# Premier League Five Year Review 2016 – 2019



M20180667 – Miguel Falcão e Silva M20180510 – Guillaume Ribeiro M20180542 – Jorge Teixeira







## **Data Description**

The data used for this project consists in five seasons (2014/15, 2015/16, 2016/17, 2017/18, 2018/19). Each season represents 380 games, for a total of 1900 games played in five years. Every row contains the following information:

Variables	Definition
Div.	Season Code
Date	Day of the Match
Home	Home Team
Away	Away Team
FT Home Goal	Goals Scored by Home Team at Full Time
FT Away Goal	Goals Scored by Away Team at Full Time
FT Result	Result at Full Time
HT Home Goal	Goals Scored by Home Team at Halftime
HT Away Goal	Goals Scored by Away Team at Halftime
HT Result	Result at Half Time
Referee	Name of the Match Referee
Home Shot	Home Team Shot Fire on Goal Area
Away Shot	Away Team Shot Fire on Goal Area
Home Shot	
Target	Home Team Target Fire on Goalkeeper Area
Away Shot Target	Away Team Shot Fire on Goalkeeper Area
Home Fouls	Total Number of Fouls by Home Team
Away Fouls	Total Number of Fouls by Away Team
Home Corners	Total Number of Corners by Home Team
Away Corners	Total Number of Corners by Away Team
Home Yellow	
Card	Total Number of Yellow Cards by Home Team
Away Yellow Card	Total Number of Yellow Cards by Away Team
Home Red Card	Total Number of Red Cards by Home Team
Away Red Card	Total Number of Red Cards by Away Team
B365 Home	Odd for the Home Team - Win
B365 Draw	Odd for Draw
B365 Away	Odd for the Away Team - Win

#### **Visualization and Interaction Choices**

The Premier League is England 's football association's top division. The competition was formed in 1992 and the last edition – season 2018/19 – was the 27th edition (Premier League - Home Page. n.d.). Every season, 20 teams fight for title of champion (19 games home, 19 games away) in a total of 380 games. The season itself goes from August to May with most games occurring on Saturday and Sunday afternoons (Premier League - Home Page. n.d.). Of the 20 teams that compose the Premier League, some of the most iconic teams such as Manchester United, Liverpool, Arsenal, Chelsea, Tottenham and Manchester City are amongst the biggest clubs in Europe.

According to Jones (2019) money report, Europe's Top 20 Team had a revenue of £8.3 billon in the season 2017/2018, which represents 6% increase from the previous season. The



revenue originated 43% from broadcast rights, 40% from commercial and 17% from matchdays. The Primer League has nine teams in the top 20 of the Big Five Leagues (Spain, England, Germany, France and Italy), from which Six teams are in top 10, the highest being Manchester United in 3<sup>rd</sup> place.

The reason why we choose the Premier League data is the most profitable. According to Holmes (2018), the European football market "is now worth £22.8 billion, with the English Premier League being the market leader. Premier League reported record revenue of £4.5 billion for 2016/17, which marked a 25 per cent increase on the previous year. Revenue was boosted by the three-year domestic TV rights deal worth £5.1 billion, signed in 2015". As of 2018, the Premier League generated over £1.7 billion ever year in TV rights. During this five-year period (2015-2019), the Premier League assisted to one unexpected event and many recordings broken:

- 2014/2015 Fourth Title for Chelsea
- 2015/2016 First Title for Leicester City
  - o First outsiders to became champion in one season.
- 2016/2017 Fifth Title for Chelsea
  - o Record wins by any team in a Premier League season (30)
- 2017/2018 Third Title for Manchester City
  - First side to win Premier League with 100 points. Most difference to second place (19 points)
  - Record wins by any team in a Premier League season (32)
  - Highest goal difference (79).
- 2018/2019 Fourth Title for Manchester City
  - o First team to successfully defend the Premier League title for 10 years.
  - One of the most thrilling title races in the competition's history by one-point difference.
  - Record on top two combined points (Man City and Liverpool 195)

(Retrieved from Premier League - Home Page, n.d.)

