Information Visualization in the NBA: The Shot Chart

Stephen Chu
University of California, Berkeley
chu.stephen@gmail.com

ABSTRACT

In this paper I describe a new interactive shot chart prototype to be used for NBA game analysis. The prototype aims to visualize the performance of specific combinations of players, a feature lacking in previous shot charts available. The tool attempts to focus particularly on how players on the court affect each other by allowing users to filter from 1 to 10 players and visualizing how individual and team performances change with the substitution of players. I discuss the shot chart's implementation, functionality, potential uses, and upcoming challenges, as well as ways to improve the collection of NBA statistics to account for player interactions.

Author Keywords

NBA information visualization, shot chart, lineup-dependent data

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

An increasing emphasis is placed on the use of statistics to analyze NBA game performance. Teams have been hiring more statisticians in hopes that this investment will grant them competitive advantages. According to David Biderman, half of the NBA teams have hired a statistician and these teams win more games than teams without statisticians [4]. However, there still are successful teams that have not emphasized the importance of statistics.

A general skepticism towards the heavy use of statistics still lingers in NBA font offices. In the recent 2010 MIT Sloan Sports Conference, some of NBA management expressed beliefs that statistics should not have significant impacts on team decisions because there has been no effective way to account for the team and player dynamics that influence the game [11].

The current state of NBA statistics focuses primarily on the individual player's production. This hinders the understanding of how players interact. The challenge in proving that statistics have significance in the NBA involves developing information visualization tools to provide evidence of how the 10 players on the court affect each other.

RELATED WORK

David Turo applied treemap algorithms to uncover important information from NBA data. The treemaps were particularly successful in uncovering outliers and in cause-effect analysis in the data [5].

Unfortunately, there has been little work done in the area of basketball information visualization, and this is reflected by the stagnant state of NBA data collection and visualizations. The statistics emphasized in today's analysis of the game are the same statistics used several decades ago. There has been little collective effort to move beyond these statistics.

APPROACH

To visualize performance of specific lineups, matchups, and team dynamics, I attempted to improve the shot chart, a common visualization used in basketball analysis that displays shooting information.

Shot Charts

The shot chart has the potential to be a useful basketball analytical tool because it can contain much of the game's important information in a single place. Also, the chart can be used to discover the strengths and weaknesses of offenses and defenses. By extending the functionality of shot charts, I hope to create a tool that will allow users to understand a team in greater depth. For example, by understanding how the team's field goal percentages change by substituting different players into the on-court lineup, we can understand how players affect the shooting percentages of his teammates.

The current shot charts available do not have the necessary functionality for in-depth analysis of how well specific combinations players work together or play against each other. Current shot charts only provide generalizations of how well individual players or teams as a whole perform, preventing the analysis of specific player lineups.

Two of the most popular shot charts are provided by NBA.com and ESPN.com. Both of these shot charts show how well an individual player or team shoots against another team as a whole, but these visualizations do not provide information for specific lineups. In the following subsections, I discuss the strengths and weakness of the NBA.com and ESPN.com shot charts.

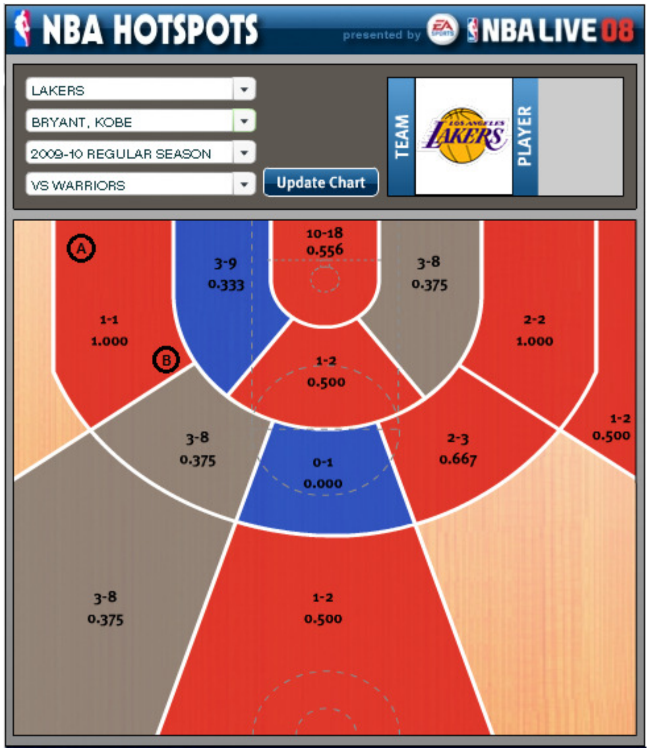

Figure 1. A shot chart for Lakers player Kobe Bryant against the Golden State Warriors in the 2009-2010 Regular Season.

