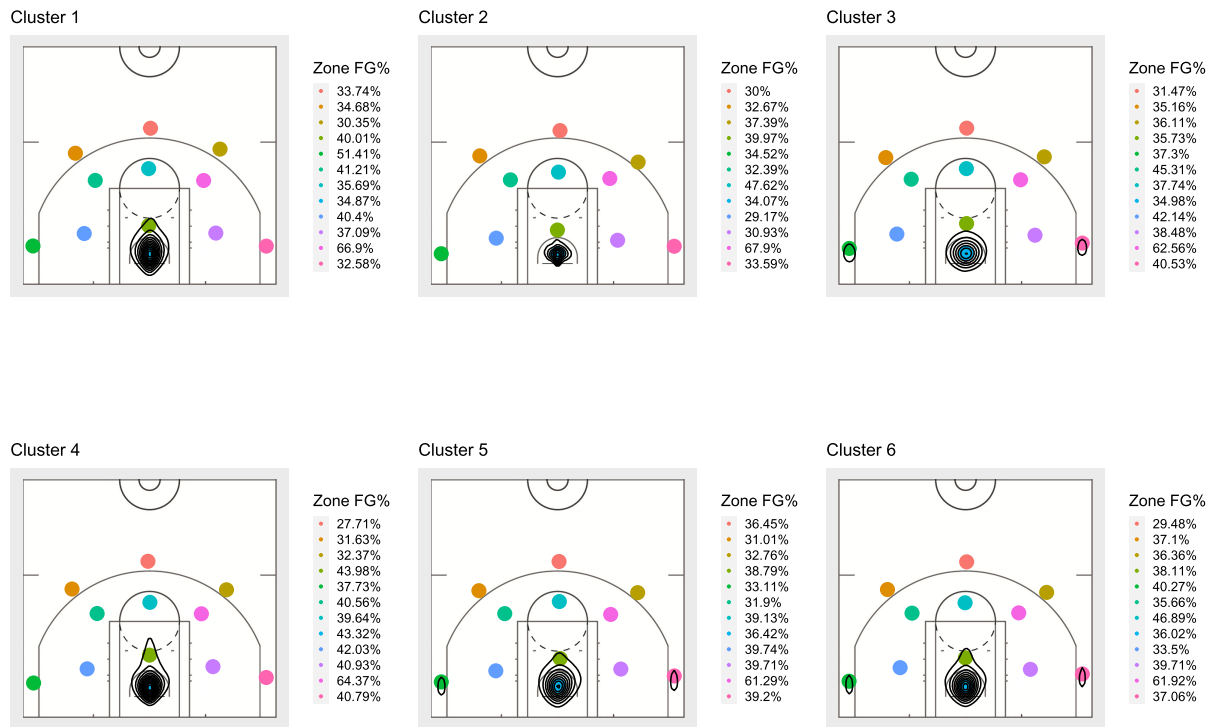


Figure 6: The shooting charts expanded analysis. The colored dots are the centers for each of the zones. Each cluster’s collective field goal percentage in each zone is shown on the right. The contour lines show where the players in each cluster tended to shoot from more frequently.

expanded analysis by enrolling the player who made more than 100 shooting attempt over the offensive court. With this threshold, we have 359 players in our analysis. We find nineteen clusters among those players. We provide the shooting pattern in Figure 6, Figure 7, and Figure 8. We also provide the cluster demographics for each cluster in Table 7 and cluster shooting tendencies in Table 8.

These nineteen clusters can be classified into three main groups. The first group includes cluster 1, 2, 4, 11, 15, 16 and 17. The second group includes cluster 3, 5, 6 and 12, and the remaining clusters are in the third group. For the first group, most of shot attempts are in the interior and close to the basket. The shooting tendency has indicated some interesting difference among different clusters as shown in Table 8. For the second group, they have in general a larger number of shots beyond the three point line with limited area than first group, especially on the corner. In the third group, players take relatively high frequency of shots from beyond the three-point arc.

