

A CLUSTER ANALYSIS OF NBA PLAYERS

DWIGHT LUTZ
DWIGHT.LUTZ@GMAIL.COM
303-229-8959

ABSTRACT. To revolutionize the concept of basketball “positions”, we use a multivariate cluster analysis to group NBA players based on several statistics. Once these clusters have been formed, we analyze how each cluster affects winning in the regular season. Along with the main effect of each cluster, we analyze the 2 and 3-way interactions of the clusters. The ultimate goal is to determine which types of players are most relevant to winning, and also what combinations of players affect winning. This will give us a way to scientifically measure team chemistry.

1. Introduction

Every NBA player is given a label that is meant to describe what that player should do on the court. There are 5 of these labels and we call them positions. These labels are decided first by the player’s physical size. If he’s big and strong, then he is going to be a power forward or center. If he’s small and fast then he is labeled as some sort of guard. From there, a general view of the player’s skills helps to sort him into one of the 5 specific positions. If he’s a guard and shoots it well, he’s a shooting guard. If he handles the ball and passes well, he’s a point guard.

The point is, this process is far from scientific even though it has a big impact on coaches’ decisions with regards to playing time, and general managers’ approaches to acquiring new players. This paper aims to use a multivariate cluster analysis to redefine the positional labels that are given to NBA players. We will abandon the traditional method of using a player’s height and weight, since these factors can be terribly misleading. Certain small players are able to be effective inside the paint on both ends of the floor. Similarly, plenty of big men have success on the perimeter. This paper will focus less on who the player is and more on what that player actually does on a basketball court. Several statistics are used in an effort to capture the style of play for each individual, and we put very little emphasis on how effective the player is. For example, you will find Dwight Howard in the same cluster as Spencer Hawes. The two players have similar styles, not similar effectiveness.

We will develop the process that was used to form the clusters and give a specific description for each cluster. A specific number of clusters was not forced. However, intuition can quickly tell you that 5 positions seems to be far too few to describe all the different styles of play that exist on a basketball court. Conversely, using a large number of clusters will overfit the clusters to individuals, and would make analyzing differences between clusters difficult due to small sample sizes. One of the key focuses of this paper will be determining which types of players are targeted by winning teams and which types of players are more often found on losing teams. It is quite possible that a team should choose an inferior player in order to meet a certain need. Of course, this thought process is nothing new, but this paper will use a much more scientific method to determine the types of players that exist, and the relevance of those players to a team.

The first part the paper will focus on individual positions and their effect on winning. The second part will consider different combinations of the clusters that we will develop. Basketball may be the hardest game to evaluate individual players since each player’s performance is greatly affected by the other players on the court. Just as we will determine which individual positions relate to winning, we will try to understand what combinations of positions are most often found on winning teams. This type of analysis can give a general manager a tangible idea of how well a certain player will fit on a new team. Obviously, team chemistry goes far beyond the statistics that will be considered in this paper, but our methods will be an evolution of the

current thought process that NBA executives use to evaluate the compatibility of an available player with the team they are operating.

2. Cluster Analysis

In order to construct clusters that represent playing style we needed to use variables of the same variety. The following variables were used: games played, minutes played per game, percent of made field goals that are assisted, assist rate, turnover rate, offensive rebound rate, defensive rebound rate, steals per 40 minutes, blocks per 40 minutes, and the number of shots attempted per 40 minutes at each of the following locations: at the rim, from 3-9 feet, from 10-15 feet, from 16-23 feet, and beyond the 3-point line (see Table 1).

Since many of these variables are measured on different scales, we standardized all the variables using z-scores in order to put them all on the same scale, and thus gave an equal weight to each variable. Next, we experimented with several different clustering algorithms including the classical k-means [1] method and Ward's hierarchical clustering [2]. The method that we ultimately decided to use was the EM algorithm for Gaussian mixture models [1]. The EM algorithm has proven to be extremely effective in maximizing the separation of the clusters and has become the modern form of cluster analysis [1] [3].