NBA Hotspots

Figure 1 shows the format of NBA Hotspots, a shot chart that partitions the court and encodes values of field goal percentages by hue [8]. Using Hotspots gives users a general sense of how well a player or team shoots against another team in the regions labeled on the court. Hotspots also provides the gathered data used to calculate the field goal percentages. For example, in the sector enclosing the hoop, Kobe made 10 and missed 18 shots. This is important because it gives users a sense of how often a player or team shoots from a particular position. However, it would be difficult to mentally visualize the shot distribution with the numbers given. One improvement in my implementation of the shot chart that I will discuss later is the option to visualize shot distribution along with multiple other views.

Unfortunately, Hotspots does not show enough hue variance between different sectors for users to quickly understand the differences between values of the different regions. For example, the left-most colored sector is colored red and has a field goal percentage of 100%, while the middle bottom sector has a field goal percentage of 50% but is also colored red. If the explicit values were not written in the sectors, the user would not understand that these two positions have different likelihoods of making or missing a shot. Also, Hotspots does not provide a scale for users to know how the hues correspond to a range of values. Another way to encode value can be using boldness or transparency of the color, which is the option I chose for my implementation.

An important consideration when creating shot charts is how to divide the court into appropriate sectors. Hotspots provides very large sectors. Although Hotspots does provide information for how well Kobe shoots in a certain area, the size of these partitions makes it unclear how well Kobe actually shoots at specific positions in the floor. Looking at Figure 1, we see that the top, leftmost sector inside the arc (containing marks **A** and **B**) has a value of 100% by collecting data on one shot. This sector is particularly large and covers many positions that Kobe can shoot from. One can argue that different positions in this sector mean a great difference in the effectiveness of Kobe's shot. A shot attempt from location **A** might be very different from an attempt from location **B,** so generalizing the whole sector does not give an accurate field goal percentage when comparing specific locations in the sector. The problem caused by size of the region is worsened when combined with the lack of data for that particular sector. In the partition just discussed, there was only 1 shot recorded, which is not enough data to confidently encode that whole area as 100%. By decreasing the size of the divisions, the field goal percentages at specific spots on the floor would be more accurate.

Hotspots allows users to filter by teams or individual players and their opposing teams. However, Hotspots does not allow users to filter by specific combinations of players, so one cannot use this tool to understand how certain players work together. The best it can do to show collaboration is show how a team as a whole performs.

ESPN Shot Chart

Figure 2 shows another popularly used shot chart tool provided by ESPN.com [7]. This chart does not explicitly state field goal percentage on spots across the court, but instead marks the court as a map of made and missed shots. This shot chart shows much more detail about the specific shots that occurred during a single game than NBA Hotspots. This visualization displays exactly where each shot attempt originated. Furthermore, hovering over a shot gives users extra information about the shooter, the distance from the basket, and the time of the attempt. Because of the specific locations, users have an improved sense of how well a player shoots from a very specific location, and this shot chart also allows users to visualize shot distribution. The filtering of made and missed shoots lets users get a sense of where the bulk of successful shots were made, as well as which areas had the most misses.

Unfortunately, as seen in the figure, clustering is a problem, making it hard to single out certain shots. One the right half of the court, you can see that there are many shots that cluster near the basketball hoop. It is unclear whether all of the shot attempts are displayed, or if shots are covered underneath by others. Although filtering through players one-by-one will let users uncover most of the possibly covered shots, users are also not sure how shots that are taken at the exact same position with the same result will be displayed. Also, this shot chart visualization makes it

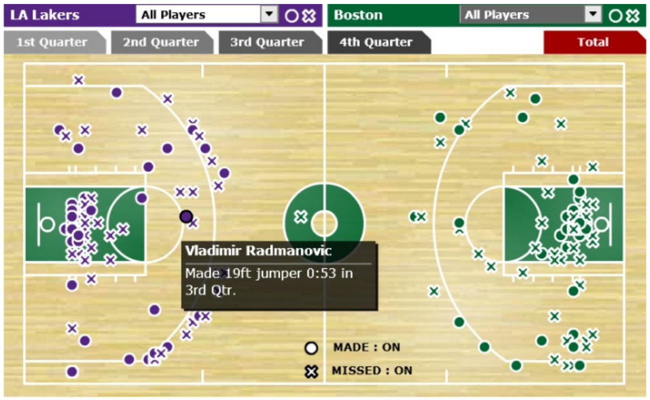

Figure 2. The ESPN shot chart for a Lakers-Celtics game.

difficult for users to figure out the field goal percentage of a certain area on the court.

Again, a problem with this visualization is that it does not allow users to filter through specific combinations of teams. In some sense it is a step ahead of NBA Hotspots because it allows us to filter by player from both teams, whereas Hotspots only allowed filtering by player from one team. ESPN Shot Chart adds filtering through quarters in the game, so it could be a useful tool in comparing a player versus player matchup in a certain quarter. However, there is no way of knowing if those players actually played at the same time, which makes the comparison more difficult to evaluate.

