

CS5044 – d3.js practical

TOP 5 FOOTBALL LEAGUES TRANSFER SPENDING



WORD COUNT: 946

Website linkS:

<https://aml29.host.cs.st-andrews.ac.uk/D3/infoVis/index.html>

<https://aesc1.host.cs.st-andrews.ac.uk/D3/infoVis/index.html>

# DATA & QUESTIONS

The football transfer market has transformed and changed over the years. Using the power of JavaScript and D3.js we will be exploring the following question:

* How do the top 5 league transfer spending’s compare with each other over the seasons and what players positions are mostly involved in these transfers?

To answer this question, we will be using data obtained from [Kaggle](https://www.kaggle.com/vardan95ghazaryan/top-250-football-transfers-from-2000-to-2018) [1]. Our chosen dataset has recorded details of the 250 most expensive transfers each season. The data collection holds data for all the leagues; however, we will be working with a subsample of it which involves only the top 5 football leagues which are 3339 from the original 4700 observations:

1. Premier League
2. Bundesliga
3. LaLiga
4. Ligue 1
5. Serie A

The dataset provides us with 10 attributes, but we will be using only four of them:

|  |  |  |
| --- | --- | --- |
| Attribute Name | Attribute Type | Attribute Meaning |
| Season | Ordinal (String) | Season transfer was made |
| Position | Nominal (String) | Player’s playing position involved in transfer |
| League\_to | Nominal (String) | The league the player is being transferred to |
| Transfer\_fee | Numerical (Integer) | The transfer fee amount in € |

*[TABLE 1.1] Showing the attributes that were used in whole visualisation*

# IMPLEMENTATION

As a beginning stage we followed a ‘5-sheet ideation process’ to brainstorm multiple options of expressing the data

When we first started undertaking the project in terms of coding, we were not familiar with the D3.js technology. For this reason, we worked on a simple solution and iteratively added more complex features to improve the final result. References can be found in the code [2] [3] [4].

# DESCRIPTION OF VISUALISATION

Each data attribute has been encoded with a specific visual variable to allow the visualisation to communicate with the user in an expressive and effective manner. In addition to this two of the attributes have been used as filters to allow the user to interact with data and the visualisation state.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ATTRIBUTE | ATTRIBUTE TYPE | VISUAL VARIABLE | REASON OF VISUAL VARIABLE | EXPRESSIVE(YES/NO) |
| Season | ORDINAL | POSITION | Showing the sequential order of the seasons | YES |
| Position | NOMINAL | POSITION | Each position indicates a different value | YES |
| League\_to | NOMINAL | COLOUR(HUE) | Different colour indicates a different league | YES |
| Transfer\_fee | QUANTITATIVE | LENTGH/POSITION | Larger length or furthest position means larger transfer amount | YES |

*[TABLE 1.2] Showing the attributes used in visualization with their visual variable used to encode it.*

Graphical user interface, chart

Description automatically generated

*[FIGURE 1.1]  Shows the 4 different perspectives of the data which the user will use*

From Layout Box [1] = We use a horizontal bar graph to show the sum of all the top 5 leagues transfer fee spending for each season. The bar with the largest length is the season with the largest spending.

From Layout Box [2] = We use a line-dot graph to explore in more detail each league's spending for each season. The position of the dots shows the transfer fee spent for each season, while the lines show the change from one season to the other.

