

# Why did they win? Visualizing NBA teams across multiple seasons

**Yannick Boesmans**

KU Leuven

yannick.boesmans

@student.kuleuven.be

**Yannick Laevaert**

KU Leuven

yannick.laevaert

@student.kuleuven.be

**Yves Langerlaert**

KU Leuven

yves.langerlaert

@student.kuleuven.be

## ABSTRACT

Visualization of sports data is very common in recent years as access to data has increased. The abundance of data and audience allows for new and innovative presentations. In this paper we describe an interactive visualization of the NBA competition over the last 30 years. With three different views, we allow the user to explore data about NBA seasons including standings, team statistics, transfers and player statistics. The d3.js framework was used to develop this visualization.

## ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI):  
Miscellaneous;

## Author Keywords

Information Visualization; CHI; NBA

## INTRODUCTION

The National Basketball Association (NBA) is the most famous professional basketball league in the worldnbawiki. As in every other sports branch, a lot of data and statistics about players, teams, games and seasons are available, mainly in large tables. Despite the large amount of data available, gathering useful insights from this data can be surprisingly challenging. In our work we try to present this data in an innovative and intuitive way, allowing the user to draw meaningful conclusions.

In section 2 we describe the goal of the visualization and the target audience. In section 3 we describe the data used, including its origins, advantages and limitations. In section 4 we give an overview of related literature and web resources, including related visualizations. This includes both visualizations of NBA or other sports data, as well as visualizations tackling an issue encountered during the development of our visualization. In section 5 we describe the visualization itself. We give an overview of the different stages of development of the visualization, as well as the major design decisions. Section ?? discusses potential improvements of the final visualization and lessons learned from the project. We conclude in section ?? . Appendix ?? contains definitions for common basketball terms used in this paper.

## GOAL AND TARGET AUDIENCE

The visualization's goal can be best summarized by the following sentence: "Why did a certain NBA-team win the Championship?" The visualization focuses on exposing relations

between team performance and their player roster over several years. It allows users to find explanations for major improvements and declines in team performance. The visualization allows exploration of NBA data by lay persons. More specifically, the visualization does not focus on premade explanations for phenomena visible in the NBA data. By providing easy and intuitive access to the data, the visualization allows users to draw their own conclusions. The visualization's target audience are lay persons. Specifically, basketball fans are the core of our target audience. It can be of particular interest with the recent surge in fantasy leagues, for which data analysis is paramount to a player's success **fantasy**; **fantasyskill**.

## DATA

The data visualized is a subset of the data available on basketball statistics site basketball-reference**basketball-reference**. A wide range of data is available on this site. This includes common basketball performance statistics, such as field goals, percentage of shots scored, shooting distance, minutes played per season, number of personal fouls made, salaries and much more. All of this data is available as time-series. These statistics are available for each player individually for each season. Aggregated data for entire teams and for entire seasons is also available. Additionally, individual game statistics are also available, including playoff games. In our visualization we only use data from 1984 onwards. This is done because of rule changes in the NBA which changed the number of teams competing and the competition's structure. To simplify implementation of our visualization, only data after the last major rule change was used. The data we use includes league standings and playoff rankings for each team, team overall statistics, the team's roster and individual player statistics for each year, including the PER (Player Efficiency Rating)**per**.

The data was gathered by scraping the basketball-reference site using the provided download capabilities. Most of the data was downloaded in csv format, while some tables had to be manually scraped. This process was automated using Python scripts. The data was then combined in a preprocessing step. In this step, each team's playoff rankings were calculated based on the matches played during the playoffs, and the rest of the data was combined into json format. The final preprocessing step combined all data into one json file. A model of the available data can be seen on figure 1.

Rework  
realted  
work

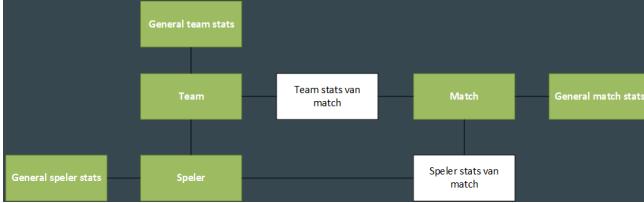


Figure 1. An overview of the data.