It is possible to conclude that the recent seasons can explain the growth in the demand for Premier League games. The base of the revenue for the Premier League are the matchdays and these have consequential outcome to commercial aspects of clubs (marketing and apparels) as well as broadcast rights (league and European games). The outcome of these matches – game statistics – need to the provide to the public, allowing them a fast and yet reliable assimilation what occurred in a season with one glance. By using R Studio, we propose to use data visualization to achieve our goal. Upon visiting our page, users can access a table and heat map.

### **Reading the Visualization**

#### Table

For the table, the user will have access to an overall view of the season. This graphic is interactive as the user can choose the season. At first, we decided to use year however since the season goes from August to May, we decided to use the Div. variable. Therefore, the user selects the div code representing on season via an input function. For this graph, we had to create a code to compute points between the different matches, set differences for tie breaker and order the final result.



A win is 3 points, a draw is 1 point and a defeat does not give point, hence 0 points. For the tie break, the code will look first into the goal difference and second to away /home game victories won. The results are lock to a position that represents the final stance in the season. The user can order by ascend or descend the following variables – points, games, wins, draws, losses, goals scored, goals taken and goals difference – in order to access the best and worst team in each variable. We created this table based on the criteria of the eight rules of thumb of data visualization. According to Cabral (2019), these rules are:

- 1. No unjustified 3d
- 2. No unjustified 2d
- 3. Eyes Beat Memory
- 4. Resolution over Immersion
- 5. Overview First, Zoom and Filter, Detail on Demand
- 6. Responsiveness is Required
- 7. Get It Right in Black and White
- 8. Function First, Form Next

(Retrieved from Class 3b – Rules of Thumb)

The table was designed for all types of users, emphasizing the point-and-click access. We opted for a black and white design since this "ensure that the most crucial aspects of visual representation" (Cabral, 2019), assuring that readability is achieved in its most simple form.

### **Heat Map**

With the heat map, the user can have access to discipline record of all seasons or a specific season, choose the between yellow and red card and choose the game type at home or away game. This was achieved via three select input, displayed in a row. According to Cabral (2019), heatmap consist in "datasets with two keys are often arranged in a two-dimensional matrix alignment where one key is distributed along the rows and the other along the columns, so a rectangular cell in the matrix is the region for showing the item values".

For this heat map (see figure 1), the two-dimensional matrix focusses the amount of yellow and red cards shown by club for each referee in the league. We decided to use yellow-orange-red range, as yellow is 0 or close to 0, red for high number and a brown color for the highest number presented in the data originally extracted.

Figure 1 – Heat Maps Objectives

Idiom	Heatmaps
What: Data	Table: two categorical key attributes (genes, conditions), one quantitative value attribute (activity level for gene in condition).
How: Encode	2D matrix alignment of area marks, diverging color- map.
Why: Task	Find clusters, outliers; summarize.
Scale	Items: one million. Categorical attribute levels: hun- dreds. Quantitative attribute levels: 3–11.

(Retrieved from Class 4a – Visualization Tables)

We developed the heat map to find clusters of the referees more prone to show cards and summarize which clubs are more subject to scrutiny. Again, we emphasized the click and point access. The users of this heat map would be a football manager or a player to have in



mind the history of the referee towards the club they represent, could be a bookmaker to adjust the odds of which team will get first card or player expelled and for the supporter to have sensibility when placing a bet for its team to win as the referee decision. Though impartial, the referee decisions can create crucial moments in the game that can determine the end result, affecting the outcome of supporter's bet and odds. Both the table and the heat map can be accessed at <a href="https://github.com/GuillaumeRibeiro/DVFinalProject">https://github.com/GuillaumeRibeiro/DVFinalProject</a>.

