NBA Shot Data Visualization

JOEL CHRISTENSEN, Utah State University, USA
SAM JOHNSON, Utah State University, USA
HAGEN LARSEN, Utah State University, USA
MARSHAL TAYLOR, Utah State University, USA

Additional Key Words and Phrases: NBA, basketball, visualization, data, interactivity

ACM Reference Format:
Joel Christensen, Sam Johnson, Hagen Larsen, and Marshal Taylor. 2022. NBA Shot Data Visualization. 1, 1 (March 2022), 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 MOTIVATION, OBJECTIVE, AND TARGET AUDIENCE

Basketball is a game filled with data. In modern professional basketball, computer vision allows player tracking by the second and produces a wealth of data that is consumed by players, coaches, pundits, and fans daily. [6]. The visualization of such tracking data has become a multi-billion dollar industry across sports, but dynamic sports such as basketball especially need intelligent visualizations [2]. In this paper, we introduce our tool that interprets such rich data and displays it intuitively and powerfully to potential users.

Our visualization is designed for use by individuals already familiar with NBA basketball. However, users may vary widely in motivation and purpose. Below, we cover different groups that may find utility from using the tool.

1.1 Audience: Casual Users

Casual NBA fans can use this tool to learn how effective their favorite players are. With the increasing popularity of sports gambling, they may be searching for good betting or fantasy basketball opportunities. The tool could help them answer questions like:

- How is *player x* performing this year? How is he performing relative to other players in the NBA?
- Where is *player x* taking his shots from? Where is he most effective?
- What trends does *player x* exhibit throughout a typical game? How does his shot selection change?

Authors' addresses: Joel Christensen, christensenjo14@gmail.com, Utah State University, Logan, Utah, USA, 84321; Sam Johnson, sam.johnson1357@gmail.com, Utah State University, Logan, Utah, USA, 84321; Hagen Larsen, A02254637@usu.edu, Utah State University, Logan, Utah, USA, 84321; Marshal Taylor, mptaylor777@gmail.com, Utah State University, Logan, Utah, USA, 84321.

1.2 Audience: Coaches

Coaching staffs should also find our interface useful. Over 75 percent of NBA coaches already use live tracking data to guide players through practices and games [5]. Coaches use player tendencies, field goal percentages, and other statistics to make decisions on who will play, what sets their team will run, and how they will defend other teams and players.

This type of game-planning is obviously useful to NBA coaching staffs at the highest level, but our interface could also be useful to lower-level basketball coaches. By analyzing NBA players, lower-level coaching staffs might learn from the best and better emulate the way that NBA coaches make decisions. Our interface may help answer the following questions:

- Where on the floor does *opposing player x* have the highest field goal percentages?
- Should we try to force *opposing player x* to drive right or left, based on his tendencies and field goal percentages?
- What plays should we call?

1.3 Audience: Professional Team Management

Front office staffs throughout the NBA and G-League will also find value in the visualization interface. The General Manager and other executives must manage their teams within specific parameters established by the NBA. The NBA promotes parity through a salary cap, limiting the amount of money that can be distributed to players each year. General managers must also pay careful attention to the style and tendencies of their players, making sure the team plays complimentary basketball. It's the GM's job to put together a winning team through signings and trades. The visualization will help front offices of all levels answer questions such as:

- How does the on-court production of *player y* compare to other players who earn similar salaries?
- What are *player y*'s skills? Are they complimentary to the other players' skills on the team?
- Do *player y*'s trends and skill sets fill a hole in our roster as it is currently constructed?

1.4 Motivation: NBA Revenues

In figure 1 we show total NBA revenue over time. We can see that NBA revenue steadily rose up until 2020 when it dropped during the pandemic [1]. In any case, revenues are still very high. The NBA is a multi-billion dollar industry. In general, the better a team plays the more revenue it will generate. This serves as a huge incentive to analyze data and gain insights to improve overall team performance. This demonstrates another key motivation to using our tool: increasing revenue.