## RELATED WORK

On the IEEE Visualization Conference in 2013, the first workshop on sports data visualization was held **ieevis**. There is shown that the area of sport produces large amounts of data, which can be used for building visualizations and making analyses of the data. In the past visualizations of sports existed already, but there were not a lot of visualizations, and they were simple in the sense that they just represented the data without trying to give insights or to approach the data on a different way. There are some, but limited, sports-specific visualizations, like the basketball shot chart. This is a visualization frequently seen when looking for visualizations about the NBA. Some shot chart visualizations try to be different than others, but the general design and idea stays roughly the same.

With our visualization we try to make it easier to interpret a collection of data about the NBA, and try to let users find patterns and insights in the data. A same kind of incentive was the start for **cox2006sportsvis** who created a visualization to help people discover meaning in statistics generated during sporting events in the case of baseball.

(From bubble view) A more traditional visualization of a play-off is the team double elimination bracket. Although this is a more common representation, and there are some nice designs like in this NBA 2012 play-offs visualisation **tournamentladder** where a sunburst visualisation is used, we wanted to create a more compact version by eliminating the recurring representation of a team. We considered enclosed circles, as used in this visualisation of the wimbledon tournament roster **enclosedcircles**. Although the enclosed circles is a compact representation as well, we have continued with our own design as we believe this view is more original and innovative. Furthermore we are convinced that our view gives a good illustration of the end positions. It's generally assumed that the team that is on top of the "ladder" is the winning team. We didn't encounter a similar view when searching for alternatives to visualize a play-off or championship. A new visualization has the advantage that it will trigger users as well.

(From field view) One good alternative was the shooting signature, **peterbeshai**. However, creating such a visualization is again a project on itself and due to time constraints we left it aside. Other possibilities such as a heatmap or spider web didn't fulfill our need to give the visitor a clear overview of how a player performed compared to its peer group.

## VISUALIZATION

The visualization consists of three parts:

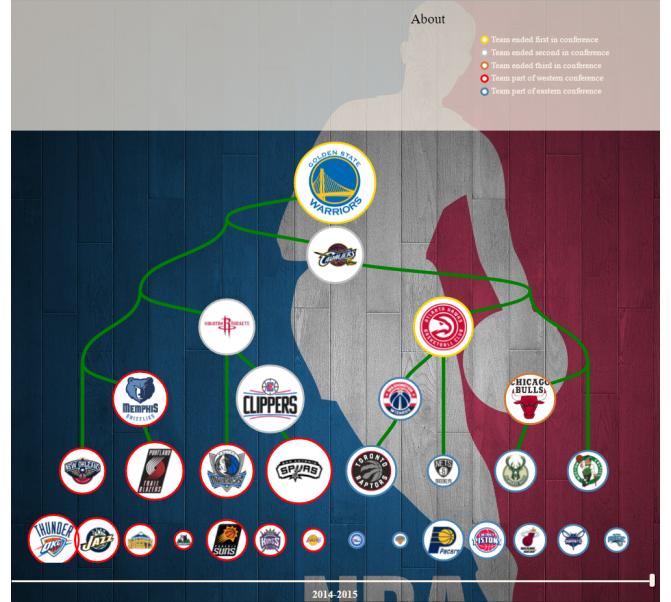


Figure 2. The play-off view without a team selected.

- The play-off view: a compact view on the play-offs per season
- The statistics (zoom) view: a more detailed view on a statistic of a selected team
- The team view: a detailed view on how good or bad a team scores on a specific field position

This technique follows the standards of information visualization. It provides an overview, allows zooming and filtering and provides details on demand **mantra**; **multipleviews**; **automatingdesign**. The play-off view provides the full overview which allows filtering by team. The statistics view provides the ability to zoom into one specific team. There some details are provided instantly, while others can be procured by hovering over certain elements.

