NBAVis: Visualizing National Basketball Association Information

Bruno Pagno, Leandro Guedes, Carla Freitas and Luciana Nedel
Instituto de Informtica, Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre, Brazil
Emails: {blpagno,lsguedes,carla,nedel}@inf.ufrgs.br

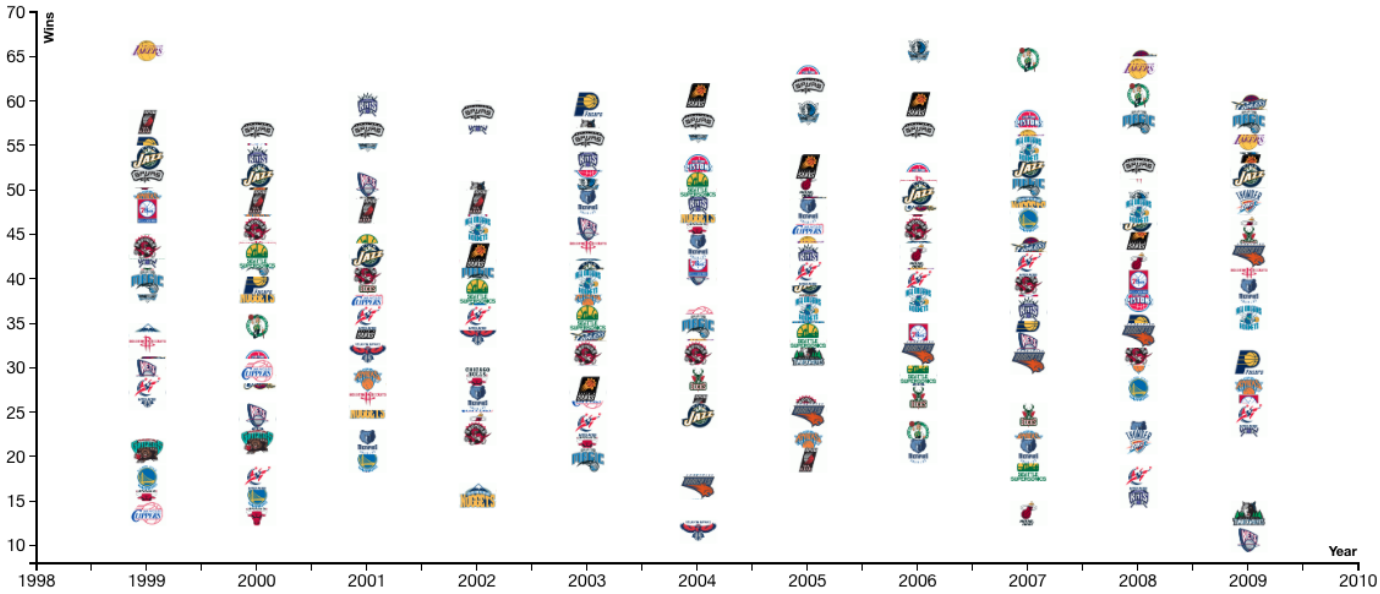

Fig. 1. Number of wins from 1999 to 2009 of NBA League teams, each team represented by its logo.

***Abstract*—In the National Basketball Association there are thousands of players, plays and several teams. This produces large amounts of data. Analyzing past data it is possible to extract important information on the sport and infer better strategies for the future. This paper introduces an ongoing work, which aims at providing data visualizations to make it easier and faster to extract useful information from a NBA dataset. By adapting known data visualization methods to this context, one can visualize large amounts of data simultaneously and rapidly extract the relevant information. Moreover, the project was developed as a website and it is possible to add new information to the dataset with no need of regenerating or recompiling the visualization code.**

***Keywords*-Information Visualization; Basketball**

I. Introduction

Sports usually produce large amounts of data about players, teams and plays. In the NBA (National Basketball Association) there has been thousands of players since its beginning, and each of them has participated at least in a few games, adding a reasonably large amount of data. These historical data are frequently used by teams to improve their game. Analyzing past data it is possible to extract important information about the sport, and infer better strategies for the future. However, statistics and raw data analysis can be a tiring process, only possible for statistics professionals.