1.5 Justification

Related visualizations like BKViz already exist and also display basketball information in a visual format [3]. Their visualization is great for displaying team-level performance, and high level statistics via a global filter. However, their visualizations seem to be disjoint from one another, and the global filter appears to take up an entire page.

We hope to further their work by including all of our visualizations on a single page, with a unified and collapsible filter. Our visualization also has the unique goal of visualizing data from a player-centric perspective, rather than a

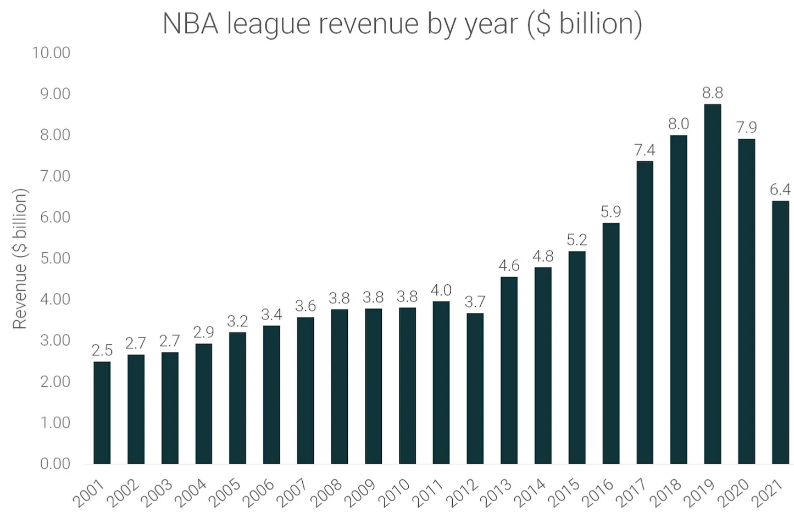

Fig. 1. League Revenue by Year

team-centric view. This will increase the granularity of the data analysis by allowing users to focus on a single player at a time.

1.6 Summary

We have described the motivation for our tool based on use cases for casual fans, coaching staffs, and professional team management. Similar tools, such as TenniVis, have proven successful in analyzing tennis performance [4]. Our tool will provide the same level of analysis for individuals in *team* sports. The entirety of ways in which the tool might be used are too vast to list. Our interactive data visualization will attempt to satisfy basketball savvy users of all interests and motivations in a meaningful way.

2 IMPLEMENTATION SUMMARY

The visualization was implemented as a web application consisting of a backend and frontend. Similar to iTTVis [7], data processing takes place in the back-end and interactive visualization takes place in the front-end. Different from iTTVis, however, the data was retrieved once using the nbastatR package and then stored in an SQLite database. Django was used as the backend framework. The backend service responds to front-end requests with data matching the user-selected criteria.

HTML, native JavaScript, and D3 were used to create the frontend which included the player selection & filtering tool and the several visualizations.

3 DATA PROCESSING

3.1 Data Source

The data we visualize is sourced from the nbastatR package for R. nbastatR pulls its data from a variety of statistics sources including the NBA Stats API, Basketball Insiders, Basketball-Reference, HoopsHype, and RealGM. These sources store a breadth and depth of statistics using both traditional stat recording techniques as well as computer vision methods such as those developed by Second Spectrum.

By leveraging nbastatR's powerful API we are able to focus on creating value through visualization and keep data processing to a minimum. In order to efficiently connect the R package data to our visualizations, we have developed a Django back-end with an SQLite database. Users can seed the player and shot data, then continue to use the interface without waiting for data to load.

3.2 Schema

The data we focus on comes from nbastatR's team_shots() function, which returns a dataframe with attributes representing shot actions taken during NBA games. Each call to team_shots() requires a date range and a season type (regular season or playoffs).

Each item has attributes including the name of the player who took the shot, their team, whether the shot was made or missed, what type of shot action it was, whether it was a two or three point attempt, the date the game was played, the current period of the game, the time remaining in the game, and the court x and y coordinates of the shot. By having all of this data available to us, we are able to create several different visualizations that provide analytic value to the user.

