Football Analytics with D3.js

Hugo Duportet MSc Data-Science
Data Visualization
University Claude Bernard Lyon 1
Villeurbanne, France
Email:hugo.duportet@univ-lyon1.fr

Adam Benharrats
MSc Data-Science Data Visualization
University Claude Bernard Lyon 1
Villeurbanne, France
Email:adam.benharrats@univ-lyon1.fr

Index Terms—Odds visualization, Sunburst, barchart, surebet, multiple barchart, friendly-user

ABSTRACT

Betting on sports, especially on football, has become commonplace today and many visualizations highlighting games statistics and championship history already exist. However, the very study of odds from different bookmakers and their evolution over several gamees is not a domain of extensive research in data visualization. We propose to unify zoomable sunburst like visualization and multiple bar chart visualizations to analyze for one game, the odds from different bookmakers and their best combination, and the mean of the past odds for the two teams considered.

AUTHOR KEYWORDS

Visualization; Interactive Zoomable Sunburst; Bar Chart; Odds Analysis.

GENERAL TERMS

Design; Innovative Odds Visualization.

1 INTRODUCTION

Football odds are predominant within all sports betting odds and bar charts are one of the simplest ways to represent them. However, using only bar charts to visualize various games odds interactively can be challenging and we introduce an interactive technique to analyze several games odds (and the previous ones) throughout a season. The odds we work with are the Home (home team win the game), Away (away team win the game) and Draw (no winner) odds. Our method merges two well known visualizations, interactive zoomable sunburst and multiple bar chart and we designed it in order to explore efficiently the different odds. It is really user-friendly in that it aims to give a clear summary of the odds for two teams meeting and nothing more and it can serve as well for the data scientist building a predition model as for the proven bettor. We did not add any other statistics to stay focus on the odds though we will add the possibility to change the type of odds (numbers of goals, result at half-time,...) thereafter. The data we used first were all the games data from the season 2015-2016 of the french first division championship "la Ligue 1".

This paper first reviews related work on zoomable sunburst and mutliple bar charts and then describes the technique that we rely on in detail.

2 RELATED WORK

2.1 Sunburst

Set up of a space with a visualization that use a radial layout which is both dynamic and simple to understand.

This type of visualisation shows hierarchy through a series of rings, that are sliced for each category node. Each ring corresponds to a level in the hierarchy, with the central

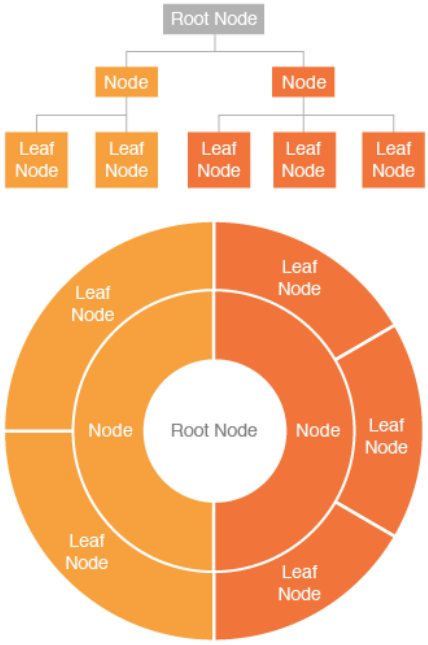

Fig. 1. Sunburst anatomy

circle representing the root node and the hierarchy moving outwards from it. Rings are sliced up and divided based on their hierarchical relationship to the parent slice. The angle of each slice is either divided equally under its parent node or can be made proportional to a value. Colour can be used to highlight hierarchal groupings or specific categories.

- Why a sunburst?