NBA Statistics

I will focus on the Lakers-Celtics games of the 2007-2008 regular season. I chose this dataset partly because the Lakers were predicted by 9 out of 10 ESPN analysts to beat the Celtics in the 2008 NBA finals [10]. However, the Celtics won surprisingly easily. In retrospect, it is easy to claim that the Celtics were a much better team, but I am still left wondering why the Lakers were picked as such an overwhelming favorite to win even though they had lost both of the regular season games against the Celtics in the regular season. The aid of a strong visual analytical tool will make it easier to understand why the Celtics matched up so well with the Lakers. Although the Lakers had strong individual players with great individual statistics, the Lakers team was much less effective against the Celtics than their averages on typical statistics sheets suggested.

Current State of NBA Statistics

A large problem in the way of visualizing and evaluating team performance and how players interact is the way that NBA statistics are collected today. Typical NBA statistics sheets focus on the individual and are good at compiling averages that the player has achieved throughout his career, but they are not good at accounting for the positive or negative impacts that his teammates contributed to that single player. The stats have just one primary variable, the single player who scored the point, made the assist, or caught the rebound. However, this disregards the other 9

2007-08 Lakers Regular Statistics

PLAYER AVERAGES

Player	G	GS	MPG	FG%	3p%	FT%	REBOUNDS OFF	REBOUNDS DEF	REBOUNDS TOT	APG	SPG	BPG	TO	PF	PPG
Kobe Bryant	82	82	38.9	.459	.361	.840	1.10	5.20	6.30	5.4	1.84	.49	3.13	2.80	28.3
Pau Gasol	27	27	34.0	.589	.000	.789	2.30	5.60	7.80	3.5	.52	1.56	1.63	2.00	18.8
Pau Gasol (TOT)	66	66	35.6	.534	.250	.807	2.30	6.00	8.40	3.2	.45	1.48	1.88	2.10	18.9
Lamar Odom	77	77	37.9	.525	.274	.698	2.60	8.10	10.60	3.5	.97	.94	2.03	2.90	14.2
Andrew Bynum	35	25	28.8	.636	.000	.695	3.00	7.20	10.20	1.7	.34	2.06	1.49	2.80	13.1
Derek Fisher	82	82	27.4	.436	.406	.883	.30	1.80	2.10	2.9	1.05	.04	1.10	2.30	11.7
Jordan Farmar	82	0	20.6	.461	.371	.679	.50	1.80	2.20	2.7	.94	.06	1.32	1.30	9.1
Sasha Vujacic	72	0	17.8	.454	.437	.835	.30	1.90	2.10	1.0	.50	.07	.72	1.50	8.8
Vladimir Radmanovic	65	41	22.8	.453	.406	.800	.90	2.40	3.30	1.9	.71	.18	1.12	2.30	8.4
Luke Walton	74	31	23.4	.450	.333	.706	1.10	2.80	3.90	2.9	.81	.24	1.35	1.80	7.2
Ronny Turiaf	78	21	18.7	.474	.000	.753	1.20	2.70	3.90	1.6	.36	1.38	.92	2.60	6.6
Trevor Ariza	24	3	18.0	.524	.333	.683	1.00	2.50	3.50	1.5	1.08	.33	.75	1.50	6.5
Trevor Ariza (TOT)	35	3	15.6	.507	.278	.653	.90	2.20	3.10	1.3	.89	.31	.66	1.30	5.5
Chris Mihm	23	5	12.1	.337	.000	.667	1.20	2.10	3.30	.6	.17	.61	.74	1.80	3.6
Didier Ilunga-Mbenga	26	0	7.5	.492	.000	.400	.60	1.00	1.60	.2	.15	.62	.31	1.50	2.5
Didier Ilunga-Mbenga (TOT)	42	0	7.8	.464	.000	.417	.70	1.00	1.70	.2	.17	.62	.29	1.70	2.0
Coby Karl	17	0	4.2	.346	.308	.800	.20	.60	.80	.5	.24	.12	.29	.20	1.8
Ira Newble	6	0	5.2	.333	.500	.000	1.00	.80	1.80	.5	.17	.17	.50	.80	1.2
Ira Newble (TOT)	49	13	14.2	.437	.327	.769	1.00	1.60	2.60	.4	.57	.16	.69	1.40	3.8
Team Averages	82	0	241.2	.476	.378	.769	11.0	33.2	44.1	24.4	8.0	5.3	14.1	20.6	108.6

Figure 3. A typical official sheet provided by the NBA and Lakers organization.

important variables on the court at the time each event was recorded. These 9 ignored variables are the other players sharing the floor. All 10 players on the court have a strong influence on each other, but the current focus on individual stats hides these influences. The NBA needs to move towards a new way of collecting statistics to account for the different combinations of players on the court.

Figure 3 shows a common statistics sheet provided by the NBA [9]. The information listed is based on the individual, except the team averages. It is easy to find the average production level of each player throughout the season, but one cannot deduce anything about how specific players interact with one another. This type of data is only good for making generalizations about a player, disregarding the influences his teammates and opponents have on his production.

Progress in Team-Related Data

Although the NBA's vast majority of the data is individual-based, there are other data vendors currently attempting to collect more player-combination specific data. These vendors include *82games.com, basketball-statistics.com,* and *basketballvalue.com* [1, 2, 3]. *82games* and *basketball-statistics* provide information for 2-player lineups and offer rankings of the best duos. However, using the data provided from these sources, we ignore the 8 other players on the floor.

