

Sports Data Visualization

Group 1: Penelope Lialios, Isabelle Gallucci, Kshitij Jhunjhunwala, Sid Valecha

Shiny app link: <https://sid-valecha.shinyapps.io/project/>

GitHub repository: <https://github.com/sid-valecha/stat436-project>

Introduction:

Most sports teams already use statistical data to support their game plans going into big games to pinpoint areas of strengths and weaknesses. They also use data analytics to scout their opponents to prepare the team for whatever is about to come their way. The main problem, though, is how this data is visualized for coaches and players. Management, coaches, and players don't have the time and energy to listen to random statistical methods and numbers but would rather have the information laid out in front of them so they know the bottom line of what they need to do. This is especially important in the NBA where there is an immense amount of data available from each player, for each stat, for each season, and from each team. By focusing on presenting the information in a digestible format, the players and coaches can focus on the task at hand - winning their games. Our focus is on the Milwaukee Bucks, and our goal is to help them develop strategies and pinpoint areas of interest to focus on in the future.

Our solution to this data visualization problem is to focus on a select few important statistics to give the team a starting point in their game planning. Our main goals were to:

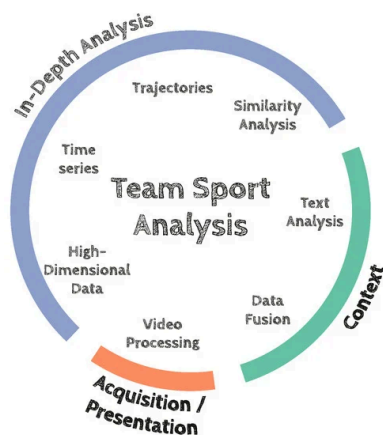
1. Identify which players are overperforming or underperforming on the Bucks
2. Identify trends in offensive and defensive performances of opponents
3. Highlight areas for improvement

Focusing on these goals will allow our team to both prepare for upcoming games as well as get insight into their own game. Since there are so many statistics in basketball, as well as large rosters, there is a challenge in how to make these visualizations concise and understandable to everyone. With these goals in mind, we narrowed down the areas of focus on our datasets to Offensive/Defensive Rating and Individual player shooting and efficiency statistics. Our job is to simplify complex data so that it is understandable and easy to act on. By doing this, we can help the Bucks improve their game plans and perform better on the court.

Literature Review

After conducting a literature review, there are an increasing number of challenges related to sports visualization as the advancement of technology continues. The main challenge is the vast amount of data spanning from different leagues, schools, and countries (Bai, 2021). Narrowing this data down into what is relevant and taking the time to develop the proper visualizations around said data has been a challenging task to complete. In basketball specifically, player performance data has become a crucial tool for teams to use during game preparation. The big role of uncertainty and luck in basketball makes it harder for analysts to make use of this data (Sarlis, 2020).

Acquiring the data is the main step in starting your sports data analysis. These datasets are extremely diverse and complex which means that they each have different purposes in the grand scheme of things (Bai, 2021). Determining your goal for the analyses is a good way to start because then you can search for datasets that best suit your purpose. Once you have the ideal dataset, finding the right analysis method is crucial. There are many you can choose from, like statistical analyses, network analysis, and motion data analysis. Choosing the right one can ensure you get the most information out of your dataset (Bai, 2021). Different analyses can be done based on the data available. Usually, teams want to know the “why” behind each performance instead of just what happened. Scouting and training would be a lot more effective if you knew why each player did what to help guide their team moving forward (Stein et al, 2017). Traditionally, analysis has purely focused on what happened so to find out why a player did a certain thing, more in-depth analyses need to take place. Stein et al offer a thorough figure that outlines the steps of team sport analysis below.



As you can see, acquiring and presenting the data is the first step in which you can scout the players and gather the data. Usually, most data analysis experts will have the data already

gathered and use the datasets created for the team scouts. Next, for context in the analysis, you merge datasets and offer background about the team and players. Lastly, the in-depth analysis uses many techniques to provide the team with as much information as possible to help them tweak their gameplans and strategies moving forward (Stein et al, 2017).

Basketball has a lot of uncertainty because so many players go in and out of the league and have performances based on age and injuries. There is usually enough past data to be helpful but the most important statistics should be ones that reflect their impact based on league averages or how efficient the player is (Sarlis, 2020). Sarlis has identified many useful metrics that we will try to use in our visualizations as well such as player impact estimate, net rating, and performance index rating (Sarlis, 2020). All of these statistics give an overview of a player's statistical performance which is useful for the analysis at the end.