A sunburst represents an efficient way of viewing quickly a significant amount of data with different levels of dependencies. Most people are trained to understand a pie chart and a sunburst is just an evolution of a base pie chart. In 'An evaluation of space-filling information visualizations'[1] authors demonstrate that treemap and sunburst performance in terms of time and accuracy of completion of a task are roughly equivalent. Nevertheless, most participants of the study chose the sunburst over the treemap.Our visualization aims to highlight surebets dynamically.
The key functions of the sunburst are the following (exemples refer to the following figure):

- *arcTween()*
- *calculateNewPath()*
- *interpolatePathForRoot()* or *interpolatePathForNonRoot()*

[1]An evaluation of space-filling information visualizations, John Stasko, Richard Catrambone,Mark Guzdial and Kevin Mcdonald, 2000

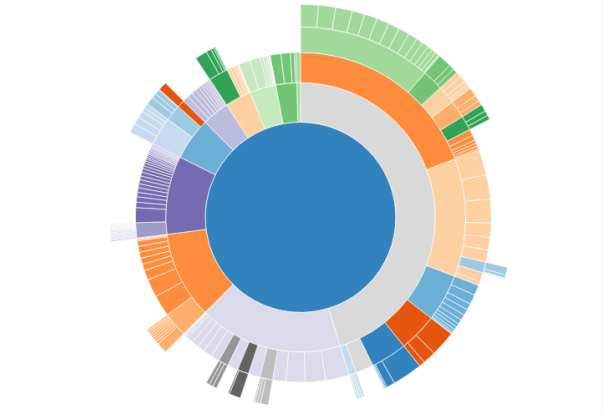

Fig. 2. Mike Bostock's Zoomable Sunburst

arcTween() is called for each click in the click event handler. Its input parameter is the data value corresponding to the clicked element.

calculateNewPath() is then called once for every path element, a total of 702 times for each click. Its input parameters are the data value and index of the path element.

The data has been formatted in a parent / child format, given the formatting, the nature of this visualization to be obvious. With tree data structures, the nodes must have a value associated with themselves and a parent in the tree.

interpolatePathForRoot() or *interpolatePathForNonRoot()* are called multiple times for each path element. Every call provides the input parameter t (for time) that represents the amount of progress in the current animation transition. The time parameter ranges from 0 when the animation starts to 1 when the animation ends. If, for example, D3.js requires 100 individual animation steps for the transition, then these functions will be called 70,200 times for each click.

Sunburst representations need the partition layout function of d3.js. It produces adjacency diagrams: a space-filling variant of a node-link tree diagram. Rather than drawing a link between parent and child in the hierarchy, nodes are drawn as solid areas (either arcs or rectangles), and their placement relative to other nodes reveals their position in the hierarchy. The size of the nodes encodes a quantitative dimension that would be difficult to show in a node-link diagram.

Our sunburst visualization was greatly based off Mike Bostock's Zoomable Sunburst example.

2.2 Grouped bar chart

Group bar charts represent another well-known method of visualization. It has a widely scope and is easily readable and understandable by most users.

Group bar charts are a way to show information about the different subgroups of several different main categories. In the example below, a group bar chart is used to display different states of the USA by age group. A separate bar represents each of the subgroups (eg, 65 years and over) and these are usually colored or shaded differently to distinguish them.

A caption should be represented to indicate to which color code what subgroup. The legend can be placed in the plot area or below the graph. Group bar charts can be used to show several subgroups of each category, but it is necessary to ensure that the chart does not contain too much information, making reading and interpretation difficult.

Grouped bar charts can be drawn in the form of horizontal or vertical charts depending on the nature of the data to be presented.

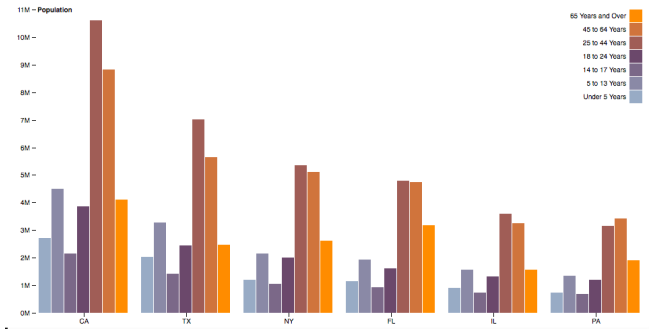

Fig. 3. Mike Bostock's grouped barchart

3 Project Description

3.1 Pre-processing

A mandatory phase of pre-processing is required to obtain the data fitted for the sunburst. We got the most interesting set of data on *http://www.football-data.co.uk/* because all the odds were already isolated and organised in the csv format and we developped a python script to transfer it to a json file ready to use with d3.js. This process allows to use exactly the informations (image file links, special attributes...) needed at any level of the visualization.