82games.com and *basketballvalue.com* provide data for how well 5-player team lineups perform, allowing viewers to judge at some level the effectiveness of different lineups from a single team. Their data sheets provide a start for looking at how players interact with each other on a team level, but there is not enough data that allows us to look into the details of the lineups. The data provided by these vendors give us the averages of the 5-man lineup as a single unit, but they do not tell us, for example, how many points

Kobe Bryant scores when he is playing with those 4 players and the locations where he attempted his shots. The data provided still give us too general of a view of team lineups, treating them as one unit. We are still limited because we do not know the exact contributions of each player while the player is in that 5-player lineup. Also, the data does not allow us to filter by 2- player, 3-player, or 4-player lineups. In some cases, we will want to know how the team performs when a certain 3-player lineup is used and the remaining 2 teammate slots can be filled by any other teammate.

Another problem with simply computing these averages of 2-man or 5-man groups is the ignoring of the opposing lineups of players. The data provided from these sources make a generalization on the opponent, so it gives little information about how well the 5-player lineup plays against a specific opponent lineup. In my data collection for the shot chart, I chose to include the 5 opponent players as part of the data.

My shot chart implementation adds functionality that allows users to see how well a team performs by filtering through combinations of players ranging from 1 player to 10 players. But to do this, I need a more fluid dataset.

Lineup-dependent Statistics

To understand player interactions and to display them through the shot chart, I need to use lineup-dependent statistics. For each common stat recorded (in this case, a shot), I need to make sure that the shot attempt is recorded along with who made the attempt, the position of the shot, and the 9 other players that were on the floor at the same time of the attempt. Each player on the court, no matter the ability, will have a significant influence on the flow and outcome of the game, so it is not enough to collect NBA statistics based on the individual, or even based on pairs of quintuplets of players. It is important to have a dataset where each shot is recorded with at least the 10 players on the court. However, we must not treat this particular lineup as a single unit. We need to log individual events within the context of the entire 10-player lineup.

Unfortunately, there has been no standardized dataset provided that has information about the positions and times of all the shots attempted in NBA games. To collect the data through my shot chart, I went through the ESPN play-by-play logs, which are sentence-by-sentence descriptions of the events of the game. Using the accompanying ESPN shot chart, I was able to record the shooter and the position of each shot. With the play-by-play, I knew the 10 players on the court at the time of the shot.

By keeping track of the 10 players on the court while also logging specific individual contributions, I have a flexible dataset that lets me look into team interactions in a new way. I know very detailed facts, such as each player's shot attempts when the player is in a certain lineup, but I also have the ability to generalize and combine the statistics to come up with the scoring average, percentages, and shot distributions and locations of any combination of players. Also, I do not have to consider all 10 players on the court at a single time. I can, for example, select one player to focus on his contributions. Because his contributions are logged with the players he played with, I can then calculate how well the team generally performs while that single player was on the floor.

IMPLEMENTATION

The shot chart prototype was built through Processing, a programming language and environment used to create images, interactions, and visualizations. Because data collection would be more tedious, the shot chart framework was built before any data collection.

Partitioning of the court

One difficulty was deciding how to partition the court. As shown previously, NBA Hotspots chose to divide the court into large sectors. However, I decided not to do that in this implementation because the large partitions would lead to an overgeneralization of the performance in an area of the court. ESPN Shot Charts did not choose to divide the court, but instead placed each individual shot as a mark on the map of the court. However, this leads to the clustering problem, where shots can be covered by other shots. There would need to be a way to peel away layers of the marks to allow users to see all the shots. Although this way showed the exact location of the shot, this method of marking the map does not make it easy for users to calculate the field goal percentages of shot locations.

I attempted to find a balance between showing detailed shot locations and easily displaying the important numbers in the areas enclosing those locations. I decided to partition the court as grid, with each rectangle area much smaller than the partitions of NBA Hotshot.

Figure 4 shows the grid structure used for the shot chart. Each shot location in the dataset corresponds to a cell in the grid. Although the use of this grid means that my shot chart will display less detail about each shot location, each rectangle is not small enough to over generalize the position of the shot. The average rectangle is approximately 4.5 feet by 4.5 feet on the actual court, which is a comfortable enough area to take a shot, while generally not being large enough that shooting from different spots in the 4.5 feet by 4.5 feet space would make a significant difference. However, this needs to be more thoroughly researched and is a problem to addressed in future prototypes of the shot chart.

A problem with using this grid structure is that the cells along the curvature of the 3-point line do not differentiate completely between 2-point and 3-point shots. This will lead to confusion and is another problem that needs to be solved.