3.3 Data Processing

As mentioned, data processing is a combination of R function calls, some Python data manipulation in a Django back-end, and final data formatting in the JavaScript front-end using D3. Because of the large amount of data obtained from the R package function calls, some initial setup is necessary by the user to seed the data. The use of a Django back-end that handles this data loading and seeding, however, provides a robustness to our interface and allows it to cut out further loading time as the application is being used.

Some amount of data cleaning and validation was necessary to present high-quality visualizations. Tuples with significant missing data and unnecessary attribute columns were discarded. Data is filtered based on the user's requested player and time span.

4 OVERALL VISUALIZATION

In figure 2 we show what the interface looks like before a player is selected. We can see that each of the four visualizations has been initialized but displays no information until after a player is selected.

In figure 3 we display the interface after Trae Young has been selected. We can see that each of the four visualizations has now populated with data for the selected player and date range.

5 PLAYER SELECTION & FILTERING TOOL

In figure 2 we presented the interface before any plotting has been done. The user can click on the menu icon in the top left corner to open up the *Player Selection & Filtering Tool* (figure 4).

Other sport visualizations have used central filtering to great success [4]. TenniVis included a large filtering panel on the left hand side of the application. We elected to hide our filtering panel behind a collapsible menu, aiming to decrease

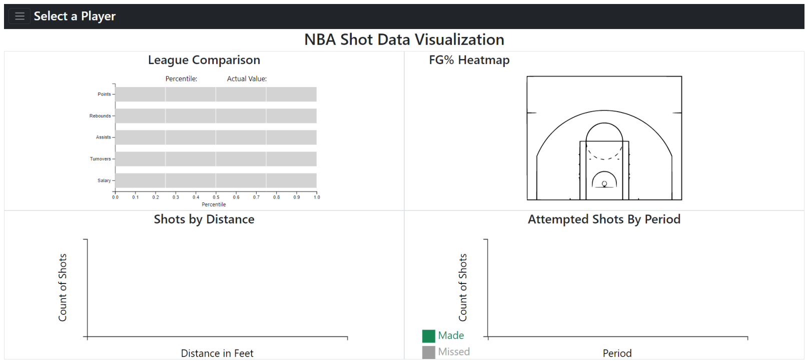

Fig. 2. Interface Before Player Selection

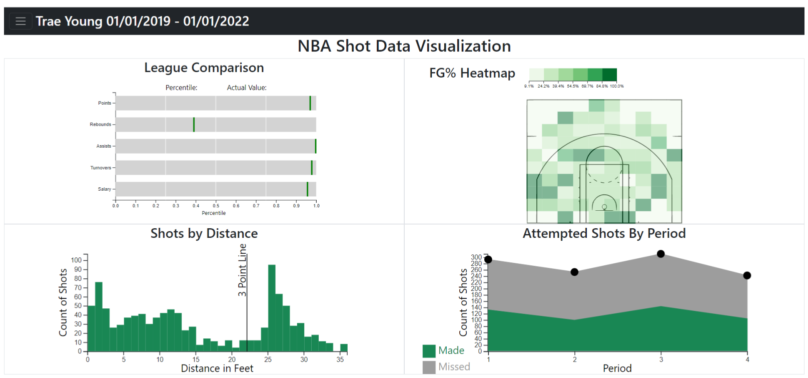

Fig. 3. Plotted Interface

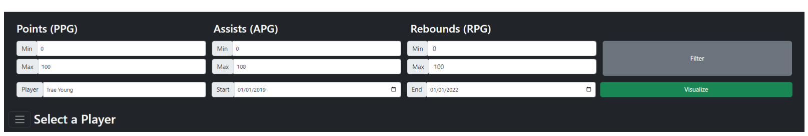

Fig. 4. Player Selection & Filtering Menu

distractions within the visualization window.

