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 Langeraert
KU Leuven
yves.langeraert
@student.kuleuven.be

ABSTRACT
This paper details the conceptualization and implementation of an interactive visualization of data about NBA seasons including teams, game results 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; See http://acm.org/about/class/1998/ for the full list of ACM classifiers. This section is required.

Author Keywords
Information Visualization; CHI; NBA

INTRODUCTION
This paper describes the visualization of NBA data made by the "The Tufters" team for the course "Information Visualization". In section **??** we describe the goal of the visualization and the target audience. In section **??** we describe the data used, including its origins, advantages and limitations. In section **??** 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 isue encountered during the development of our visualization. In section **??** 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 they win?" 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 try to offer premade explanations for phenomenons 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, fans of NBA are the core of our target audience. This means the visualization assumes most common basketball terms are known to the audience and does not provide basic terminology explanations.

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. These statistics are avalable 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.

RELATED WORK

VISUALIZATION
The visualization consists of three parts:

- The bubble 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

All three are discussed in more detail below. The three views are available to the user on a single webpage. The user start on the bubble view and is able to navigate to the statistics and team views by scrolling down or clicking on teams.

The bubble view

This view supposes to inform the user of the results of the NBA play-offs per season. A circle is used to represent a team, the size its SRS score using a '?' distribution, a stroke to illustrate it's region or medal won during regional competition and curved lines to inform the user which teams played against each other. The user has a timeline at its disposal to scroll through time. Note that the team can be recognized by the logo inside the circle. Our initial desing filled the circle with the red or blue depending on the region where the teams compete. However, the shade of colour was darker if their final ranking was higher in their region. We've replaced this with the logo's of the teams as we thought this is more user friendly and the extra information of the ranking doesn't outweight. <picture of the bubble view compared with double elimination bracket> A more popular visualization of a play-off is the team double elimination bracket. Although this is a more common representation, we wanted to create a more compact version by eliminating the recurring representation of a team. We also considered enclosed circles and a sunburst. <figure of 2 alternatives https://thetuftersblog.wordpress.com/2016/03/07/team-ranking-play-off-infovis/> 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. We didn't encouter a similar view when searching for alternatives to visualize a play-off or champi-onship. Our initial bubble view was too compact. It connected teams that played against each other directly. When searching for alternatives to reflect on our choice we discovered a similar visualization in a totally different context. <https://source.opennews.org/en-US/articles/nyts-512-paths-white-house/> <comparison of original bubble view with white house paths> This visualization gives a cleaner view of the competitors for a specific team compared to our original sketch. Hence we decided to adapt our visualization. We added an intermediate step between two teams to clearer indicate how teams competed to become NBA champion. Next to the static view, the user can interact with the visualization. When he hoovers over a circle of a team, only the games played by that team are shown by highliting the competitors (their circles) and the curved lines that connect them. This view gets fixated when the users clicks on a team. He/she then automatically gets forwarded to the statistics view where the bubble view is still represented in the right upper corner to make sure the user keeps an overview. The statistics view will be explained in more detail in the next section.

The statistics view

The statistics view gives a user a clear overview of how a statistic has been influenced compared to the year before. The view consists of 3 parts:

- Context section: the top part should inform the user of his/her context

- Arrow section: hhis section illustrates why a statistic changed

- Small multiples section: hmall multiples of team statistics and how they evolve over time

Context section

The context section should inform the user at all times of the current selections that have been made. The statistics view will always be based on a specific team and season that have been selected or that are initated. This should be clear based on the textual information in the top left. Next to that, in the top right a small bubble view is shown. It indicates the team selected and how well it performed in the play-offs during that season, as ecplained in the bubble view section. When changing teams and/or season, this view gets updated as well as the textual information.