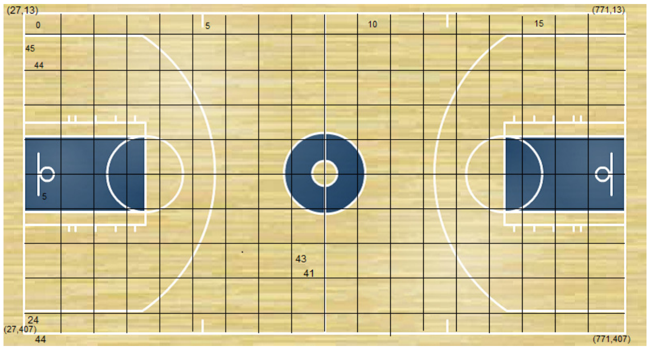

Figure 4. The locations of each shot correspond to a cell in the grid above.

Encoding values by transparency

I chose to encode values in the shot chart by transparency instead of hue. Encoding by hue does not offer a quick way to communicate to the users the values corresponding to the partitions, while the level of transparency in the rectangular box will be easier to compare areas immediately. In my shot chart, the lower the values correspond to greater transparency. A value of 0 corresponds to an alpha (opacity) value of 40 (the greatest alpha value is 255), while the maximum value displayed by the shot chart at a given time is encoded with an alpha value of 240. The alpha value for value v is equal to $40 + \frac{v}{maxValue} * 200$.

Figure 5 shows a map of field goal percentages for both teams when Kobe Bryant is in a game against the Celtics. The maximum value (100%) is labeled in a grid on the left half. I chose to not color rectangles where there was no data collected, so users would not assume that those positions on the court corresponded to some default value. Cells with a value of 0 are still colored to show that players shot from those locations.

The shot chart allows users to hover over a cell to display the value in the grid. The right half shows a rectangle with a value of 0%. Hovering over this cell shows the value (0%), the data collected in brackets [0 shots made / 2 shots attempted], and the distance from the center of the basketball hoop. Displaying the data collected partially solves the problem of clustering of shots but the details of the exact location of the shot are lost.

Handling the data

The data was taken from the Lakers-Celtics games of the 2007-2008 NBA season. The lineup-dependent statistics were stored in a HashMap. The HashMap contained entries of <*key*: 10-man lineup, *value*: collection of shot statistics>. The collection of shot statistics was implemented using an array indexed by player and the court location cell. The value retrieved were the numbers of shots made and missed by the player at the location. The majority of the values shown in the court information were calculated through array computations.

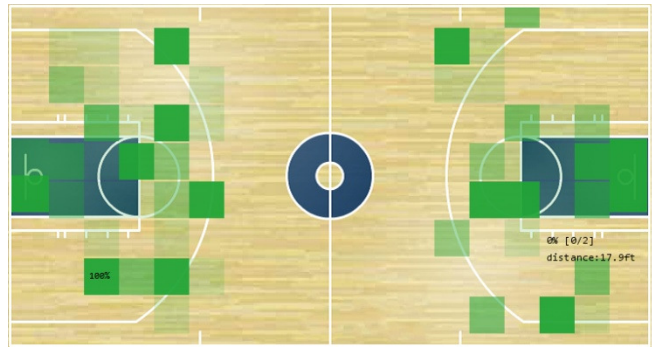

Figure 5. Each rectangle's value is encoded by opacity.

Values able to be displayed

One problem with the NBA and ESPN shot charts mentioned previously is that it is hard to get different types of important values and shooting statistics. For example, in NBA Hotshots, it is easy to determine the field goal percentage of a region, but it is difficult to visualize the shot attempt distribution of the court. For the ESPN Shot Chart, it is easier to look at the markings and get a sense of the shot distribution, but it is difficult to calculate the values of shot distribution percentages in a particular region as well as the field goal percentage of that corresponding region. My shot chart attempts to give users greater flexibility in the types of shooting data they are trying to figure out.

Field Goal Percentage and Expected Value

My shot chart has the ability to display the field goal percentage of shot locations taken by players. It is important to know how the likelihood of shots taken from particular positions in the floor changes when the lineups of players on the court changes. There is an emphasis on players to take high percentage shots. Users can switch to the expected value view to see the expected point value of a shot taken at a particular cell position. This number is proportional to the field goal percentage.

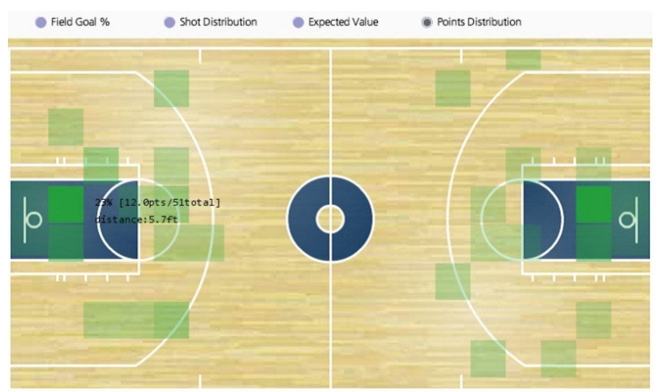

Figure 6. The points distribution map for a particular lineup of players. A large bulk of the points was scored in the low post region of court. Perhaps this explained by both teams' strong offensive post players and perimeter players who excel at slashing to the hoop.