The first thing to note is the *Player* search bar. You can click in the in the search bar and the tool will automatically display a list of players that the user can select. As the user searches for a player, that drop-down list will be updated (think Google search), showing all player names that match what has been typed so far.

After the user has selected a player, their attention should turn to the two date selection tools, *Start* and *End*. The user can change these dates to modify the time period we are looking at. Expanding the date range would result in more data for the visualizations while reducing the date range would result in less data.

The user's attention should now turn to the *Min* & *Max* integer input fields below each of the headers *Points (PPG)*, *Assists (APG)*, and *Rebounds (RPG)*. The *Min* and *Max* fields are set by default to 0 and 100 respectively and can be changed by the user.

6 INTERFACE VISUALIZATIONS

6.1 League Comparison Plot

In figure 3 we displayed the interface after a player and date range had been selected. In the top left corner we show the *League Comparison Plot*. It effectively compares the *Points*, *Rebounds*, *Assists*, *Turnovers*, & *Salary* of the selected player over the selected period with all other players over the same period. A green bar is given at this player's percentile over the data distribution of all players. Additionally, the user can hover over the plot and the selected percentile as well as the raw values will be shown above the plot.

This tool could be very useful when attempting to compare a player to the rest of the league. It might help answer questions like "How much is this player being paid and do his stats back up such a highs salary?" The inverse question would be "Why is this player being paid so little when his stats are so impressive?" This could lead to a drafting or trading opportunity.

6.2 Shooting Heatmap

The heatmap visualization is displayed in the top right of the visualization interface. This idea has some precedence in soccer, where the spatial position of players on the field has been visualized and analyzed to determine the best formations [8]. We created an adaptation for the sport of basketball where the court is drawn and a shooting heatmap is overlaid with some opacity on top of the court. The field goal percentage is represented through shades of green. Additionally, when a user hovers over a square on the court, it shows the field goal percentage above the plot.

This visualization could be quite useful for a coach to quickly see how his players are doing from different locations. If he is able to locate a location where a player under-performs, he could suggest the player practice shooting from this location in order to improve. Conversely, if he finds a location where a player performs extremely well from a given location, the coach might create new plays that give this player opportunities to shoot from this location.

6.3 Shots by Distance

Next we refer to the *Shots by Distance* visualization, which can be seen in the bottom left corner of figure 3. It simply counts the number of shots that have been taken at various distances from the basket.

This could be useful to gain insights into what type of offensive player someone is. Some types of players might shoot only three pointers while others might take shots only close to the basket.

6.4 Shot Frequency by Quarter

Next we refer to the *Shot Frequency by Quarter* visualization, which is shown in the bottom right corner in figure 3. To display this figure, we take all of the shots over the selected player and time frame and separate them into the four quarters of the game. Then, we plot the number of made shots and missed shots for each quarter through two line plots. By hovering over the black dots, we can see the actual raw values that are being plotted.

This visualization could be useful to a coach who would like to see how a player is performing across different quarters. The visualization might answer questions like, "Which quarters does this player shoot the most shots in?" or "Which quarters does this player have the best and worst shooting performance?"

6.5 Summary

Our visualization is an intuitive, web-based application that shows its users four different visualizations in a clutter-free panel format designed to provide key insights into NBA player basketball data.

A collapsable player selection & filtering menu is also included in the visualization, giving users an easy, convenient, and intuitive way to filter the visualization, each of which is tied to the same user-selected player.

7 VISUALIZATION DISCUSSION

7.1 Limitations

One of the major limitations for this visualization lies in its inability to compare two or more players directly. Improvements could be made to select multiple players from the player selection menu, and encode their data with a categorical channel in the same reference frame. This could help front offices across the league feel confident when a roster decision comes down to one of a few players.

Another major limitation is the inability to view the same statistics for a more general group, such as position groups or teams. Implementing a different selection menu could enable us to display the same information for an entire team, as opposed to an individual. This would help coaches understand the team's weaknesses and strengths in greater detail. We could also allow for the selection of other groupings, such as starting lineups, so coaches and staff can make the best decisions about which players should play on the court together at the same time.