The organization of these three views is in a vertical layout with fixed scrolling. Navigation with the keyboard is supported as well.

The choice was made to use a thematic background instead of opting for the traditional minimalistic approach to information visualization. This was done to improve the attractiveness of the visualisation on the one hand and the users' memory of it on the other hand **aesthetics**.

All three views are discussed in more detail below. The three views are available to the user on a single webpage. The user starts on the play-off view and is able to navigate to the statistics and team views by scrolling down or clicking on teams. A different season can be selected with the left and right arrows or by clicking on the timeline. The timeline is a fixed part of the visualisation. It is visible in the three different views.

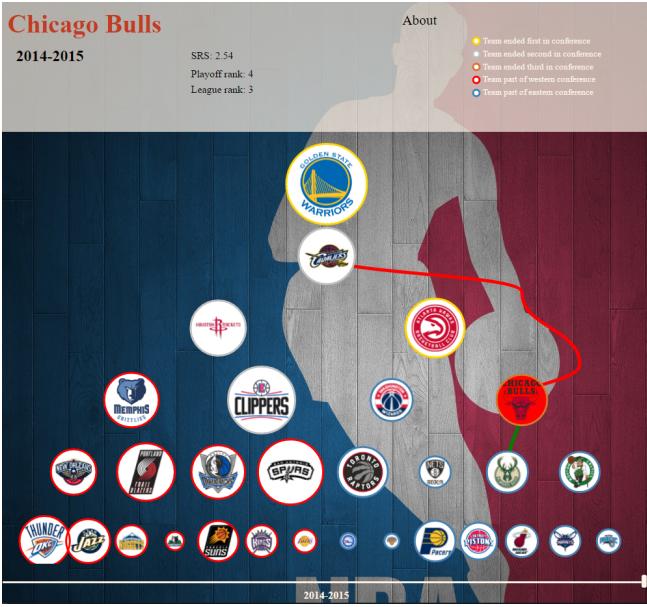


Figure 3. The play-off view with a team selected.

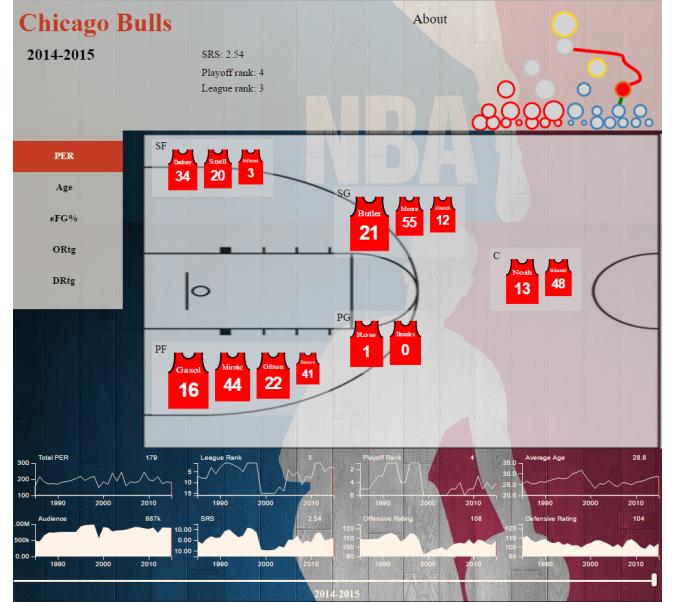


Figure 5. The team view.

### The play-off view

This view informs the user of the roster and end positions in the NBA play-offs for a selected season, in general or for a selected team. The view consists of two parts:

- Context section: the top part informs the user about the selected team and shows a legend of the play-off view.
- Bubble section: this section represents the play-offs of the selected season.

The play-off view is shown in figure ??.

#### Context section

The context section contains a legend for the bubble section. The bubble's borders are gold, silver or bronze if the team ended respectively first, second or third in the regular competition. Teams that didn't end in a podium place during the regular competition have a red or blue stroke color indicating the team's region, respectively the western or eastern conference. The context section also contains a link to an "about" page where the user can get an explanation of the visualization with its different views, can get some background information about the NBA and the terminology used, and can check the source of our used data. If the user hovers over or selects a team in the bubble section, the context section shows more information about that team, as shown in Figure ???. The information shown includes the Simple Rating System score (SRS), the end position in the play-offs and the end position in the regular competition.