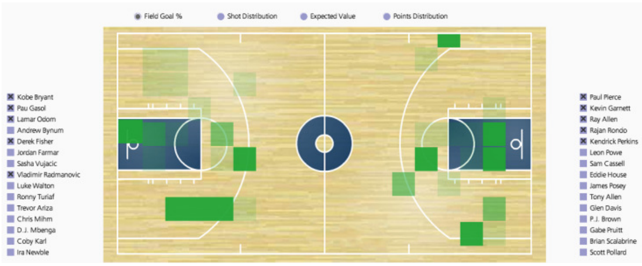

Figure 7.a.

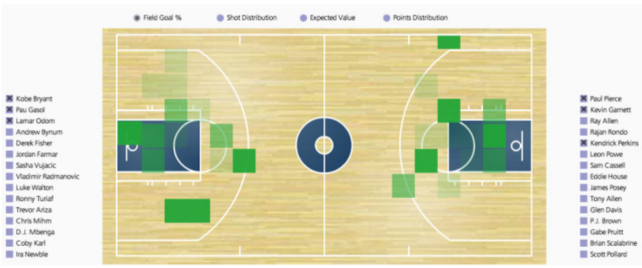

Figure 7.b.

Shot Distribution and Points Distribution

Basketball analysts and teams should also know the shot distributions and points distributions, as well as how they change along with lineup changes. The shot distribution mode of my shot chart displays where a certain player or lineup prefers to shoot. This is can be compared with the field goal percentage map to check if players are shooting more from high percentage regions.

The points distribution view shows the locations of where the points are actually being scored. It is important to know where the bulk of your points are being scored, as well as where the bulk of points are being scored against your team. This is especially important information to find weaknesses in a team's own defense.

Filtering by players

A deficiency in previous shot charts was the lack of ability to filter performance by groups of players. NBA Hotspots and ESPN Shot Chart only provide filtering for individual players or the team as a whole. The new shot chart allows filtering by players, from as many as 1 player to 10 players, 5 players maximum per team. The court reflects the information gathered when all the selected players are on the court at the same time.

The left half of the court corresponds to shots made by the Lakers and the right half contains the Celtics' shots.

Figure 7.a shows the field goal percentage information when the 10 players selected are on the court at the same time. Figure 7.b displays the information when the 6 selected players are on the floor at the same time. The flexibility of filtering through player combinations gives us a new way of looking at how team dynamics change with the substitution of players.

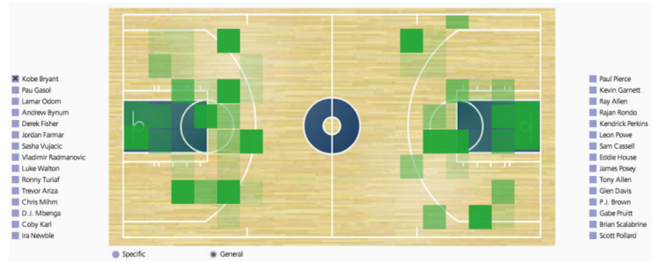

Figure 8. The general mode allows for selecting players and visualizing the entire shooting performance of both teams when the selected players are on the court.

Because we need to be able to find specific information of each individual (the shots he attempted and the corresponding shot locations) and allow the filtering of any possible lineup on the floor, we cannot rely on the datasets supplied by the sources mentioned previously. They do not offer datasets that allow for the flexibility of computing specific information for each possible lineup. The NBA provides primarily individual based stats, while the 5-player and 2-player averages provided by sources such as *82games* are missing information about the opponent lineup, as well as each individual contribution of the 2-player or 5-player combination.

The No-Stats All-Star

One new emphasis in basketball analytics is the attempt to discover players that may not be as productive individually as others, but whose teams generally benefit when they are on the court [6]. These players may not score many points but their presence greatly and positively affects their teams in ways the current method of collecting NBA statistics has failed to quantify.

To solve this problem, I added an extra mode to the shot chart that allows users to visualize how the two teams compare when the selected player or players are on the court. Instead of showing just the information of shots contributed by selected players, this mode allows for the selection of a player and the map will display the entire shooting information for both teams when that player is on the court.

In Figure 8, the shot chart configuration is in the "general" mode. Kobe Bryant is selected, and the court map shows the field goal percentages of all the Lakers and Celtics shots when Kobe Bryant is on the court. To produce this information, each lineup that includes Kobe must be accounted for, so the information displayed may contain contributions from more than 5 players for each team.

To further enhance the information provided by the shot chart, I display extra stats of the player and team performance at the bottom of the chart. The lower left half of the tool shows shooting percentages of the players selected as well as the overall team percentages when the player is on the floor. The lower right half shows stats that

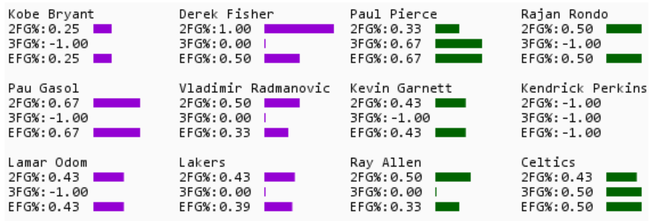

Figure 9.a. The extra information displayed on the lower left of the shot chart tool. Displays shooting percentages of the players and the team. A value of -1 means there was no data collected for that player.