The cluster analysis was performed on all players from the 2010-2011 season having played in at least 30 games and averaging at least 10 minutes a game. This left us with 329 players with which to build the clusters. All data was collected from hoopdata.com [4]. The Mclust function was used in the R Project for Statistical Computing, and this function uses the Bayesian Information Criterion to determine the parameters of the model and how many clusters to use [3]. This is how we settled on using 10 clusters. The most important trait of the cluster solution is the separation of the clusters, which we examine graphically in Figure 1. All variables seem to be showing strong variability from cluster to cluster, which mathematically validates the use of each variable. If there was a variable that remained relatively constant across clusters, then we would consider removing it. In order to increase the sample size, we also considered players from previous seasons, but we needed put these players into one of the clusters that we had just created. For this we used Fisher's Linear Discriminant [2] to place the players from the 2008-2009 and 2009-2010 seasons into one of the 10 clusters. Discriminant analysis is a classification technique used to determine group membership of a new set of observations. From this group, there were 646 players that met the games and minutes played criteria described above, giving us a total of 975 observations. When using Fisher's Linear Discriminant to predict cluster membership, we had a misclassification rate of about 16%. Moving forward, we can expect 84% of players from the 08-09 and 09-10 seasons to be classified into the correct cluster. Although there is a bit of error here which could potentially be improved upon with techniques such as the Support Vector Machine, we are rather pleased with accuracy of our cluster predictions.

We left out many details regarding clustering with EM, Fisher's Linear Discriminant, and the Mclust function. For more insight on the mathematical theory behind these topics, see Bishop [1], Izenman [2], and Fraley and Raftery [3] respectively.

Before we begin to talk about which clusters affect winning, it is important to realize the relevance of the clusters that we have created. One could look at the names of the clusters that I have given in Table 2 and begin relate them to the 5 traditional positions. When doing this, you might be surprised by which clusters certain players fall into. For instance, I labeled one cluster as "Ball Handlers" since that cluster had high assists, high turnovers, very few assisted field goals, high steals, low rebounds, and most of their shots are from the perimeter. Those characteristics describe a typical point guard. Uninterestingly, many of the league's point guards find themselves in that cluster. However, Hedo Turkoglu (on the Magic only) is also in that cluster. All of Hedo's stats agree with the description of this cluster given above (except for the steals). Hedo was constantly creating his own shot, and doing so for others. Whether it was effective or not, Turkoglu was playing point guard, or as we've labeled it, "Ball Handler", for the Orlando Magic last season. Another interesting result is that neither of the two Laker point guards, Steve Blake or Derek Fisher, find themselves in the "Ball Handler" cluster, they are both in the "Combo Guard" cluster. The members of this cluster score mostly on assisted field goals, but rarely turn the ball over. It has been well documented that Phil Jackson's triangle offense requires less need for a traditional point guard than a usual NBA offense. It is quite possible that Steve

Blake, a more traditional point guard for most of his career, struggled last season because he was being forced to play a position he was not used to playing. It would be hard to see this phenomenon from the traditional perspective we have of NBA positions, since he played point guard both before the Lakers and with the Lakers.

These are just two examples of many situations where this analysis allows GM's and coaches to more efficiently acquire and play the correct players for their team. Could the Magic have gone stretches of games or the season with the 6'10" Turkoglu at point guard? Coach Stan Van Gundy would have been called crazy for trying it, but Hedo was already doing it. Turkoglu's style of play is evidence that there would have been a smooth transition if the Magic had no other point guard on the floor. This does not imply that the Magic should have dumped all other point guards to let Turkoglu play the position, it simply means that Turkoglu's effect on the game most likely would not have changed if he was the only point guard. Recall, the result is not that Hedo is an effective point guard, it is that his style of play resembles that of point guard, which is why he was categorized as one.

For more detail on how each cluster is defined, see Table 3 which gives the average z-score for each statistic for each cluster. Also note some typical members of each cluster in Table 2.