#### Bubble section

This section presents the roster and end positions of the play-offs for a selected season. Every team is represented by a circle with the logo of the team inside it. The size of the circle represents the SRS score using an own interpretation of perceptual scaling because the (Flannery) Appearance Compensation is not usable with a domain of both negative and

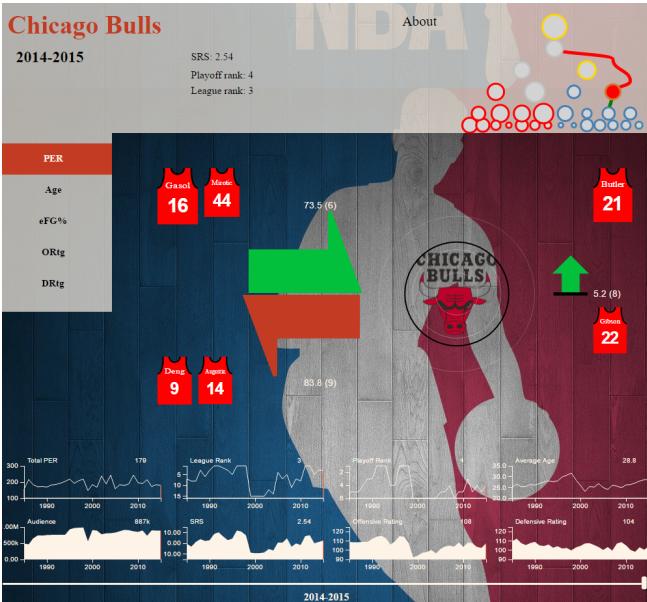
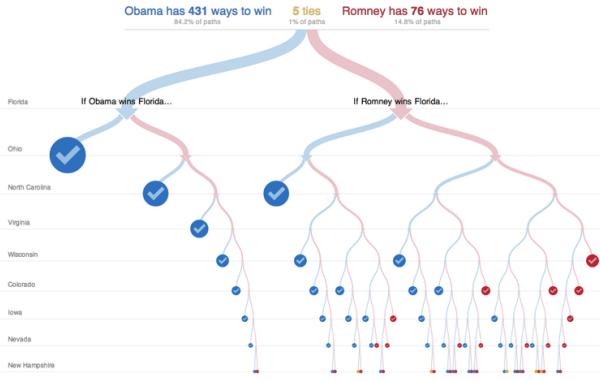


Figure 4. The statistics view.



**Figure 6.** The path of president candidates in their road to become president.

positive values. An exponential function with exponent 1.5 was used. This way the proportion between the area of a bigger and the area of a smaller circle will be greater than the proportion of values the areas represent.

The curved lines between teams represent that those teams have encountered each other in the playoffs. Each row represents an end position in the play-offs. The higher a team is, the further it has come. From the top to the bottom the rows represent the winner of the final (rank 1), loser of the final (rank 2), losers of the semi-finals (rank 3), losers of the quarter-finals (rank 4), losers of the eighth finals (rank 5). The last row represents the teams that did not make it to the play-offs (rank 6).

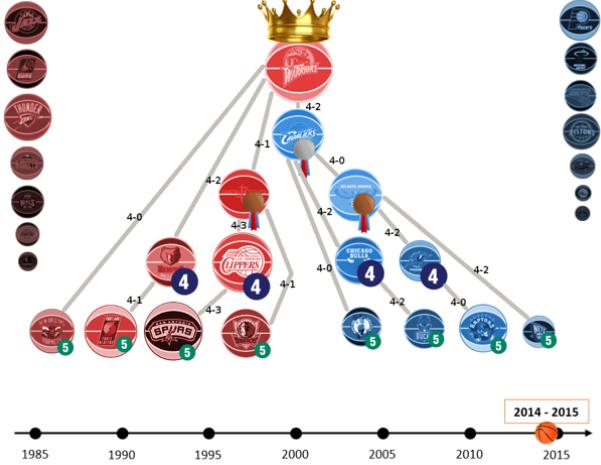
This simple presentation intuitively puts the winning teams higher than the losing teams. The winner of a certain game can be easily deriving from their relative vertical position. It is also clear that each team needs to have played and won a series of matches against one other team at every level below them ???. This view is further reinforced by highlighting a team and its matches and opponents when hovering over it. Additional team information is then also shown in the context part.