Compared to the clustering results in Section 4.1, the number of players we consider here is doubled, with the cluster number increases by half (from 13 to 19), and the average cluster size increases by about half as well (from 12.8 to 18.9). While a larger number of clusters/players certainly increases the difficulty of interpretation, we do observe some common patterns between these two sets of clusters. For example, the small clusters (with ≤ 6 players in the original analysis and ≤ 10 in the expanded analysis) often consists of “outlier” players who either make a lot or very few overall shots or 3-pointers. For instance, cluster 1 in Table 4 makes the most number of shot attempts (and 3-pointers) among all 13 clusters, and cluster 5 make the least shot attempts

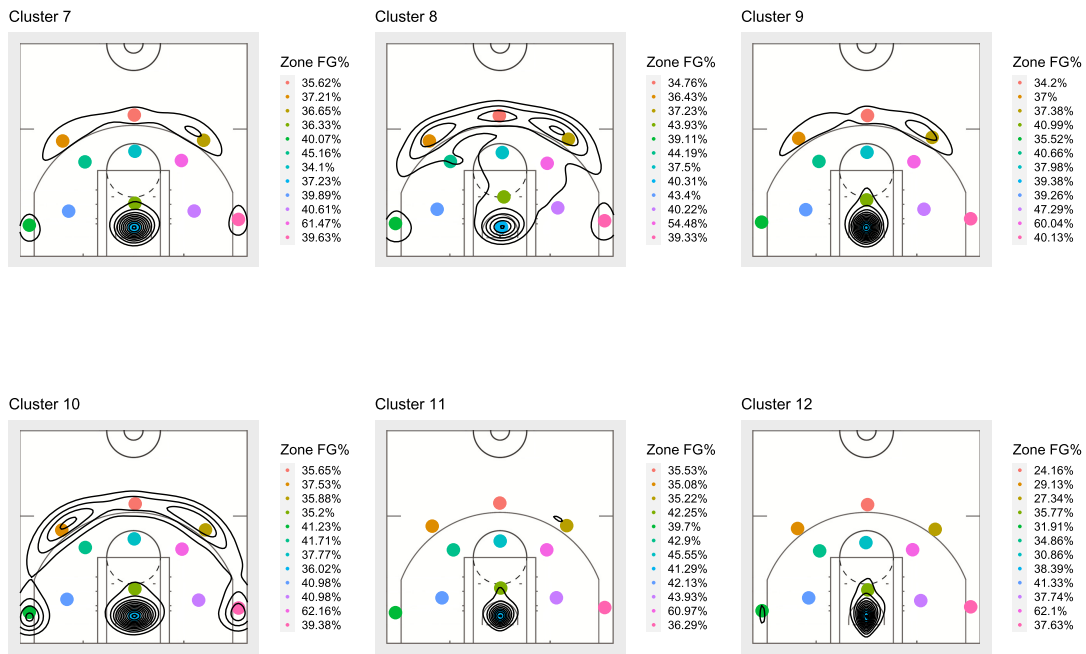


Figure 7: The shooting charts expanded analysis. The colored dots are the centers for each of the zones.