3.2 Sun burst

The first level of the visualisation represents the ten games played in a given day of a season of the "Ligue 1". The

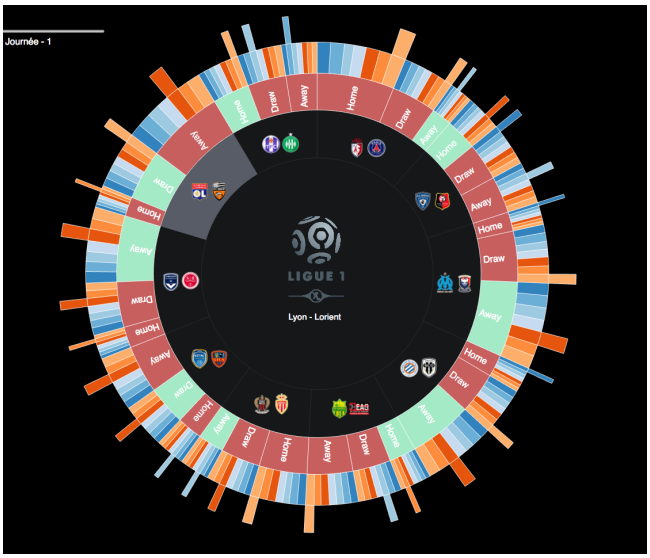

Fig. 4. Sun burst - root level

day and season can be chosen using the slider and the drop down menu located at the top-left corner. The root level includes the ten games and their children. Each children of a game consist in three odds, Home, Draw and Away, which are themselves the mean of their children, the odds of several bookmakers for the parent game. The last layout of the sunburst represents the greatest odds of all the bookmakers whose combination may append to be a sure bet. The color of the arc corresponding to the odds (Home, Draw, Away) of the true result of each game is coloured in green and the length of each arc is proportional to the odds value.

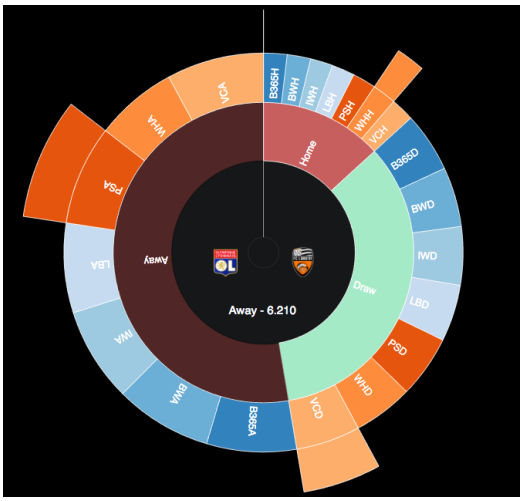

Fig. 5. Sun burst - game level

This sunburst by himself offers a novel and aesthetic way to illustrate the odds of different bookmakers for a given game.

Sure bet

This technique is based on the odds differences of different bookmakers on a given sports event. The surebet ensures the bettor takes advantage of the odds of at least two bookmakers to launch bets which, whatever the outcome of the game, will generate a certain profit (guaranteed profit). For a given game, we consider the maximum of the combination of the three (Home, Draw, Away) odds for all the bookmakers which is equivalent to:

$$BestBet = max(\frac{1}{Home} + \frac{1}{Draw} + \frac{1}{Away}) \qquad (1)$$

Theoretically, there is a sure bet if $BestBet > 1$ but some bookmakers withdraws a percentage from the betting and we chose to display only the best odds for each bookmaker in our visualization.