#### Previous designs

Our initial design used the team circles' lightness to indicate their final ranking in their region. The hue was used to show the league they were part of. According to the Ranking of Perceptual Tasks **perceptualranking** this is a good property for ordered values like the regular competition ranking. But the combination of the logo's of the teams with the colors resulted in an overload of information and made it more difficult to compare the end ranking of different teams.

As an alternative we tried the same view without the logo's. This was deemed inconvenient because you cannot instantly see which team is represented by which circle. Therefore we've chosen to only use the teams' logos.

Our initial play-off view was too dense because of the connections between teams as well. It connected teams that played against each other directly. When searching for alternatives to



**Figure 7.** First design of the play-off view.



**Figure 8.** First implementation of the play-off view.

Hierarchy  
 = sorting  
 = facilitate  
 trends of  
 winners position  
 have a meaning  
 in the play-off view  
 when selecting circle  
 =>  
 zoom/more detail  
 Bubble principle of  
 - Symmetry (whole = play-off) - connectedness arc = games played Pre attentive characteristics - size of circles - hue to highlight selections or indicate increase decline arrows next view

reflect on our choice we discovered a similar visualization in a totally different context **whitehousepath**. This visualization gives a cleaner view of the competitors for a specific team compared to our original sketch. Hence we decided to change our visualization. We added an intermediate step between teams to more clearly indicate how teams competed to become NBA champion.

add statistics  
 view image here

We initially also considered a map view to show the teams and their regular competition ranking on a map of the USA. However, this view had no additional value compared to the play-off view except for the geographical position, which has no importance for our visualisation. Furthermore we've found a visualisation of the NBA with a map view **mapviewvisualization** similar to our idea, so we decided to not implement this map view.

### The statistics view

The statistics view gives a user a clear overview of how a statistic has been influenced in the selected season compared to the season before. The statistics view is showed in Figure 4. The view consists of 3 coordinated parts:

- Context section: the top part informs the user of the context of the selected team
- Arrow section: this section illustrates why a statistic changed compared to the previous season
- Selected statistics section: team statistics and how they evolve over time

#### Context section

The context section is partly identical to that of the play-offs view. The team's information is also displayed in the top left. In the top right, the bubble section of the play-offs view is displayed in a reduced, minimalistic form. This view is repeated to create a sense of continuity between the different views and provide context of the current position in the visualization. The team shown in the statistics view is highlighted on this smaller bubble section as well. Replicating this section also alleviates short-term memory issues. It prevents the need to scroll back and forth between the different views constantly.

#### Arrow section of selected statistic

This section illustrates a change for a chosen statistic for the selected team compared to the previous season. This gives a user more insight into why a statistic changed over time and how this change has impacted team performance. The selected team is represented by a circle in the same way as explained in the play-off view. The change for a statistic is represented by three arrows:

- Arrow on the left pointing towards the circle: indicating the influence of players who joined the team
- Arrow on the right of the circle: indicating how the team internally changed, e.g. by training its players.
- Arrow on the left pointing away from the circle: indicating the influence of players who left the team

The jerseys next to the arrows are the jerseys of players that had the biggest contribution to that arrow. On the left, different statistics , including player PER can be chosen from. Below this, player information is shown when hovering over a player's jersey.