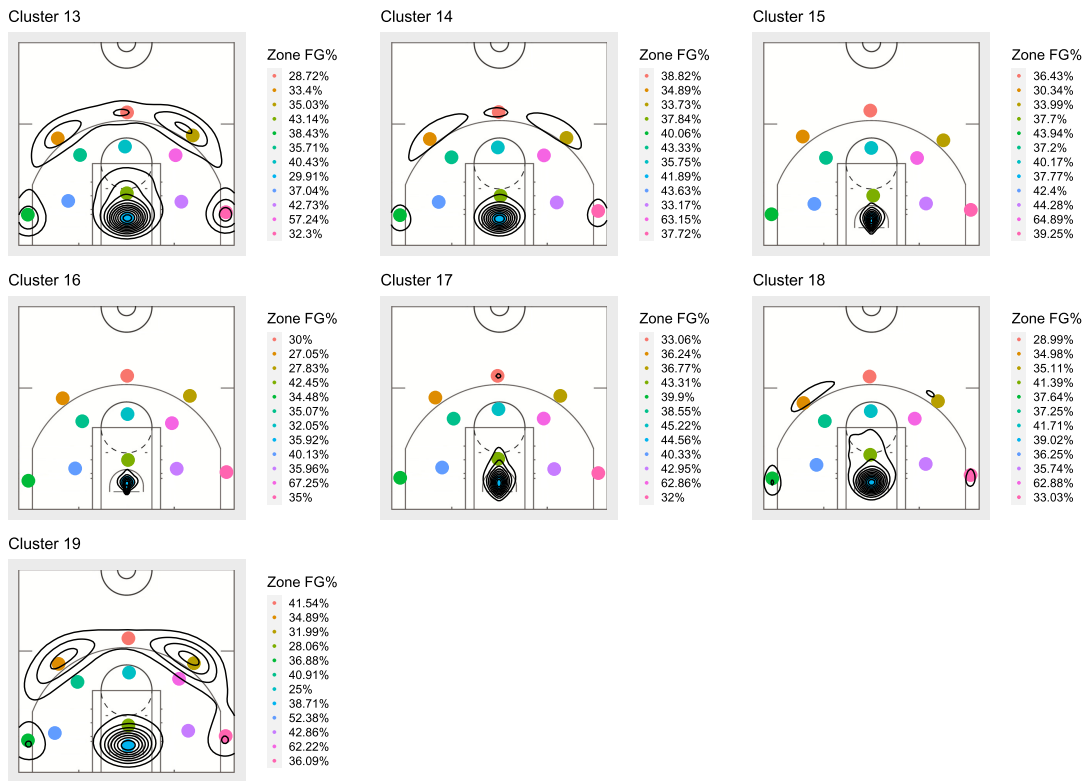


Figure 8: The shooting charts expanded analysis. The colored dots are the centers for each of the zones.

Table 7: Cluster Demographics for expanded analysis: Mean and standard deviation for height measured in inches, weight measured in pounds, salary in millions of dollars, and NBA experience measured in seasons played.

| Cluster | | Height (in.) | | Weight (lbs.) | | Salary (million USD) | | Exp (yrs played) | |
|---------|------|--------------|------|---------------|-------|-------------------------|-------|---------------------|------|
| Number | Size | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 1 | 8 | 80.75 | 3.41 | 235.50 | 29.32 | 18.03 | 11.05 | 4.00 | 2.67 |
| 2 | 9 | 81.22 | 1.92 | 240.33 | 19.80 | 12.05 | 8.90 | 4.78 | 3.07 |
| 3 | 35 | 78.77 | 2.86 | 219.83 | 20.84 | 6.62 | 6.65 | 4.77 | 4.04 |
| 4 | 29 | 79.66 | 3.83 | 226.59 | 28.51 | 10.08 | 8.86 | 6.14 | 4.61 |
| 5 | 11 | 78.09 | 1.97 | 219.73 | 19.33 | 7.21 | 8.49 | 3.64 | 3.61 |
| 6 | 11 | 78.82 | 2.56 | 218.45 | 22.30 | 9.50 | 9.62 | 3.91 | 3.08 |
| 7 | 25 | 77.56 | 2.97 | 211.00 | 25.25 | 8.18 | 9.51 | 6.48 | 5.57 |
| 8 | 18 | 76.44 | 3.47 | 209.22 | 21.48 | 6.06 | 8.72 | 7.67 | 6.44 |
| 9 | 11 | 77.18 | 3.43 | 214.27 | 16.54 | 13.50 | 10.24 | 5.64 | 3.26 |
| 10 | 33 | 78.94 | 2.51 | 217.88 | 17.92 | 5.71 | 5.04 | 4.94 | 3.79 |
| 11 | 60 | 77.13 | 3.82 | 210.95 | 25.65 | 8.95 | 9.19 | 4.40 | 3.70 |
| 12 | 12 | 79.92 | 2.57 | 227.17 | 23.50 | 6.85 | 6.55 | 4.83 | 4.73 |
| 13 | 17 | 77.76 | 2.56 | 215.71 | 20.84 | 5.75 | 7.52 | 4.35 | 3.72 |
| 14 | 16 | 78.00 | 2.71 | 217.06 | 11.86 | 7.07 | 6.52 | 3.62 | 3.03 |
| 15 | 12 | 80.25 | 2.70 | 226.25 | 22.40 | 8.58 | 8.33 | 3.83 | 3.33 |
| 16 | 18 | 81.89 | 2.22 | 247.72 | 25.01 | 10.39 | 8.62 | 4.83 | 2.87 |
| 17 | 13 | 80.77 | 3.56 | 234.77 | 32.78 | 14.39 | 8.95 | 4.92 | 2.78 |
| 18 | 17 | 76.76 | 2.77 | 210.88 | 19.31 | 3.91 | 4.35 | 4.47 | 4.60 |
| 19 | 4 | 76.50 | 1.73 | 217.50 | 14.43 | 8.91 | 4.33 | 5.50 | 3.42 |

(and 3-pointers). In Table 8, cluster 1 makes the most number of shot attempts and 4th highest number of 3-pointers among 19 clusters, while cluster 19 has made a good amount of 3-pointers and overall shots, but the overall shooting percentage (.41) is the lowest among all clusters. Some of the large clusters also share similar patterns. For example, the second largest cluster from the original analysis, cluster 11 in Table 3 and 4, takes very similar values in terms of average player height/weight and shooting attempts/percentage with the largest cluster from the expanded analysis, cluster 11 in Table 7 and 8.