3. The individual effects of the clusters on winning and losing

This section of the paper is our best effort to determine which types of players on a basketball court are most important to a team's overall success. Before we begin, I want to reemphasize that we are not accounting for the quality of the players. Two players from the same cluster might be vastly different in terms of effectiveness, but they will always be similar in style. When we label a player as a "shooter" that means he shoots a lot; it does not necessarily mean that he makes a lot of shots. The better players will always produce more wins, but we will show you which types players seem to be found more often on the winning franchises on average.

First we must define winning and losing. We did not use wins, but instead used average point differential per game for all 90 teams from the last three regular seasons. The first method was to analyze the effect of the clusters on point differential as a one-way Analysis of Variance (ANOVA) [5] with ten levels of the cluster factor and each of the teams being the 90 observations. From this we got a p-value of .0111 for the F-test. This tells us that there is a significant cluster effect on team point differential. Now we investigate which clusters are the most relevant. We calculated all pairwise 95% confidence intervals for the differences between the cluster means using Bonferroni's correction. Unfortunately, zero was in all of these 95% confidence intervals, so we could not determine that point differential was significantly greater for any one cluster than another. However, we did compute the mean point differential for each cluster and those results agree with the more interesting analysis that we give below.

Next, we considered only the winning teams, or those with positive point differential. From the pool of winning teams, we wanted to find out which types of players were most common. Note that each cluster does not have the same number of players; the number of players per cluster is given in Table 4. Many inferences can be made from the varying sizes of the clusters. For example, there is an abundance of Ball Handlers (8) but the "Durable Shooters" are quite scarce. The Durable Shooters (10) shoot many 3's and some long 2's, but also have high steals and play an incredibly large amount of games, thus we describe this group as durable. This is an area where general managers can find potential value. If a team understands how abundant Ball Handlers are, that team is less likely to overpay for one. However, it is important to understand how much each position effects winning in order to understand value. It is quite possible that Ball Handlers are a critical piece to the success of a franchise, in which case it makes sense to pay a high price for one.

So the question of interest is, of the players in each cluster, what percentage of them are on winning teams? Table 4 gives these percentages ordered from highest to lowest. Durable Shooters (10) are the type of player that most often are members of winning organizations as 66.7% of these players are a winning team. So the Durable Shooters are the most rare breed of player and they are the most owned player by winning teams. These two aspects together make Durable Shooters extremely valuable. How about the Ball Handlers we mentioned earlier? Only 44% of these players find themselves on winning teams. When you combine that

with how large of a group they are, it seems that Ball Handlers are rather expendable.

If a certain type of player had no effect on a team's point differential, then we would expect 50% of the players in that cluster to be on winning teams. So we should compare each percentage in Table 4 to 50%. Actually, since only 48% of all players were on winning teams, we should compare each percentage to 48%. For whatever reason, there were more qualifying players on the losing teams. An important question we need to answer is, are the percentages of players on winning teams for each cluster significantly different than 50%? By significantly, we mean in the statistical sense of the term. To determine this, we did one-sample T-tests for each cluster and the two-sided p-values for these tests are given in Table 4.

We can conclude that there is strong evidence ($p\text{-value} < .1$) that the following clusters have a positive impact on point differential: Durable Shooters (10) and Combo Guards (1), whereas these clusters have a negative impact on point differential: Aggressive Bigs (7) and Perimeter Scorers (9). Also, there is some evidence ($p\text{-value} < .15$) that Ball Handlers (8) and Big Bodies (6) have a negative impact on point differential as well. Although the other 4 clusters have a percentage of their players on winning teams that is a bit different from 48%, these results are not statistically significant, and thus inconclusive. Collecting more data could potentially allow us to determine if the results from these remaining clusters are meaningful.