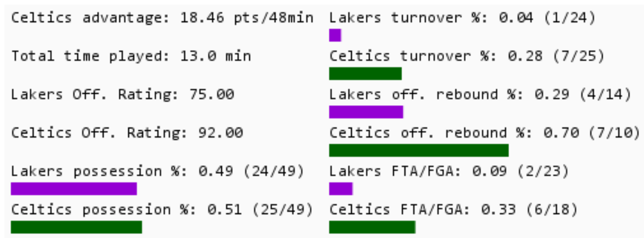

Figure 9.b. The extra information displayed on the lower right of the shot chart tool.

allow for team comparison, such as total time the lineup specified has spent on the court together, offensive rating (points scored per 100 possessions), turnover rate (turnovers per possession), and which team has the advantage by points per 48 minutes, the length of the game.

Using this extra information will make it easier to judge how teams match up with each other from a given lineup. Also, it makes it easier to find a "No-Stats All-Star," a player who generally improves his team's performance when he is on the court, even if his individual numbers are not impressive. Figures 9.a and 9.b. show the extra information displayed at the bottom of the tool.

RESULTS

The major result of this project was the development of a new type of shot chart that can visualize specific lineup matchups, allowing users to take a closer look into how players interact with each other. This shot chart has the power to be very specific by displaying individual contributions in certain lineups, but also has the ability to combine data to generalize the entire team's performance when the selected players are on the court. This shot chart demonstrates the variety of information able to be displayed by shot charts, which have historically been too simple for detailed analysis of NBA basketball.

The new shot chart prototype provides the flexible filtering of 1 to 10 players. Previous implementations of the shot chart were limited in their ability to filter between different groups of players and usually only allowed viewing individual player stats or the whole team's stats.

The tool demonstrates the ability to show different types of values using shot information such as location and whether or not the shot was made. Previous shot chart implementations do not allow easy viewing of different types of shot data. For example, NBA Hotspots do not show detailed shot locations and shot distributions, while ESPN Shot Charts do not display field goal percentages. In my prototype, users have multiple views of shot information. These views include field goal percentage, expected value, shot distribution, and points distribution.

The tool finds a balance between shot location specificity and clustering by portioning the shot chart into the small cells and encoding values by opacity.

The aim of the tool is to provide analysts an avenue to evaluate NBA teams with a tool that gives a more in-depth look at the performance of specific lineups on the court. Using this tool, you can see how team performances are affected by substituting players in and out of lineups. Using the field goal percentage and expected value, users can see how players' field goal percentages change at different locations on the court. Using the shot distribution, users will be able to track how the shot location preference changes with different players, and with the points distribution, users find where the bulk of the points are scored.

Using this tool can uncover strengths and weaknesses of offenses. For example, it is important for a team to spread their offense across the court. The user can use the shot distribution view to see if the team appropriately spreads the floor with its shots. With the field goal percentage and expected value view, the user can figure out the areas of high percentage shots when a certain player combination is on the court.

This tool also helps evaluate team defense. One can use this tool from the opponent's angle. For example, looking at the points distribution view for the opposing team will show the user weaknesses in his or her team's defense. This tool can uncover the most common locations where the opponent scores, which exposes deficiencies in defending these areas.

This tool is also an attempt to allow teams to change their lineups based on the opponent player's on the floor. Using this tool can give us information on which lineups match up best with the current opposition. However, using only this tool is not enough justification to make any major substitutions. This tool is meant to supply additional information to team management, and is not supposed to be the primary source for decision making.

DISCUSSION

The work done for this project shows that there is still much progress to be made in the area of NBA information visualization. The NBA's methods of data collection need to change if statistical analysis of game performance wants to take into account team and player interactions. Statistics based on the individuals can only go so far in helping evaluate the game.

Furthermore, the current visualizations provided by the most well-known and used sources for NBA information are not sufficient for visualizing important shifts in performance between different player combinations on the court. The NBA and ESPN shot charts discussed in the paper only give overly generalized information about the individual players and teams as a whole. They ignore the fact that the game always has 10 important players that affect each action and event in the game, no matter the ability of the player. The mere presence or absence of players can have a significant impact in matchups, so there need to be visualizations that give users the ability to focus on these lineup-dependent statistics.

There is much more work to be done to improve visualization tools for NBA analysis. Tools need to be able to visualize a vast variety of data in one place. The developed prototype shows that with current shot location data taken from the ESPN shot chart and ESPN play-by-play logs can generate at least four different views of shot information can be produced. This is only for one type of event in the game, so certainly our tools can be extended and improved greatly. Increasing the flexibility and adding functionality to visualizations such as simple shot charts can make them much more useful.