5 Discussion

In this paper, we propose a flexible nonparametric Bayesian clustering approach for jointly modeling the shot attempts and goal percentage in NBA games. The clustering results provide useful insights in addition to player position and demographic information. For example, latent factors such as nature of shots that were taken (driving lay-ups compared with post-ups in the paint) are apparent in the clusters although they cannot be directly measured.

Several future research directions remain open. First, we only use raw counts in the data analysis and it will be of interest to adjust for players' number of games (or minutes). For example, two players could appear to be low volume shooters, but one player might be a high

Table 8: Cluster Shooting Tendencies for expanded analysis: The proportion of made and attempted 3-pointers and averaged across all shots (2-pointers and 3-pointers). The goal percentage (Made/Att) is the percent across all shots taken by players in that cluster. Attempts (Att) is the average number of attempts per player in that cluster.

| Cluster | 3-pointers | | All FG | |
|---------|------------|--------|----------|---------|
| | Made/Att | Att | Made/Att | Att |
| 1 | 0.35 | 284.88 | 0.49 | 1116.75 |
| 2 | 0.34 | 90.78 | 0.57 | 725.33 |
| 3 | 0.36 | 203.49 | 0.44 | 450.60 |
| 4 | 0.33 | 71.07 | 0.49 | 486.28 |
| 5 | 0.34 | 192.91 | 0.45 | 554.91 |
| 6 | 0.36 | 214.45 | 0.46 | 729.45 |
| 7 | 0.37 | 283.88 | 0.44 | 523.68 |
| 8 | 0.37 | 157.44 | 0.41 | 290.78 |
| 9 | 0.37 | 458.00 | 0.45 | 1059.82 |
| 10 | 0.38 | 242.67 | 0.42 | 390.52 |
| 11 | 0.36 | 225.18 | 0.46 | 647.33 |
| 12 | 0.29 | 64.67 | 0.46 | 280.08 |
| 13 | 0.34 | 113.65 | 0.41 | 218.18 |
| 14 | 0.36 | 338.19 | 0.45 | 666.75 |
| 15 | 0.36 | 91.42 | 0.52 | 502.08 |
| 16 | 0.30 | 29.72 | 0.56 | 609.56 |
| 17 | 0.35 | 230.54 | 0.48 | 858.77 |
| 18 | 0.34 | 110.94 | 0.44 | 351.71 |
| 19 | 0.35 | 397.25 | 0.41 | 613.25 |

volume shooter who was injured or plays few minutes while the other player plays many minutes but does not carry a heavy offensive load for their team. Intuitively these players should not be clustered together, but under the current framework they possibly could be. Secondly, our method can be applied for analyzing other related data sets, such as to investigate clustering patterns in the playoffs. In the NBA, 16 of the 30 teams qualify for the annual postseason knockout tournament that ultimately determines the season champion. During playoffs, the level of competition is increased and the top players tend to play longer minutes while the less “valuable” players may play fewer minutes in the playoffs so the number of shots they take around the court will be reduced. It could be interesting to see if clustering patterns change as competition increases in quality and intensity and how it varies for players of different talent and skill levels. In our model, we use a Dirichlet process mixture for clustering, and it will be of interest to consider other types of nonparametric priors such as the mixture of finite mixtures prior (Miller and Harrison, 2018).

Many other statistics such as rebounds (offensive and defensive) and assists are recorded for every game and could be added to the model. This would allow the clusters to capture players’ overall offensive contributions. In addition, several defensive statistics could also be added as auxiliary information to help us achieve a better understanding of the relationship between the offense and defense of NBA players. We leave these directions for future exploration.

Supplementary Material

The real data and R code used to reproduce the results in this paper can be found https://github.com/ewongtoi/nba_shot_charts. Additional tables are presented in the Online Supplementary Material.