Now let us consider these results from a basketball point of view. Even though we have strong mathematical evidence, does it make sense that Durable Shooters (10) are stockpiled by winning teams? Let's recall how this cluster is defined. This group plays a high amount of minutes and plays more games than any other cluster. This type of consistency is sure to help a team to stay on balance throughout the course of a season. There have been many cases of teams struggling for extended stretches because of injuries, but the Durable Shooters simply do not get hurt. There is also something to be said about this cluster's high amount of minutes played per game. With long-distance shooters, coaches often find it difficult to play these types of players extended minutes because of their lack of contribution on the defensive end and in other areas. However, this group has convinced its coaches to leave them on the floor despite the fact that their most relevant trait is shooting without creating their own shot (they have a high percent of assisted field goals made). This group also collects a large number of steals which allows us to believe that these players have at least somewhat of an impact on the defensive end of the floor. Most importantly, this groups turns the ball over less than any other cluster but one. These are all traits common to a team that has an efficient offensive flow, and since these players are competent defensively, it seems of little surprise that winning teams collect these types of players.

Now let's look at a cluster that is found more often on losing teams. The Aggressive Bigs (7) cluster has the lowest percentage of its players on winning teams, a result that is quite significant given the small p-value. This is group is mostly made up of power forwards and centers who rebound on both ends and block shots extremely well. This group also like to be aggressive offensively. They take well more shots at the rim, from 3-9 feet, and from 10-15 feet than average. They even shoot some from 16-23 feet. So why do we find this type of player on losing teams? The answer is not quite as clear as it was for Durable Shooter group. Consider this, the Aggressive Bigs cluster is very similar to the Defensive Bigs (2) cluster, which has a positive (although statistically insignificant) percentage of players on winning teams. The main difference between the two clusters is that the Aggressive Bigs take a lot more shots from 3-23 feet. Both clusters are composed mostly of bench players as they both have low minutes played per game. Perhaps the members of the Aggressive Bigs cluster are too aggressive offensively and tend to disrupt the team's offensive flow. Furthermore, the Big Bodies (6) cluster is another similar group that also has a negative impact on point differential. Players in the Big Bodies cluster play extremely few minutes per game, but tend to shoot many shots from 16-23 feet. This is even more evidence that big men who come off the bench and are overaggressive offensively tend to find themselves on losing teams. Whether they are losing because they are shooting too much, or they are shooting too much because they are losing is an entirely different question, but the two traits are undoubtedly intertwined.

4. The effects of pairs and triples

As mentioned earlier, every player on a basketball court is greatly affected by the other players on his team. Many people call basketball the ultimate team game since all 5 players need to have a certain basketball

chemistry in order to be successful. In the previous section, we discussed how individual types of players impact a team's success, which is essentially the "main effect" of the clusters on winning. In this section, we will focus on which combinations of clusters are most often seen on winning teams. In other words, we will consider the "interaction effects" of the clusters on point differential. Ideally, we would not only consider what pairs of clusters affect a team's point differential, but also which triples, groups of 4, 5, and even 6 clusters are most often found on winning teams. For the sake of brevity, we will only cover pairs and triples in this paper. After all, we saw 218 different three-way combinations on the different teams.

Although there are many possible strategies to analyze the interactions between clusters, we chose to maintain consistency and analyze the combinations in a similar fashion as the individual clusters, which is to determine which combinations show up far more often on the winning teams than on losing teams. From there, we did one-sample T-tests to determine if the percentage of each combination that was found on winning teams was significantly different from 48%.

Table 5 gives the ten 2-way combinations with the highest percentage of players on winning teams. Included in the Table 5 are the p-values for the two-sided T-tests. Clearly, the results for the combinations are much stronger than those for the individual clusters. Mathematically, this result is biased towards the performance of the individual clusters. When considering a combination of two individual clusters that are very common on winning teams, a few occurrences of that combination will appear on winning teams even if no true interaction between the two clusters exists. However, from a basketball standpoint, it seems rather intuitive that two players can have a greater marginal impact on a game than one player can. This concept is one of the reasons that basketball analysts have begun using five-man performances alongside a player's individual output. In order to account for the bias, we included a column in the table called "Increase". This column describes the impact (positive or negative) that adding the worse performing cluster to a team (that already has a player from the better performing cluster) has on the team's point differential. So adding a 10 to another 10 increases a team's chance of having a positive point differential by $78.95 - 66.67 = 12.3\%$.