There has been some effort to change the way NBA statistics are collected. Some of these contributors such as *82games* were discussed. However, the statistics they provide are still lacking in details. They have moved in the right direction by collecting data for 2-man and 5-man group combinations. However, their data still generalizes by treating these combinations as one unit, rendering users unable to pick out the individual contributions made in the player combinations. Also, their data does not take into account for the 5 opposing players sharing the court.

There has been a growing emphasis on analyzing statistics in the NBA, despite current skepticism. Unfortunately for the public, NBA teams keep any statistics-related developments private to keep a competitive edge. This privacy between NBA teams has most likely led to the slow progress in NBA statistics to incorporate lineup-dependent statistics. It is possible that an NBA team has already made significant strides in the way it collects data but this information is not available to the public. However, if the articles of professional NBA sport writers and analysts are any indication, lineup-dependent statistics are still not used in the evaluation of player and team dynamics. There is still a great focus in the sports community on individual player stats, and there needs to be a shift of focus towards lineup-dependent statistics if the sports community is interested in developing new ways to analyze the game.

FUTURE WORK

There are many extensions and improvements to the system that must be made.

Improved partitioning of the court

Currently, the court is treated like a grid of roughly 4.5 feet by 4.5 feet rectangles. However, I need conduct more research to answer the following questions: Is 4.5 by 4.5 feet enough space for a player to take a shot? Should the region's area be decreased? Should the court be divided into discrete sectors? During data collection, there were times when a shot would fall along the edges of multiple cells, so perhaps it is better to treat the court on the pixel level.

Furthermore, the rectangular grid structure is not accommodating to the three-point line curve. In some cells, there is an overlap between 2-point and 3-point shot locations, so future versions of the shot chart need to make clear the boundaries between these two types of shot locations.

Improved encoding of values

The prototype doesn't standardize the encoding of values. For example, the alpha value is proportional to the maximum value of the current data being displayed on the court map. Therefore, the max alpha value could correspond to a value of, for example, 50%, but once the user deselects the player and new information is displayed, the max alpha value might correspond to a value different from 50%. This could confuse users initially, and without hovering over the cells to display the exact information contained in the location, users might assume incorrect information.

Improved methods of data collection

Currently, the prototype uses data from the Lakers-Celtics games of the 2007-2008 season. Unfortunately, not all of the data from these games was included. I had to manually collect the data by referring through the ESPN play-by-play logs to make sure that I knew which 10 players were on the court at the time of a shot. I had to refer to the ESPN Shot Chart for each shot recorded in the play-by-play to find out the shot location. This proved to be an extremely slow process. It would be much easier if companies with the resources could record lineup-dependent statistics. However, making this shift requires much more work for game stat recorders, and there needs to be a way to standardize this type of data.

Furthermore, the current dataset provided by official sources is lacking in many details. Exact shot coordinates have still not been provided. Important factors such as which players guard each are also not recorded.

Verification of the tool's usefulness

Because of the lack of data, I could not determine the tool's actual and potential utility. Solving this problem will include improving the tool based on feedback from NBA teams. Then, the tool must be evaluated and used by NBA teams to verify if it has any significant worth.

One argument against the heavy use of statistics in the NBA is that teams do not always get to play each other for enough time to make accurate judgments from the data. Team rosters change rapidly, even within the season. Some teams play against each other only two times each season. This may not be enough time to gather information about the specific lineup matchups to produce convincing evidence that influences team management decisions. Each game is only 48 minutes, so I must research how useful data is if it is gathered for a certain amount of time. Furthermore, there will most certainly be some lineups and players that are rarely used. Lineup-dependent statistics and tools such as the one developed will be less useful for players who play a limited amount of time.

ACKNOWLEDGMENTS

Many thanks to Maneesh Agrawala for teaching CS294-10: Visualization in the Spring 2010 term.

REFERENCES

1. 82games. http://www.82games.com/.
2. Basketball-statistics.com. http://basketball-statistics.com/.
3. Basketballvalue.com http://basketballvalue.com/index.php.
4. Biderman, David. Are Statheads the NBA's Secret Weapon? http://online.wsj.com/article/SB10001424052748
5. D. Turo. Hierarchical Visualization with Treemaps: Making Sense of Pro Basketball Data. In *Conference on Human Factors in Computing Systems.* 441-442. 1994.
6. Lewis, Michael. The No-Stats All-Star. http://www.nytimes.com/2009/02/15/magazine/15Battier-t.html?_r=1
7. Los Angeles Lakers vs. Boston Celtics - Shot Chart - November 23, 2007 – ESPN. http://scores.espn.go.com/nba/shotchart?gameId=271123002
8. NBA Hotspots. http://www.nba.com/hotspots/
9. NBA Lakers Stats. http://www.nba.com/lakers/stats/.
10. NBA Playoffs 2008- NBA Finals - Boston Celtics vs. Los Angeles Lakers – ESPN. http://sports.espn.go.com/nba/playoffs2008/series?series=lalbos
11. Sloan Sports Conference 2010. http://sloansportsconference.com/2010/