#### Selected statistics

Below the arrow section some selected statistics are shown over time. The line charts enables a user to identify peaks or drops that influenced the outcome of a team in this season, following seasons or previous seasons. A user interacts with the visualization by shifting the bar indicating the current year displayed. By doing this the view gets updated and the user can see the impact on the team in the context section and why the statistic changes in the arrow section. This view provides a lot of additional information to users by allowing them to see correlations between statistics and phenomena quickly and easily.

#### The team view

This view shows each player of the team on its field positions. Players are represented by their jerseys. Jerseys aligned next to each other share the same position, which is indicated as well. The size of a jersey would illustrate a players performance on this spot. This view as well as the statistics view should give the user the ability to search for explanations why a team is performing better or worse over seasons. On the other hand, a user can see what impact a change in team characteristics has on its overall performance. This view should give a user insight on how strong or weak a particular team is for a specific field position. A user can then scroll through time to see how field positions evolve and what the influence is over time.

#### Previous designs

Initially we designed this view more elaborate. When the user would click on a jersey, specific player information would be shown. A number of player statistics would be shown as small multiples. A bell curve should have informed how a player scores compared to his team, other players in the league on this position or all other players in the league. We considered other charts for the small multiples as well. A box plot or bullet was less clear for us to indicate where a player stands compared to the group he was compared with. Alternatives for the small multiples were evaluated as well. This part of the visualization was deemed less important than the others and has not been fully developed because of time and resource constraints.

#### Improving the result

Talk about how we improved the data-ink ration in the end..

#### FUTURE WORK

As with most projects, the result is never 'finished', meaning that there are always improvements possible but due to timing or resource constraints they haven't all been implemented. We suggest a number of features that might improve the visualization:

- Smooth transitions between states could improve the user experience. For example when changing year, a transformation of the arrow in the statistics view would help the user

sketch of small multiples here

Is this necessary? I don't know what to write here

see a difference between years. At the moment the arrows change at the blink of an eye, making it difficult for us humans with a short term memory to make the comparison. This would have been especially useful in the transitions between different years on the bubble view.

- To further support our short term memory we could also 'save' the previous state. Eg. in the field view we could visualize ghost jerseys to indicate which players left the team or position and highlight players that joined the team. This would help the user see how the team changes over time.
- At the moment no attention has been payed to the process, provenance and history of the user's exploration session. A user thus needs to memorize his action to be able to share or reproduce his actions. This could be improved by adding URL parameters. This way a user at least can share or save a state of an analysis. One step further is to provide a play-feature. The visitor could them share his exploration and replay it. A simple implementation would ask the visitor for a start date and an end date and would then 'play' the changes of a particular team and characteristic over time.
- We could support users further in looking for patterns by focussing more on comparisons. We could integrate a view where a visitor is able to compare two teams in the same year, or in two different years.

## Technology

To create the visualization, the d3 javascript framework **d3.js** was used in combination with html5 **html5** and jquery **jquery**. Note that D3 is based on HTML, CSS, Javascript and SVG. This allows easy access to the visualization as most modern browsers are capable of handling these webstandards. The choice not to use the d3 framework for the entire visualization was made to ease the layout configuration of the visualization. Instead, html was used to do the global layout of the visualization. We opted to structure the site with multiple divs as containers. To each container we allocate a svg with a specific visualization.

When working with d3.js we encountered some obstacles:

- A synchronous call need to be combined with synchronous calls. More specific, when a visitor uses the timeline to scroll through time, the whipping of previous visualizations was not synced with the creation of the visualisation. This resulted in whipping sections when a visualization was not created yet. Hence the view resulted in multiple figures overlapping.

## LESSONS LEARNED

We've learned a couple of lessons during the creation of this visualization:

- Find a story to tell with your visualization
- Each separate component should be a sufficient informative visualization on itself
- Creating a custom visualization costs time and opportunities

In what follows we will discuss our experience in more detail.

sketch  
of how  
the  
page is  
structured  
here

## Story telling

At the beginning of this project we were able to design the 3 separate views quite quickly. Our main problem was to integrate these views as a whole. When we created drafts of two views being the play-off view and the statistics view, we suddenly noted patterns. The Golden State Warriors became NBA champion a couple of years after a great uplift in SRS score. This was because of new players joining the team in 2009. In the years following, the team kept on attracting talented players and increased the inherent team score to become NBA champion in 2015. With the team view, one can even notice that the team gets stronger in center team positions during this period. Users are able to explore the data for patterns. On top of that, we provide the insight on how the changes were impacted by changes in teamplayers.