With regards to specific clusters, there are many conclusions to be had. The Durable Shooters (10) cluster seems to interact well with several other clusters. We find this as no surprise given the breakdown of the characteristics of that cluster above. The other cluster that is prominently featured in Table 5 is the Combo Guards (1). This group spreads the floor by shooting from range and rarely creating their own shot in a similar fashion to the Durable Guards, but they handle the ball more as they have high assists and high turnovers. These two groups seem to interact well with each other. Both of these clusters play a high amount of games, spread the floor, and rarely create their own shot. In other words, they can easily complement players of any other style of play. The same is simply not true for other types of players. For example, the Perimeter Scorers (9) group does not coexist well with many other clusters. In fact, this group is peppered throughout Table 6, which are the 10 combinations most often found on losing teams. Furthermore, the combination of two Perimeter Scorers has the lowest percentage of appearances on winning teams of any possible pair. Also, the negative impact of adding a second Perimeter Scorer to a team is the greatest of any addition of a single cluster to a team that already has one type of player. As seen in the "Increase" column of Table 6, adding a second perimeter scorer (assuming the team already has one) drags the team's probability of having a positive point differential down by about 19%.

Why is the Perimeter Scorers cluster so incompatible with others? This cluster takes a ton of shots. Note in Table 3 that they take more shots than average from every distance. They rarely score assisted field goals and essentially are one-on-one players. The argument has been made for years that a team needs to move the ball on offense and cannot have players who stop the ball to go one-on-one. All of our research supports this idea, and suggests that having two players like this is exponentially worse. A case in point of this is the situation with Dwyane Wade and LeBron James of the Miami Heat. The struggle for these two to coexist has been well documented, and it so happens that both players are in the Perimeter Scorers cluster. If LeBron and Wade were equally talented to the average player in their cluster, we would have given them less than a 23% chance to have a positive point differential, holding the skill level of all other teammates constant. Since LeBron and Wade are far above average, they were able to excel from the standpoint of winning, but their

early struggles would have been no surprise to this research.

Although there are many other conclusions we could make about the 2-way combinations, in fact there were 36 combinations significantly different from 48% at the $\alpha = 0.1$ level, we will move on to the three-way combinations, or triples. Table 7 gives the top ten triples that are most found on winning teams along with their P-values for two-sided T-tests. Again, there's no surprise that Durable Shooters (10) and Combo Guards (1) play a prominent role. However, notice that the Perimeter Scorers (9) are in every single one of these top triples! We spent the last paragraph talking about how poorly they mesh with other clusters, but now they seem more valuable than ever. The point is, when a team is able to put the correct players around a Perimeter Scorer, the chances of having a positive point differential increase dramatically. The 3-way combination of Perimeter Scorer (9), Durable Shooter (10), and Combo Guard (1) is found on winning teams 24 out of 25 times (Table 7), which seems far from an accident. From a basketball standpoint, it seems that you have a guard willing to give the Perimeter Scorer the ball and then spot up. Then the Perimeter Scorer gets to play one-on-one and dish off to the Combo Guard or Durable Shooter if the defense puts too much focus on his one-on-one move. Remember, the Perimeter Scorers have the 4th highest assist rate of any cluster. Another cluster that's worth mentioning here is the Defensive Bigs (2). This group performs consistently well individually and in 2 and 3-way combinations. The interesting part of this is, this group shoots far less than any other of the "big men" clusters. They interact extremely well with the offense-oriented clusters (Perimeter Scorers, Durable Shooters, even the Floor Spacers (4)). This result suggests that it makes sense to have "specialty players", or certain players on the team that excel on offense and other players that excel on defense.