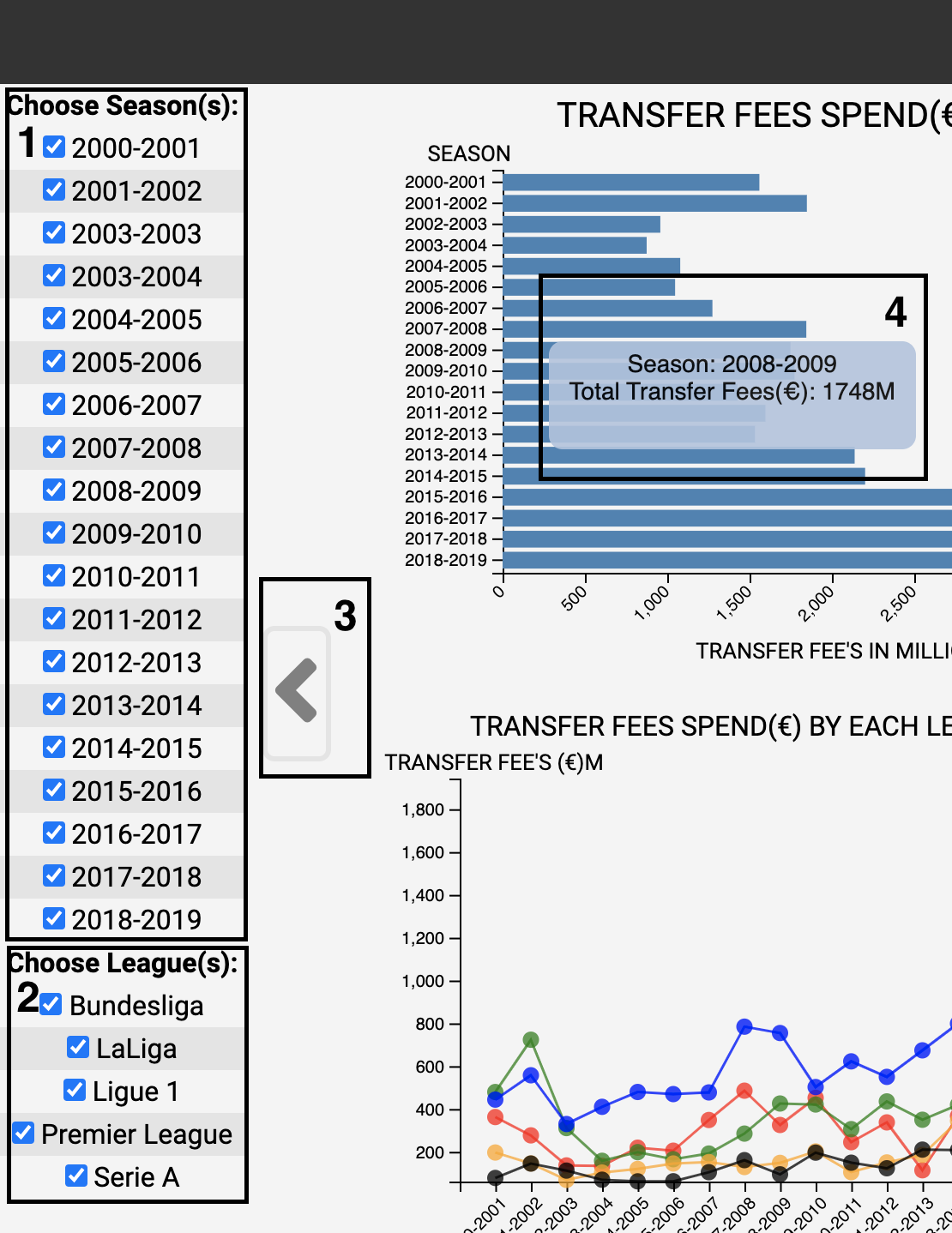
From Layout Box [3] = We use a horizontal lollipop graph to show the sum of all the top 5 leagues transfer fee spending for each unique footballer’s position. The ‘lollipop’ with the largest line length and the positioned circle further away from y-axis is the position mostly spent on.

From Layout Box [4] = We use a dot graph to explore in more detail in which positions each league is investing in. The different leagues are encoded using the visual encoding colour.

From Layout Box [5] = This is the legend box for the line graph and dot graph. Indicates which league each colour is presenting.

|  |  |  |  |
| --- | --- | --- | --- |
| ATTRIBUTE | ATTRIBUTE TYPE | INTERACTION TYPE | REASON |
| Season | NOMINAL | (CHECKBOX)FILTER | To allow the user to filter out data not interested in. Checkbox to allow user compare seasons that are not sequential. |
| League | NOMINAL | (CHECKBOX)FILTER | To allow the user compare only leagues interested in. |

*[TABLE 1.3] Showing the attributes used in visualization as interactions for the user*



*[FIGURE 1.2] Shows the interactions the system supports*

From Layout Box [1] = The user can use the checkboxes to filter in/out the seasons mostly interested in.