This paper proposes data visualizations to make it easier and faster to extract useful information from a NBA dataset. By adapting known visualization methods to this context, we can visualize large amounts of data simultaneously and rapidly extract the relevant information. Also, because the project was developed as a website, we are able to add new information to the dataset without having to regenerate or recompile the visualizations.

A. Related work

In 2013, the first workshop on sports data visualization took place during the IEEE VIS in Atlanta, Georgia. This shows that sports visualizations are currently in focus of visualization researchers, beyond what we have been seeing in the communication media.

There has been a variety of works in sports information visualization. Page and Moere [1] approached the subject showing team visualizations focused on different class of users (athletes, spectators and judges). Even though they do not discuss much about team performance data, they show how

teams have much information that can be used by different users. Cava and Freitas [2] present a visualization for soccer matches with custom glyphs in an adjacency matrix visual representation. Their work show a good use of glyphs in visualization. Season data and its importance for the sport are discussed by Cox and Stasko [3] while Beck, Burch and Weiskopf [4] present time varying visualizations but in a far more complex context then the work proposed in this project.

Also, visualization of sports data is often seen in games. Specifically in Football Manager [5] and Pro Evolution Soccer [6] it is possible to find star plots showing player stats.

B. Technique overview

For the visualizations reported herein, we use an open and free distributed dataset from DatabaseBasketball [7] containing information from players and teams over several NBA seasons. This dataset, originally a csv file, was converted into a database to make it better structured and easier to manipulate. Then, we developed a scatterplot visualization for showing NBA teams compared by their number of wins (see Fig. 2). This plot allows the visualization of the overall success rate of a team and its stability along the years. For the situations where several teams have a similar performance and the visualization does not clearly show their differences, there is an option where the user can highlight a single team just passing the mouse on it. Fig. 2 shows San Antonio Spurs as an example.

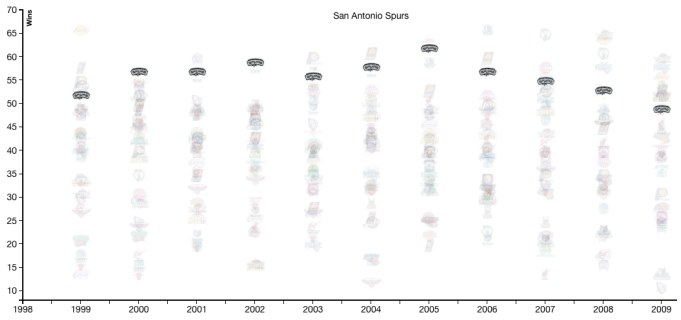

Fig. 2. Highlighting San Antonio Spurs overall performance from 1999 to 2009.

Sometimes there is too much fragmented data for a straightforward comparison. Because of this, we created the relevance coeficient that shows the general performance of the player. It is calculated by summing his normalized attributes like number of assists and points, and subtracting his normalized negative attributes when it means a unwanted action like number of turnovers or personal fouls. This is shown in the individual player scatterplot visualization (Fig. 3). This helps the visualization to display the consistency of the player throughout the years he has played, and also serves as a mean to compare players with larger or lower impact.

So, for individual players, we provide two visualizations: (1) a scatterplot with their relevance coefficient spread through the years they have played, an example being shown in Fig. 3 using data about Ray Allen, and (2) a starplot with their most relevant characteristics, which is exemplifyed in Fig. 4.