References

- Aldous DJ (1985). Exchangeability and related topics. In: *École d'Été de Probabilités de Saint-Flour XIII—1983*, 1–198. Springer.
- Antoniak CE (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 2(6): 1152–1174.
- Blackwell D, MacQueen JB, et al. (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics*, 1(2): 353–355.
- Bradley JR, Cressie N, Shi T (2011). Selection of rank and basis functions in the spatial random effects model. In: *Proceedings of the 2011 Joint Statistical Meetings*, 3393–3406. American Statistical Association, Alexandria, VA.
- Bradley JR, Holan SH, Wikle CK (2015). Multivariate spatio-temporal models for high-dimensional areal data with application to longitudinal employer-household dynamics. *The Annals of Applied Statistics*, 9(4): 1761–1791.
- Dahl DB, Vannucci M (2006). Model-Based Clustering for Expression Data via a Dirichlet Process Mixture Model. In: *Bayesian Inference for Gene Expression and Proteomics* (KA Do, P Muller, eds.), 201–218. Cambridge University Press, Cambridge.
- de Valpine P, Paciorek C, Turek D, Michaud N, Anderson-Bergman C, Obermeyer F, et al. (2021a). *NIMBLE: MCMC, Particle Filtering, and Programmable Hierarchical Modeling*. R package version 0.12.1.
- de Valpine P, Paciorek C, Turek D, Michaud N, Anderson-Bergman C, Obermeyer F, et al. (2021b). *NIMBLE User Manual*. R package manual version 0.12.1.
- de Valpine P, Turek D, Paciorek C, Anderson-Bergman C, Temple Lang D, Bodik R (2017). Programming with models: writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26: 403–413.
- Ferguson TS (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2): 209–230.
- Franks A, Miller A, Bornn L, Goldsberry K (2015). Characterizing the spatial structure of defensive skill in professional basketball. *The Annals of Applied Statistics*, 9(1): 94–121.
- Hu G, Yang HC, Xue Y (2021a). Bayesian group learning for shot selection of professional basketball players. *Stat*, 10(1): e324.
- Hu G, Yang HC, Xue Y, Dey DK (2021b). Zero inflated poisson model with clustered regression coefficients: an application to heterogeneity learning of field goal attempts of professional basketball players. *Canadian Journal of Statistics*.
- Ibrahim JG, Chu H, Chen LM (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16): 2796.
- Jiao J, Hu G, Yan J (2021). A bayesian marked spatial point processes model for basketball shot chart. *Journal of Quantitative Analysis in Sports*, 17(2): 77–90.
- Li H, Calder CA, Cressie N (2007). Beyond moran's i: testing for spatial dependence based on the spatial autoregressive model. *Geographical Analysis*, 39(4): 357–375.

- Lim J (2021). Incorporating Bayesian variable selection into the spatial mixed effects models with process augmentation: Bayesian wavelet neural network (bwnn), Ph.D. thesis, The Florida State University.
- Long JD, Mills JA (2018). Joint modeling of multivariate longitudinal data and survival data in several observational studies of huntington’s disease. *BMC medical research methodology*, 18(1): 1–15.
- Miller A, Bornn L, Adams R, Goldsberry K (2014). Factorized point process intensities: A spatial analysis of professional basketball. In: *International conference on machine learning*, 235–243. PMLR.
- Miller JW, Harrison MT (2018). Mixture models with a prior on the number of components. *Journal of the American Statistical Association*, 113(521): 340–356.
- Moran PA (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2): 17–23.
- Neal RM (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2): 249–265.
- Pitman J (1995). Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields*, 102(2): 145–158.
- Reich BJ, Hodges JS, Carlin BP, Reich AM (2006). A spatial analysis of basketball shot chart data. *The American Statistician*, 60(1): 3–12.
- Xu Z, Bradley JR, Sinha D (2019). Latent multivariate log-gamma models for high-dimensional multi-type responses with application to daily fine particulate matter and mortality counts. arXiv preprint: <https://arxiv.org/abs/1909.02528>.
- Yang HC, Bradley JR (2022). Bayesian inference for spatial count data that may be over-dispersed or under-dispersed with application to the 2016 us presidential election. *Journal of Data Science*, forthcoming: 1–13.
- Yin F, Hu G, Shen W (2022a). Analysis of professional basketball field goal attempts via a bayesian matrix clustering approach. *Journal of Computational and Graphical Statistics*, forthcoming: 1–23.
- Yin F, Jiao J, Hu G, Yan J (2022b). Bayesian nonparametric estimation for point processes with spatial homogeneity: A spatial analysis of nba shot locations. In: *Proceedings of the 39th International Conference on Machine Learning (ICML)*.