Tooltip

We added a tooltip that displays the information relative to the arc considered. If the mouse is over a level greater than the game level, it displays the value of the odds corresponding in addition to the arc information. At level 2, each arc represents a game and the tooltip returns the mean of the sum of all his children, namely the mean of the bookmakers odds for each of the 3 game issues.

3.3 Multiple bar chart interaction

Clicking on a game arc modifies the visualization in order to display the odds information of the prior games of each team in the form of a group bar chart. It computes the mean of the odds of the different bookmakers for each prior game and the stroke of the winning team is coloured in green again.

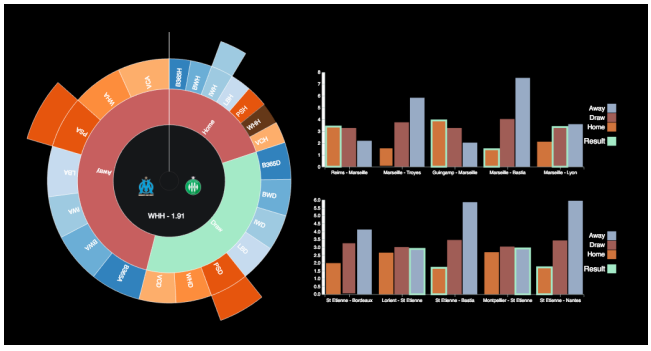

Fig. 6. Sun burst at game level and multiple bar charts for home and away team

This two bar charts offer an efficient way to sharply analyze the evolution of the bookmakers confidence in home and away teams for their respective previous games and the true result associated with these. It gives the user

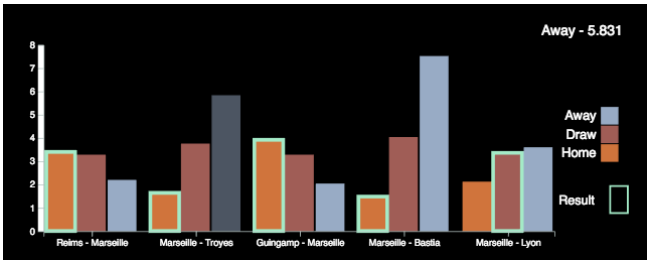

Fig. 7. Mutliple Barchart of 5 previous games

another possible indicator of a value bet (the fact that bookmakers can over or under-estimate the odds for a team) and it can be used to highlight the most predictable teams that is to say the teams with the highest true result / lowest odds ratio.

4 Discussion

Our main idea at the beginning was to create a prediction model and to visualize its predictions against the true results of different games. We thought that the variance of the odds between bookmakers was going to be a significant factor, but it did not add a lot of information to our predictions and we chose to only focus on the visualization. We thought this visualization as a tool to help us find and choose new various factors to build our prediction model. There are two main aspects that we would like to improve. It would be really interesting to analyse the evolution of odds within time for a same match i.e., bookmakers are constantly modifying the odds for a given match and treating these data as temporal series could lead to find new factors for a prediction model or to create an equivalent of a sure bet but with only the odds of one bookmaker. Moreover, it is possible to add a drop-down menu to modify what is displayed in the double mutliple bar chart and add more odds (half time result, number of goals...) in the future to widen our field of interpretation.

5 Conclusion

We have presented a novel way of visualizing football odds by merging two existing visualizations, Interactive Sunburst and Multiple Bar Chart. We have shown that our method propose an innovative way to navigate between several games and days of championship to better understand the odds evolution through time and to return the best value for each odds of a given game Our contributions are: (i) the unification of Interactive Sunburst and Multiple Bar Chart, (ii) an evaluation of the visualization with sports data. This work has shown that unifying two visualization techniques can substantially improve their efficiency and bring out the best of the data we study

Acknowledgments

We would like to thank Mr. Vuillemot for his support (even during the christmas holidays!), Mr. Rico for his advice in javascript, and Stack Overflow!

References

[1] Sunburst and bar chart *https://github.com/mbostock*