Table 8 gives us the other end of the spectrum, or the triples most often found on teams with a negative point differential. It's clear that the Aggressive Bigs (7) performs extremely poorly, and we discussed earlier why this might be the case. Let us consider some of the clusters that are interacting with the Aggressive Bigs. Notice that several of these clusters are comprised of big men. The Defensive Bigs (2), Elite Bigs (5), and Big Bodies (6) are clearly all big men of different varieties. These results question the age old basketball theory that a team can never have too much height. Several 3-way combinations made up of multiple big men often find themselves on losing teams.

The other cluster that is prevalent in Table 8 is the Versatile Swingmen (3) group. This the epitome of a the type of player that does many things well, but does not do one thing great. They rebound well from the perimeter, they get a few assists and steals, even a few blocks. Most of their shots are assisted and happen at the rim or outside the 3-point line. NBA analysts often call these types of players "glue guys", and claim that every winning team needs a player or two like this. However, having three Versatile Swingmen seems to be rather detrimental, as only 4% of these triples are on teams with positive point differential. Also, this group does not seem to combine well with big men that are aggressive offensively, as we see the (3,7) and (3,6) popping up in both the losing triples (Table 8) and losing doubles (Table 6). These results show that we may have a tendency to overrate the intangible qualities and perimeter defense that this type of player brings to the table.

5. Conclusion

The results that have been described in this paper are only a subset of the information that can be gained from this analysis. However, there are potential criticisms to this analysis, such as the thought that the quality of the player is more important than the type of player. After all, we can all agree that the Miami Heat did not have compatible parts, but they still won 58 games and had a +7.5 point differential. Also we recognize a lack defensive statistics used in the cluster analysis. This was mainly due to data limitations. Also, our results would be more accurate if we had not used a binary dependent variable to represent winning. Despite all of this, we have been able find some very powerful trends with regards to how different types of players affect winning. We cannot say that acquiring Durable Shooters is the undisputed way to win NBA championships, but we cannot deny that their presence helps teams win games either. Our analysis of the combinations of the clusters gives a new way to think about team chemistry and how the compatibility of a team's pieces will affect winning on average. Overall, we feel that the development of these clusters produces a more scientific way to think about the positions of individual players, and the overall structure of a team's roster.

6. Appendix: Figures and Tables

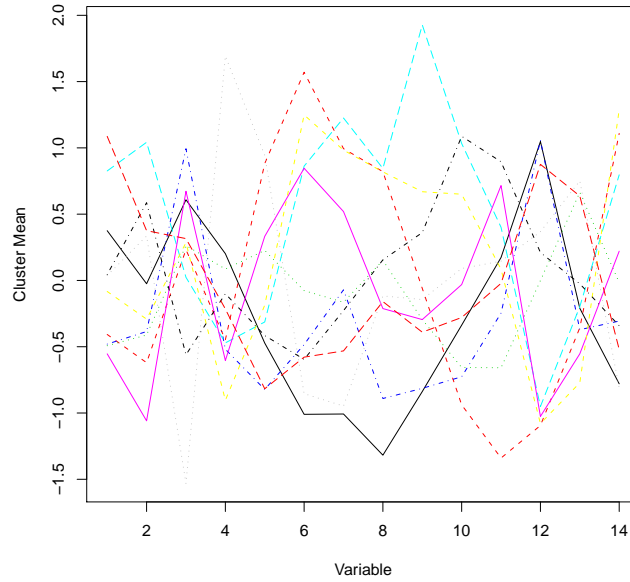


FIGURE 1. We are interested in the separation of the clusters for each variable. Each line represents one of the 10 clusters, and we hope to see each cluster have a unique line such that no two lines fall on top of each other.