In the end, we will use this background to make a statistical visualization of the Milwaukee Bucks as a whole and how they compare to other teams in the league. This will include using statistics that accurately reflect their impact in an easily digestible way for the teams to understand. The main problem with sports data is the enormous amount of variables and datasets that make the interpretations of the games so difficult. With increased knowledge about the variables and different statistical analyses, data visualization can be an easier way to make the information concise and understandable.

Design

For our research synthesis, while translating the abstract concepts from our literature review into practical applications for the Milwaukee Bucks, we focused on creating visualizations that not only display data but allow for interactive exploration and understanding of complex datasets. This approach helps in resolving the issues identified in the introduction by making the data actionable and insightful.

Design Implementation

1. Player Performance Analysis

- **Scoring Trends:** This line graph tracks each player's scoring over time, enhanced by a linear regression trendline to indicate performance trends. This visualization

aids coaches in identifying periods of high or low performance, which is crucial for making strategic decisions about player training and game-time utilization.

- **Shooting Efficiency:** A density plot displays distributions of Field Goal Percentage (FG%) and Three-Point Percentage (3P%) for selected players. This plot helps in evaluating shooting skills and pinpointing players who excel or need improvement in specific shooting metrics.
- **Points Per Minute (PPM):** Showcasing a player's efficiency, this line graph represents how many points a player scores per minute played. It is vital for assessing the impact of players during their time on the court, especially useful for rotations and understanding player productivity in real game scenarios.

2. Team Performance Comparisons

- **Offensive and Defensive Ratings vs. Opponents:** These line charts display the Bucks' offensive and defensive ratings compared to those of selected opponents. The interactive nature of these charts allows for real-time adjustments based on the opponent selected, providing strategic insights into the Bucks' performance relative to their competitors. This comparison is crucial for pre-game strategy and post-game analysis, helping to highlight strengths and areas for improvement against specific teams.

Visual Design Choices

The following design choices were made to ensure effectiveness, clarity, and user engagement:

- **Interactivity:** Utilizing Shiny's interactive capabilities, users can dynamically select players and opponents to instantly see updated data, making the tool flexible and responsive to different analytical needs.
- **Clarity and Accessibility:** Through careful selection of colors (e.g., team colors for the Bucks and contrasting colors for opponents), our visualizations are not only aesthetically pleasing but also easy to understand at a glance. Clear legends, annotations, and a

consistent visual theme across all plots ensure that the data is accessible to all users, regardless of their data literacy levels.

- **Contextualization:** By embedding season boundaries and using trend lines in our plots, we provide context that helps in making sense of the data over time, enhancing the strategic value of the visualizations.
- **Consistent Branding:** To maintain brand consistency and enhance team spirit, we used the Milwaukee Bucks' official color palette throughout the visualizations. This not only reinforces team identity but also ensures that the visualizations resonate more with the team management and players.

Conclusion

Sports analytics is a rapidly growing field with the power to unlock information about players and teams, enhance their performance, and self-evaluate what has yet to be available in previous years. Although analytics are not an absolute reflection of a player's talent or a team's chances of success, they can be useful if utilized correctly, especially when including well-thought-out, clear visualizations. Our application was designed to do exactly that: to equip our client, the Milwaukee Bucks, with the most accurate information possible to prepare their roster and coaching staff for their future games against other NBA opponents. Our visualization for team analysis should be seen as a valuable starting point for the coaching staff in particular to compare themselves to different teams and overall league performance alike, allowing them to take away the strengths their strategies could highlight or weaknesses they can expose. Further, readers should view the efficiency statistics for Bucks players as a guide for the athlete's performance history on both sides of the court and a reference to see who could pose the best threat to the opposition's greatest attributes. Through interactive visualization techniques and informative statistical findings, these visualizations could be useful tools for player and team success.

References

- Bai, Z., & Bai, X. (2021). Sports Big Data: Management, Analysis, Applications, and Challenges. *Complexity*, 2021(1). <https://doi.org/10.1155/2021/6676297>
- Sarlis, V., & Tjortjis, C. (2020, November). Sports analytics -- Evaluation of basketball players and team performance. *Information Systems*, 93. <https://doi.org/10.1016/j.is.2020.101562>
- Stein, M., Janetzko, H., Seebacher, D., Jager, A., Nagel, M., Holsch, J., Kosub, S., Schreck, T., Keim, D. A., & Grossniklaus, M. (2017). *How to Make Sense of Team Sports Data: From Acquisition to Data Modeling and Research Aspects* (1st ed., Vol. 2). Geospatial Data. <https://doi.org/10.3390/data2010002>