## Each component should be informative

In order to have a good visualization as a whole, each component should be a clear and informative visualization on itself. During the project, we started noticing that most components in a visualization lack references and hence are not informative enough. Eg. each small multiple of a statistic in the statistics view on itself gives sufficient information to stand on itselvs. The bar in the small multiple indicates the score in the selected season and enables the user to evaluate that score over time. Have there been better scores, or worse scores. Is this score part of an upward movement during years? This should be weighed against the duplication of data and use of screen space. Coordinates views alleviate this need significantly as well, as each component is strongly supported by other components.

## Custom visualizations

Before exploring what is out there in the d3.js world, we made our own sketches. This is how we came up with the 3 custom views that make up our visualization. When comparing our design with alternatives we could not find something that satisfied our needs. Nonetheless, they gave us inspiration to finetune our designs. Although we are satisfied with our result and we are convinced it was the right choice to reach our goal, it had some draw backs. We noted that other teams had more freedom in exploring multiple different visualizations which enabled them to evaluate different options with trial and error. Much of the time developing was spent on the technical details for these custom visualizations. This was especially true for the bubble component of the playoffs view. No out of the box solution for this could be found so everything from the determination of the positions of the bubbles to how to connect them had to be developed by us.

## CONCLUSION

We believe the visualization we have created can be of tremendous value for fans and fantasy basketball players. Personally we have gained several insights into the performance of basketball teams by using the visualization and we believe others could learn even more. The most interesting conclusion we were able to make concerns the importance of individual players in a basketball team. We were able to find strong correlations between having two to four very strong players and

the team's performance. It seems having one star player is not enough. On the other hand, having more than 5 good players no longer has much influence either.

While the visualization has a lot of potential for improvement, we think the innovative play-offs representation combined with detailed team information in an integrated fashion has much value.