TABLE 1. Variable Abbreviations and Formulas

Variable	Abbreviation	Formula (if needed)
Games Played (minimum of 30)	GP	-
Minutes played per game (minimum of 10)	Min	-
Percent of made field goals that are assisted	% Ast	$\frac{\text{made field goals that are assisted}}{\text{total made field goals}}$
Assist Ratio	AR	$\frac{\text{Assists} \times 100}{\text{FGA} + (\text{FTA} * .44) + \text{Turnovers}}$
Turnover Ratio	TOR	$\frac{\text{Turnovers} \times 100}{\text{FGA} + (\text{FTA} * .44) + \text{Turnovers}}$
Offensive Rebound Rate	ORR	$\frac{100 \times (\text{Player ORebs} * (\text{Team Min} / 5))}{(\text{Player Min} * (\text{Team ORebs} + \text{Opp DRebs}))}$
Defensive Rebound Rate	DRR	$\frac{100 \times (\text{Player DRebs} * (\text{Team Min} / 5))}{(\text{Player Min} * (\text{Team DRebs} + \text{Opp ORebs}))}$
Attempted field goals at the rim per 40 minutes	Rim	-
Attempted field goals from 3-9 feet per 40 minutes	3-9	-
Attempted field goals from 10-15 feet per 40 minutes	10-15	-
Attempted field goals from 16-23 feet per 40 minutes	16-23	-
3-point field goals attempted per 40 minutes	3s	-
Steals per 40 minutes	Stls	-
Blocks per 40 minutes	Blks	-

TABLE 2. Cluster labels and typical members

Cluster	Cluster Label	Typical Members
Combo Guards	1	Steve Blake / Mario Chalmers / Rudy Fernandez
Backup Bigs	2	DeAndre Jordan / Ben Wallace / Brendan Haywood
Skilled Swingmen	3	Shane Battier / Lamar Odom / Paul George
Floor Spacers	4	Channing Fry / Matt Bonner / Mike Miller
Elite Bigs	5	Amare Stoudamire / Elton Brand / Brook Lopez
Big Bodies	6	Jason Collins / Antonio McDyess / Kurt Thomas
Aggressive Bigs	7	Brandon Bass / Carlos Boozer / Tyler Hansbrough
Ball Handlers	8	Chris Paul / Kyle Lowrie / Devin Harris
Perimeter Scorers	9	Rudy Gay / Dwyane Wade / Eric Gordeon
Versatile Shooters	10	James Harden / Ray Allen / Nicolas Batum

TABLE 3. The values in the table represent the average z-score for each variable for the given cluster. Recall, positive z-scores refer to above average values and negative z-scores refer to below average values. The further the z-score is from 0, the further from the mean that statistic is.

Cluster (#)	GP	Min	% Ast	AR	TOR	ORR	DRR	Rim	3-9	10-15	16-23	3s	Stls	Blks
Combo Guards (1)	0.38	-0.02	0.61	0.20	-0.47	-1.01	-1.01	-1.32	-0.84	-0.34	0.17	1.05	-0.21	-0.78
Defensive Bigs (2)	-0.41	-0.62	0.23	-0.45	0.88	1.57	0.99	0.82	-0.06	-0.94	-1.34	-1.09	-0.36	1.11
Versatile Swingmen (3)	-0.49	-0.41	0.26	0.09	0.23	-0.08	-0.16	0.15	-0.30	-0.66	-0.66	0.00	0.67	-0.02
Floor Spacers (4)	-0.48	-0.39	0.99	-0.52	-0.82	-0.48	-0.07	-0.89	-0.81	-0.73	-0.25	1.06	-0.37	-0.31
Elite Bigs (5)	0.83	1.04	0.02	-0.47	-0.31	0.86	1.23	0.85	1.93	1.03	0.40	-0.95	-0.20	0.80
Big Bodies (6)	-0.55	-1.06	0.67	-0.60	0.33	0.85	0.52	-0.21	-0.30	-0.03	0.72	-1.03	-0.55	0.22
Active Bigs (7)	-0.08	-0.29	0.29	-0.90	-0.19	1.25	0.98	0.82	0.67	0.65	0.08	-1.08	-0.77	1.29
Ball Handlers (8)	0.04	0.34	-1.53	1.69	0.94	-0.85	-0.95	-0.12	-0.14	0.09	0.15	0.37	0.75	-0.81
Perimeter Scorers (9)	0.04	0.59	-0.56	-0.10	-0.42	-0.60	-0.23	0.16	0.36	1.09	0.90	0.21	-0.03	-0.34
Durable Shooters (10)	1.09	0.38	0.32	-0.21	-0.82	-0.58	-0.53	-0.16	-0.39	-0.28	-0.02	0.88	0.64	-0.52