From Layout Box [2] = The user can use the checkboxes to filter in/out the leagues mostly interested in.

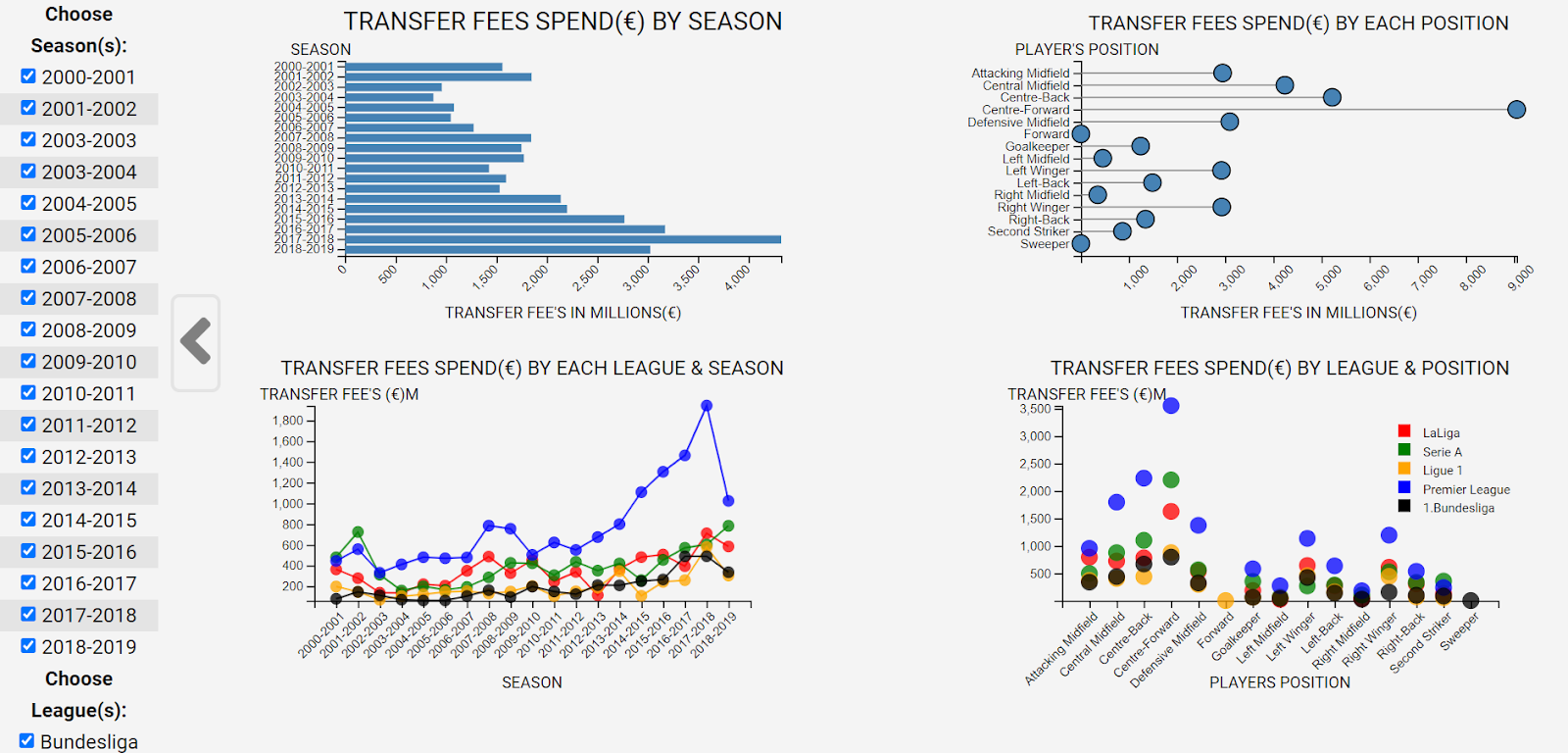
From Layout Box [3] = The user can hide the filters, so the graphs can take more space.