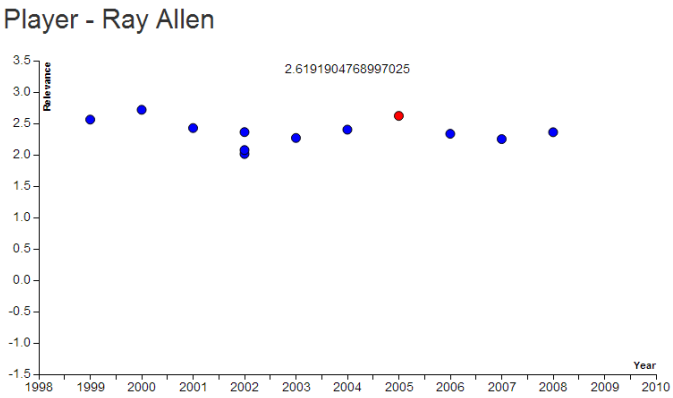

Fig. 3. Ray Allen relevance information timeline.

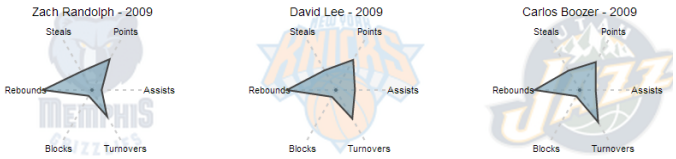

Fig. 4. Starplot example: Players on 2009.

II. Technical Details

In this section, we detail how data is structured, and how the visualizations and the main structure were built.

A. Data structure description

The dataset used in this work comprises regular season information about teams and players in the NBA since 1945. Both teams and players records contain information about points, rebounds, assists, steals and other relevant data for knowing their performance history.

B. Visualizations

We used two known visualizations in different situations for displaying information currently in our NBA dataset. They were chosen because, according to Chambers et al. [8], people are very used to them in data analysis.

1) Scatterplot: Scatterplot is a very used graphical method [8, p. 45], making easier to the user perceiving and understanding what the visualization is about. As shown in Fig. 2, we used two axis, one for the year and other for the number of wins of each team, while in Fig. 3 we plot the player performance factor for each year. The use of the NBA team logo makes the visualization much clear and easy to understand, eliminating the need of a complex caption.

2) Starplot: Star plots are a popular representation of multidimensional data, commonly used in games for showing characters attributes. We considered it a good way to depict each player information because the form of the plot already shows his qualities. Chambers et al. [8, p. 61] indicate that star diagrams are easier to judge than bar charts. As shown in Fig. 4 and Fig. 5, we used a six-axes star plot showing the most relevant information about each player.

C. Web presentation

The visualizations are available in a website. This makes it possible to create a navigation flow that is easy to understand by the end user. We separated the visualizations in teams and players. This way it is possible to display teams and their players' information. Also, there is a page showing a players ranking, where they can be ordered according to their attributes. Fig. 5 is actually part of the 2009 assists ranking, where one can see similar glyphs, since the three starplots shown are from players that are near in the ranking.

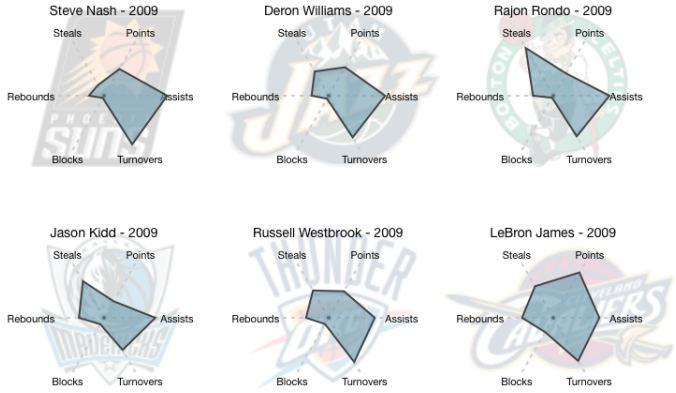

Fig. 5. Starplots for part of the assists ranking of some NBA players in 2009.