A third limitation is the inability to display historical players within the context of the era they played in. Throughout the history of the NBA, league-wide tendencies have evolved and changed. For example, there are generally more 3-point shots taken today than in years past. Contextualizing an individual within the era they played in helps us understand the true impact of a player relative to their peers. A player taking 5 3-pointers per game in the 90's would've been the highest-volume 3-point shooter in the league. Today, 5 3-pointers per game is closer to the league-wide average.

7.2 Lessons Learned

We learned several lessons during the project. One major lesson learned was related to the structure of our data points. A major oversight that inhibited our ability to represent historical players within their appropriate context was the way in which we structured players in our database. Statistical averages such as PPG, APG, and RPG were represented as attributes on a player record, which minimized our ability to view statistical averages for a specific time frame. We learned that how we structure data greatly impacts our ability to accomplish our objective.

We also learned about the need to keep administrative functions out of the way of the visualization. Our original player selection pane was in the middle of our four visualizations. When we presented the project that way, we received feedback to encapsulate the player selection pane to an expandable menu, which greatly improved the clarity of the visualization.

8 INSTRUCTIONS TO RUN

Begin by connecting to the USU network and navigate to http://vizus-toy.cs.usu.edu/2022_cs5820/nba-data.

Expand the player selection menu by clicking the menu icon in the top left corner of the screen. Optionally, enter a minimum and maximum value for Points Per Game (PPG), Assists Per Game (APG), and Rebounds Per Game (RPG). After entering the search criteria, click 'Filter'. Clicking on the player name input, you will see a drop-down list appear. The names in this drop-down list are restricted to those players matching the criteria you entered.

Select a player from the list that you would like to visualize. Identify a date range to restrict the time-frame of data points displayed, or use the default values. Click 'Visualize', and collapse the player selection menu by clicking the same icon you used to open it.

You will now see that the visualization has populated with data points for the selected player. Explore the visualizations and observe the interactions that display as you hover over different data points.

Tip: If you are struggling to find players with interesting visualizations, try looking at data for Stephen Curry, Rudy Gobert, Royce O'Neale, Bojan Bogdonovic, or Luka Doncic.

REFERENCES

[1] Dimitrije Curcic. Nba revenue statistics (2001-2022), Nov 2021.

[2] Raffay Hamid, Ram Krishan Kumar, Matthias Grundmann, Kihwan Kim, Irfan Essa, and Jessica Hodgins. Player localization using multiple static cameras for sports visualization. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 731–738, 2010.

[3] Antonio G. Losada, Roberto Therón, and Alejandro Benito Santos. Bkviz: A basketball visual analysis tool. *IEEE Computer Graphics and Applications*, 36:58–68, 11 2016.

[4] Tom Polk, Jing Yang, Yueqi Hu, and Ye Zhao. Tennivis: Visualization for tennis match analysis. *IEEE transactions on visualization and computer graphics*, 20(12):2339–2348, 2014.

[5] Yusuke Sano and Yohei Nakada. Visualization for potential pass courses and quantification for offensive and defensive players in basketball. In *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, pages 1–6, 2021.

[6] Egor F. Smirnov, Nikita S. Ivanov, Maksim Zavgorodnii, Ivan Yu. Khramov, and Victoria V. Chernyshova. Recognition of the basketball players position using live cameras. In *2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*, pages 695–698, 2021.

[7] Yingcai Wu, Ji Lan, Xinhuan Shu, Chenyang Ji, Kejian Zhao, Jiachen Wang, and Hui Zhang. ittvis: Interactive visualization of table tennis data. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):709–718, 2018.

[8] Yingcai Wu, Xiao Xie, Jiachen Wang, Dazhen Deng, Hongye Liang, Hui Zhang, Shoubin Cheng, and Wei Chen. Forvizor: Visualizing spatio-temporal team formations in soccer. *IEEE transactions on visualization and computer graphics*, 25(1):65–75, 2018.