From Layout Box [4] = The user can use the mouse pointer to go over graph positions and obtain details on each point. The pointer gives a detailed summary of what each point is showing.

# INSIGHTS FROM THE VISUALISATION

## General Insights

Users are able to manipulate the different graphs by interacting with the filters.



*[FIGURE 1.3] An Overview of the Visualisations with Filters*

At first glance, without removing any filters, it is evident that the spending tends to increase each season (bar graph), with the most expensive position clearly being centre-forward (lollipop graph).

## Specific Season Insights

Diagram

Description automatically generated with medium confidence

*[FIGURE 1.4] Comparison Between Transfer Fees Spend by League/Season and League/Position*

Using the Season filter provided [1], a user can select specific seasons and assess each leagues spending. From [2] and [3] and with the legend box, we can observe the premier league is spending the largest and the positions they are mostly investing in are ‘Centre Forward’ and ‘Centre Back’.

## Comparing century spending

Chart

Description automatically generated

*[FIGURE 1.5] Filtered data for the first century, investigating position spending*

*Chart

Description automatically generated*

*[FIGURE 1.6] Filtered data for the second century, investigating position spending*

Looking at the lollipop graphs from the two comparisons above, it may be noted that the disparity between how much is spent in the centre forward position relative to the other position may have seen a decrease in the second decade of the 2000s.

## Comparing the slope change of spending of each league

Chart, diagram

Description automatically generated

*[FIGURE 1.7] Filtered data first & last season to compare slope change of league spending*

Filtering data for only the two seasons will allow a user to visualise the slope change in terms of transfer fee spending for each league. It is evident that season 2000-2001 Serie A was the highest spending league, however in 2018-2019 the Premier League was the highest spending league. We can see there is a steeper slope change for the Premier League. In addition to this, from the bar graph it is noted that the spending across all leagues has doubled in 2018-2019.

Diagram

Description automatically generated with low confidence

*[FIGURE 1.8] Filtering data for first & last season only for premier league*

To visualize the same data for specific league(s), a user could simply filter in only the leagues interested in by ‘checking’ the appropriate filter boxes of leagues.

# CRITICAL DISCUSSION

Although the final design/dashboard gives an overview and a detailed perspective of the data, it still has some limitations. Firstly, no zoom interaction has been implemented, for some users the graphs might be considered to be small. Secondly, we would have liked to implement a details-on-demand feature that would show to the user the most expensive football player at each data point. Thirdly, a large part of the data is time-based. Another limitation is not having the ability to see an automatic time-lapse that would show the change of these characteristics in real time. This would allow the user to examine the change from one season to a next in an automatic way. These are the three features we will further explore and implement to accommodate the user’s needs.

# REFERENCES FOR CODE

[1] Datasets, K. (n.d.). *Kaggle*. Retrieved from Kaggle: <https://web.microsoftstream.com/video/18f8f66e-c21d-4718-92eb-c2c86dfe2318>

[2] D3.JS. (n.d.). *bl.ocks.org*. Retrieved from <https://bl.ocks.org>

[3] Stack-Overflow. (n.d.). *Stack-Overflow*. Retrieved from https://stackoverflow.com.

[4] D3-Gallery. (n.d.). *D3-Gallery.org*. Retrieved from <https://www.d3-graph-gallery.com>

# REFERENCES FOR EXPRESSIVENESS

[5] Jonathan C. Roberts, C. H. (n.d.). *etching designs using the Five Design-Sheet Methodology.* Retrieved from http://fds.design/wp-content/uploads/2015/10/fds- presentation-final-ieeevis2015.pdf

[6] Book for effective visuals : **Design for information: an introduction to the histories, theories, and best practices behind effective information visualizations**