TABLE 4. Cluster counts and percentage of each cluster on winning teams

Cluster	10	1	2	5	4	8	3	9	6	7
# of Players	57	85	110	86	120	172	84	120	64	77
% on winning teams	66.67%	62.35%	54.55%	52.33%	45.83%	44.19%	44.05%	41.67%	40.63%	37.66%
p-value	0.011	0.022	0.343	0.669	0.363	0.128	0.278	0.068	0.135	0.029

TABLE 5. Top ten pairs most found on winning teams

Combination	Winning Count	Total	Percentage	Increase	P-value
10 10	15	19	78.95%	12.3%	0.007
1 10	32	42	76.19%	9.5%	0.000
2 10	42	56	75.00%	8.3%	0.000
3 10	37	52	71.15%	4.5%	0.002
1 1	22	31	70.97%	8.6%	0.017
1 2	77	113	68.14%	5.8%	0.000
1 5	57	87	65.52%	3.2%	0.003
9 10	32	49	65.31%	-1.4%	0.031
5 10	38	59	64.41%	-2.3%	0.026
1 4	62	100	62.00%	-0.4%	0.016

TABLE 6. 10 pairs most found on losing teams

Combination	Winning Count	Total	Percentage	Increase	P-value
3 9	37	106	34.91%	-9.1%	0.002
4 7	34	100	34.00%	-11.8%	0.001
3 7	28	84	33.33%	-10.7%	0.002
6 9	28	86	32.56%	-9.1%	0.001
3 6	15	49	30.61%	-13.4%	0.005
7 9	28	108	25.93%	-15.7%	0.000
6 7	13	51	25.49%	-15.1%	0.000
7 7	6	24	25.00%	-12.7%	0.011
6 6	3	13	23.08%	-17.5%	0.047
9 9	15	66	22.73%	-18.9%	0.000

TABLE 7. Top 10 3-way combinations most often found on winning teams

Combination	Winning Count	Total.Count	Percentage	P-value
1 9 10	24	25	96.00%	< .001
1 1 9	28	30	93.33%	< .001
2 9 10	41	44	93.18%	< .001
3 9 10	27	29	93.10%	< .001
4 9 10	23	25	92.00%	< .001
2 2 9	32	35	91.43%	< .001
1 2 9	90	99	90.91%	< .001
5 9 10	29	33	87.88%	< .001
1 5 9	66	76	86.84%	< .001
2 4 9	76	88	86.36%	< .001

TABLE 8. 10 3-way combinations most often on losing teams

Combination	Winning Count	Total Count	Percentage	P-value
7 7 9	3	35	8.57%	< .001
1 7 7	6	73	8.22%	< .001
2 6 6	3	39	7.69%	< .001
3 7 7	7	105	6.67%	< .001
5 5 7	2	31	6.45%	< .001
2 7 7	6	103	5.83%	< .001
5 7 7	4	75	5.33%	< .001
3 3 3	1	25	4.00%	< .001
4 7 7	5	138	3.62%	< .001
3 6 6	1	36	2.78%	< .001

7. References

- [1] Christopher M. Bishop, Pattern Recognition and Machine Learning, New York, NY, 2006.
- [2] Alan J. Izenman, Modern Multivariate Statistical Techniques, New York, NY, 2008.
- [3] Chris Fraley and A.E. Raftery, "Model-based Methods of Classification: Using the mclust Software in Chemometrics", Journal of Statistical Software, vol. 18, issue 6, Jan. 2007.
- [4] Hoopdata - NBA Statistics and Analysis. Oct. 2009. Web. 13 Jan. 2012. <<http://www.hoopdata.com/>>.
- [5] Gary W. Oehlert, A First Course in Design and Analysis of Experiments, New York, NY, 2000.