Arrow section of selected statistic

This section illustrates a change for a chosen statistic for the selected team compared with the season before. This should give a visitor more insight into why a statistic changed over time. The selected team is represented by a circle in the same way as explained in the bubble 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 (team member current season, not previous season)

- Arrow on the right of the circle: indicating how the team inherent changed (team member this season and previous season)

- Arrow on the left pointing away from the circle: indicating the influence of players who left the team (team member previous season, not current season)

<figure of the statistics view> The shirts next to the arrows are the shirts of players that had the biggest contribution to that arrow. Note that these shirts don't change in size. The box on the right enables the user to change between statistics. Next to that it informs the user of which statistic has been chosen. When the user hits the circle in the middle representing the teams, he is guided to the team view. This view will be explained in more detail in the next section.

Small multiples of different statistics

Below the arrow section small multiples are shown of multiple statistics over time. The line charts enables a user to identify peaks or drops that possibly influenced the outcome of a team is 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. An alternative to scroll through time is to use the timeline at the bottom of the page.

The team view

This view as well as the statistics view should give the user the ability to search for explanations why a team is 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 shows each player of the team on its field positions. Shirts aligned next to each other will thus share the same position. The size of a shirt would illustrate a players performance on this spot. The PER score is initiated. 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, see how field positions evolve and what the influence is over time. Initially we designed this view more elaborate. When the user would click on a shirt, specific player information would be shown. Next to some general text information, we would have visualised a number of player statistics 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. 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 peergroup. <Sketch of the small multiples>

Technology

To create the visualization, the d3 javascript framework **d3.js** was used in combination with html5 **html5** and jquery **jquery**. This allows easy access to the visualization as most modern browsers are capable of handling these technologies. 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. Eg. The statistic view is build out of x divs filled with different svg's created with the d3 framework. <sketch of how the page is structured> When working with d3.js we encountered some obstacles: - A synchrounous call need to be combined with synchronous calls. More specific, when a visitor uses the timeline to scroll through time, the whiping of previous visualizations was not synced with the creation of the visualisation. This resulted in whiping 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 use case as soon as possible to build your visualization around
- 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.

A good use case

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 bubble view and the statitics view, we suddenly noted patterns. The Golden State Warriors became NBA champion a couple of years after a great uplift in SRS score. This due to a increase inside the team and 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 (with ups and downs) 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. The visualization however is missing at the moment if this reinforcement of the field positions is due to the attraction of talented people or by utilizing available talent within the team.

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. This is a lesson learned from our work, but also when evaluating the work of peers. We started noticing that most components in a visualization lack references and hence are not informative enough. Eg. each small multiple of a statisctic in the statistics view on itself gives sufficient information to stand on itselfs. 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? For this reason we can argue that the circle representing the teams SRS score in the statisctics view is not the best choice. We could have chosen a bar chart with an indication of the best and worst score for that specific team over time or for the best and worst team in that season.

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 makes 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. We could have prepared our data in a format that was more compatible with out of the box examples, but this would have had too much of an impact on our result. Next to that, although we've tried to write our code as general as possible,

we're not convinced this could be shared in such a way that other people building a visualization could easily re-use our work to visualize other sports.

CONCLUSION

References

basketball-reference Basketball Refernce. *Basketball Reference*. Apr. 2016. URL: http://www.basketball-reference.com/.

d3 Mike Bostock. *D3 - Data Driven Documents*. Apr. 2016. URL: https://d3js.org/.

html5 Wikipedia Community. *Wikipedia - HTML5*. Apr. 2016. URL: https://en.wikipedia.org/wiki/HTML5/.

jquery JQuery. *JQuery*. Apr. 2016. URL: https://jquery.com/.

per Wikipedia Community. *Wikipedia - Player Efficiency Rating*. Apr. 2016. URL: https://en.wikipedia.org/wiki/Player_efficiency_rating/.

peterbeshai Peter Beshai. *Buckets: NBA Shot Visualization*. Apr. 2016. URL: http://peterbeshai.com/buckets/app/.

APPENDIX

TERMINOLOGY

PER

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