## References

- aesthetics** Andrew Vande Moere and Helen Purchase. "On the Role of Design in Information Visualization". In: *Information Visualization* 10.4 (Oct. 2011), pp. 356–371. ISSN: 1473-8716. DOI: [10.1177/1473871611415996](https://doi.org/10.1177/1473871611415996). URL: <http://dx.doi.org/10.1177/1473871611415996>.
- automatingdesign** Jock Mackinlay. "Automating the Design of Graphical Presentations of Relational Information". In: *ACM Trans. Graph.* 5.2 (Apr. 1986), pp. 110–141. ISSN: 0730-0301. DOI: [10.1145/22949.22950](https://doi.org/10.1145/22949.22950). URL: <http://doi.acm.org/10.1145/22949.22950>.
- basketball-reference** Basketball Reference. *Basketball Reference*. Apr. 2016. URL: <http://www.basketball-reference.com/>.
- cox2006sportsvis** Andy Cox and John Stasko. "Sportsvis: Discovering meaning in sports statistics through information visualization". In: *Compendium of Symposium on Information Visualization*. Citeseer. 2006, pp. 114–115.
- enclosedcircles** Peter Cook. *Wimbledon 2013 Player bubbles*. May 2016. URL: <http://wimbledon.prcweb.co.uk/playerbubbles.html>.
- fantasy** Gabriel S. Dzodom and Frank M. Shipman. "Data-driven Web Entertainment: The Data Collection and Analysis Practices of Fantasy Sports Players". In: *Proceedings of the 2014 ACM Conference on Web Science*. WebSci '14. Bloomington, Indiana, USA: ACM, 2014, pp. 293–294. ISBN: 978-1-4503-2622-3. DOI: [10.1145/2615569.2615649](https://doi.acm.org/10.1145/2615569.2615649). URL: <http://doi.acm.org/10.1145/2615569.2615649>.
- fantasyskill** Ed Miller and Daniel Singer. *For daily fantasy-sports operators, the curse of too much skill*. Sept. 2015. URL: <http://www.mckinsey.com/industries/media-and-entertainment/our-insights/for-daily-fantasy-sports-operators-the-curse-of-too-much-skill>.
- html5** Wikipedia Community. *Wikipedia - HTML5*. Apr. 2016. URL: <https://en.wikipedia.org/wiki/HTML5>.
- ieeveis** IEEE. *What's the score? The 1st Workshop on Sports Data Visualization*. May 2016. URL: <http://workshop.sportvis.com/>.
- jquery** JQuery. *JQuery*. Apr. 2016. URL: <https://jquery.com/>.
- mantra** Ben Shneiderman. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations". In: *Proceedings of the 1996 IEEE Symposium on Visual Languages*. VL '96. Washington, DC, USA: IEEE Computer Society, 1996, pp. 336–. ISBN: 0-8186-7508-X. URL: <http://dl.acm.org/citation.cfm?id=832277.834354>.
- mapviewvisualization** Lemon Chiu. *NBA-Visualization*. May 2016. URL: <http://lemonchiu.github.io/NBA-Visualization/>.
- multipleviews** Michelle Q. Wang Baldonado, Allison Woodruff, and Allan Kuchinsky. "Guidelines for Using Multiple Views in Information Visualization". In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. AVI '00. Palermo, Italy: ACM, 2000, pp. 110–119. ISBN: 1-58113-252-2. DOI: [10.1145/345513.345271](https://doi.org/10.1145/345513.345271). URL: <http://doi.acm.org/10.1145/345513.345271>.
- nbawiki** Wikipedia Community. *Wikipedia - National Basketball Association*. Apr. 2016. URL: [https://en.wikipedia.org/wiki/National\\_Basketball\\_Association](https://en.wikipedia.org/wiki/National_Basketball_Association).
- per** Wikipedia Community. *Wikipedia - Player Efficiency Rating*. Apr. 2016. URL: [https://en.wikipedia.org/wiki/Player\\_efficiency\\_rating](https://en.wikipedia.org/wiki/Player_efficiency_rating).
- perceptualranking** Jock Mackinlay. "Automating the design of graphical presentations of relational information". In: *Acm Transactions On Graphics (Tog)* 5.2 (1986), pp. 110–141.
- peterbeshai** Peter Beshai. *Buckets: NBA Shot Visualization*. Apr. 2016. URL: <http://peterbeshai.com/buckets/app/>.
- playoffs** Wikipedia Community. *Wikipedia - NBA playoffs*. May 2016. URL: [https://en.wikipedia.org/wiki/NBA\\_playoffs](https://en.wikipedia.org/wiki/NBA_playoffs).
- srs** Doug Drinen. *A very simple ranking system*. May 8, 2006. URL: <http://www.pro-football-reference.com/blog/?p=37>.
- tournamentladder** @playoffsbracket. *NBA Playoffs Bracket*. May 2016. URL: <http://www.nbaplayoffsbracket.com/2016/index.php>.
- whitehousepath** Mike Bostock and Shan Carter. *512 Paths to the White House*. Nov. 2, 2012. URL: <http://www.nytimes.com/interactive/2012/11/02/us/politics/paths-to-the-white-house.html>.

## APPENDIX

### TERMINOLOGY

#### NBA

National Basketball Association - The largest, most well-known basketball of the USA.

#### PER

Player Efficiency Rating - An all-in-one basketball rating, boiling down all of a player's contributions into one number **per**.

#### SRS

Simple Rating System - a rating that takes into account average point differential and strength of schedule. The rating is denominated in points above/below average, where zero is **averagesrs**.

#### Playoffs

The National Basketball Association (NBA) playoffs are a best-of-seven elimination tournament among 16 teams in the Eastern Conference and Western Conference (called divisions, pre-1970), ultimately deciding the winner of the NBA **Finalsplayoffs**.

#### Jersey

The shirt worn by basketball players.