#### Issues with R Studio

We faced some limitations regarding R Studio. Creating all three visualizations had their difficulties as it was complicate to identify and update the errors. As mentioned previously, the season 2018 and 2019 was one of the most competitive as Manchester City and Liverpool. We decided to create an interactive map to show our users how the season unfolded from August 2018 to May 2019, for which we decided to use an interactive slideshow similar this:

https://preview.redd.it/60ne82rxgwk21.gif?format=mp4&s=59e31a54443aeb39f5ccbc460112d963b975d3fe

According to Segel & Heer (2010), "interactive slideshow follows a typical slideshow format but incorporates interaction mid-narrative within the confines of each slide (...) this structure allows the user to further explore particular points of the presentation before moving ahead to the next stage of the story". We opted for this type of visualization as it "allows for interaction mid-narrative, a more balanced mix of author-driven and reader-driven approaches" (Segel & Heer, 2010) and, as the visualization is single-frame interactivity is achieved month by month, as we tell the story of the 2018/2019 season, the user can see movement in the standings and make assumptions of what occurred.

In the situation we tried to replicate, the user can see the monthly development of the season via a progress bar, see the different position of all teams that competed in the season via the crest, which provides a consistent visual platform for the user. We developed the interactive map to demonstrate how important narrative visualization can be, the ability to tell stories with data graphics. As the user click on the interactive map, it starts until it completes the entire season. As the interactive map took a great time to upload and it was not easy to access club logo images in to use in the interactive map. We were not able to complete this as the code used had the following issues:

```
> animate(q, nframes = 300, fps = 15, start_pause = 30, end_pause = 30, height = 800, width = 800 )
Error in animate(q, nframes = 300, fps = 15, start_pause = 30, end_pause = 30, :
    could not find function "animate"
>
> anim_save("prem position.gif", path = "documents")
Error in anim_save("prem position.gif", path = "documents") :
    could not find function "anim_save"
```

After some research, we realized that the both codes belong to a package gganimate, which our version 3.3.3 did not support. As extra credit, we decided to two other extra graphs for a total of five graphs – a parallel graph to measure the coefficients between the odds and the end results of the games and stacked bars to work as game performance indexes – which we had to drop due to time limitations. Yet, the most difficult task was to assemble the three works in one single file, mainly issues on separating creating and separating three tabs.



#### Conclusion

The objective of this work was to create visualization solution for a database of our choice. We decided to analysis the last five seasons of the Premier League from which we conclude: Leicester winning the championship was an unexpected event as the club finished mid table in the following season, winning teams have a lower amount of card penalization that rest of the competition and season 2018-2019 title run reached a final decision when match week 29 when Liverpool draw with Everton and Manchester City won Bournemouth.

Based on our research to support the decision to choose the Premier League, we found out extensive material on data visualization for analysis purpose, especially the Deloitte infographic and slideshows. It is our strong believe that Data Visualization will be used further on in football analysis cross-functionally. Nowadays, clubs used data visualization tools to represent financial situations, game statistics and player's physical performance. In the future, once can tag the player performance to correlate game statistics, measure its impact towards the club financial situations, providing all parties with an accurate value of the player to year to date by accessing just the name or photograph of the players in the club's website.



## References

Cabral, P. (2019). Class 3b – Rules of Thumb.

Cabral, P. (2019). Class 4a – Visualizing Tables.

Cabral, P. (2019). Class 5c – Storytelling.

Data Files: England. (2019, May 15). Retrieved from <a href="http://www.football-">http://www.football-</a>

## data.co.uk/englandm.php

Holmes, E. (2018, June 7). 'Big five' European soccer leagues boast record revenues for 2016/17. Retrieved from <a href="http://www.sportspromedia.com/news/big-five-european-soccer-leagues-record-revenues">http://www.sportspromedia.com/news/big-five-european-soccer-leagues-record-revenues</a>

Jones, D. (2019, January 24). Deloitte Football Money League 2019 Deloitte UK. Retrieved from <a href="https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/deloitte-football-money-league.html">https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/deloitte-football-money-league.html</a>

Premier League - Home Page. (n.d.). Retrieved June 20, 2019, from <a href="https://www.premierleague.com">https://www.premierleague.com</a>

Segel, E., & Heer, J. (2010). Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1139-1148. doi:10.1109/tvcg.2010.179