III. Implementation

At first we set up an website using Ruby on Rails [9]. Rails is a framework that allows the developer to create dynamic web applications with database support. Using the framework we parsed the NBA dataset, and used it to build the SQL database, with explicit relations. This way we can easily retrieve the players from a specific team or other similar information. The framework and database also allowed us to create any page using only a subset of the original data in each visualization.

To create the visualizations we used the D3 [10] Javascript library. D3 is a tool that allows the developer to create several visualizations in a easy and straightforward manner. We also used the D3 Star Plot [11] extension for the star plot.

The last step was uploading the application to the Heroku [12] host, and can be accessed on this link [13]. Heroku offers free hosting with limitations and integrates greatly with Ruby on Rails applications. However, this server limits the amount of database entries so we had to remove older entries and cannot expand the amount of information in the current setup. We are studying where to host the application for supporting a larger dataset, which would surely allows us to create more consistent and better visualizations.

IV. Discussion

It is well known that sports accumulate large amounts of data [1] [2]. This alone is a reason enough to build visualizations from such data, because they tend to be complex. The visualizations provided so far in this project make it easy for the user to view which teams and players have better or worse performances along all seasons or in specific years.

An example of information that can be inferred by the player starplot visualizations is that different players that play on the same positions tend to have similar glyphs as can be seen in Fig. 4. Zach Randolph, David Lee and Carlos Boozer are all Power Forwards. All of them show a large number of points, turnovers and rebounds compared to steals, assists and blocks. It becomes clear what the player role is, and how good that player tends to be, but it also allows the user to visualize which players tend to have different style of play.

V. Conclusion

In this paper, we showed the use of known visualization techniques applied to a sports dataset. We also show that it is possible to extract information and to infer conclusions simply from observing the visualizations.

For future work on the software we intend to add a single visualization, specific for direct comparison of teams and players. This would allow better insights without the need of navigating through different visualizations. This could be easily accomplished with two or more star plots drawn with different colors overlapping each other for making instantly noticeable the difference of players or teams. Another possibility of future work is to explore other multidimensional visualization techniques, for example those based on dimension reduction, to allow observing possible clusterings of teams and players.

Acknowledgment

We are grateful to CAPES, CNPq and FAPERGS for funding our current studies, through student scholarships and research grants.

References

[1] M. Page and A. Moere, "Towards classifying visualization in team sports," in *Computer Graphics, Imaging and Visualisation, 2006 International Conference on*, July 2006, pp. 24–29.
[2] R. Cava and C. D. S. Freitas, "Glyphs in matrix representation of graphs for displaying soccer games results," in *SportVIS - Workshop on Sports Data Visualization*. Atlanta, Georgia, USA: IEEE VIS 2013, October 2013.
[3] A. Cox and J. Stasko, "Sportvis: Discovering meaning in sports statistics through information visualization," 2006.
[4] M. Beck, F. Burch and D. Weiskopf, "Visual comparison of time-varying athletes performance," in *SportVIS - Workshop on Sports Data Visualization*. Atlanta, Georgia, USA: IEEE VIS 2013, October 2013.
[5] Sports Interactive, "Football manager," http://www.footballmanager.com/lang/en_GB, 2004–2014.
[6] Konami, "Pro evolution soccer," 2004–2014.
[7] Databasebasketball. [Online]. Available: http://www.databasebasketball.com/
[8] J. Chambers, W. Cleveland, B. Kleiner, and P. Tukey, *Graphical Methods for Data Analysis*, 1983.
[9] Ruby on rails. [Online]. Available: http://rubyonrails.org/
[10] Data driven documents. [Online]. Available: http://d3js.org/
[11] D3 star plot. [Online]. Available: https://github.com/kevinschaul/d3-star-plot
[12] Heroku. [Online]. Available: http://heroku.com/
[13] Nbavis. [Online]. Available: http://nbavis.herokuapp.com