

A soccer ball with black and white panels is shown hitting a white goal net. The background is a clear blue sky. The net is in the foreground, and the ball is slightly behind it, creating a sense of depth.

# DATA422-DATA WRANGLING

Anirudh Revathi-36899816

Gus Pawson-63949878

Julian Maranan-81456731

Pretty Mathew Punnoor -34541658

Uditha Mayadunna-39768450



We will look at how many points and how much money the top 7 teams in the Premier League and La Liga earned and spent from 2015 to 2022.

We focused on the top 7 teams because finishing in the top 7 qualifies a team for European competitions, such as the Champions League and Europa League.

We'll also explore whether there's a connection between the money spent on wages and the points earned by the teams. In other words, we want to see if higher wages are linked to more points

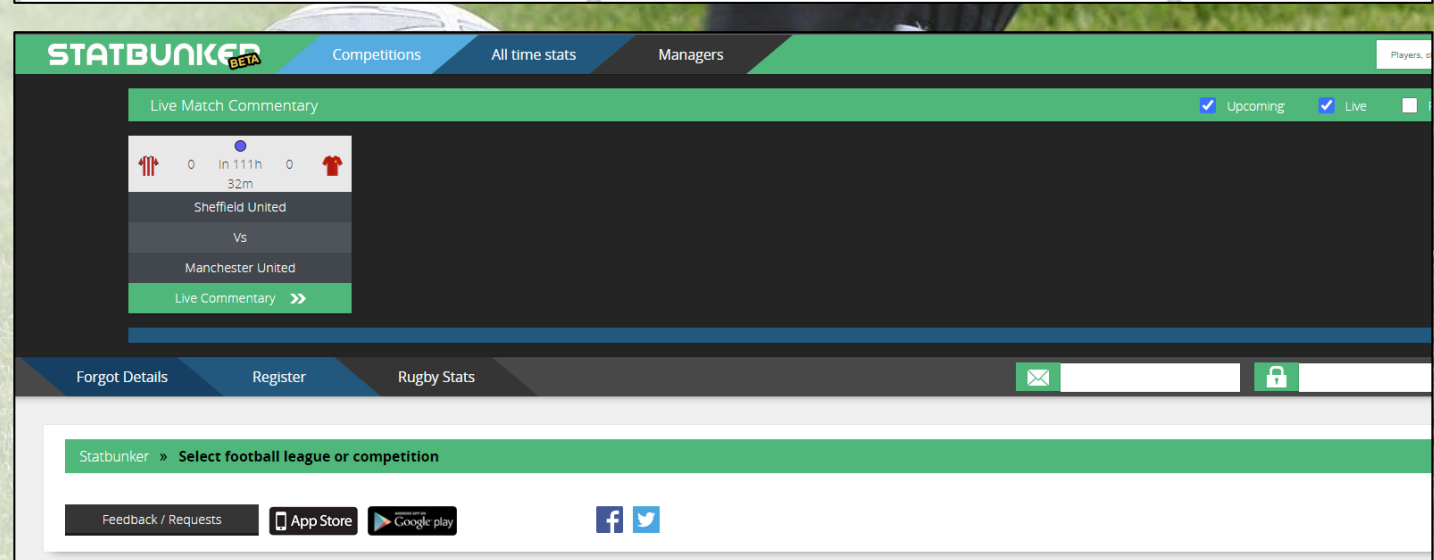
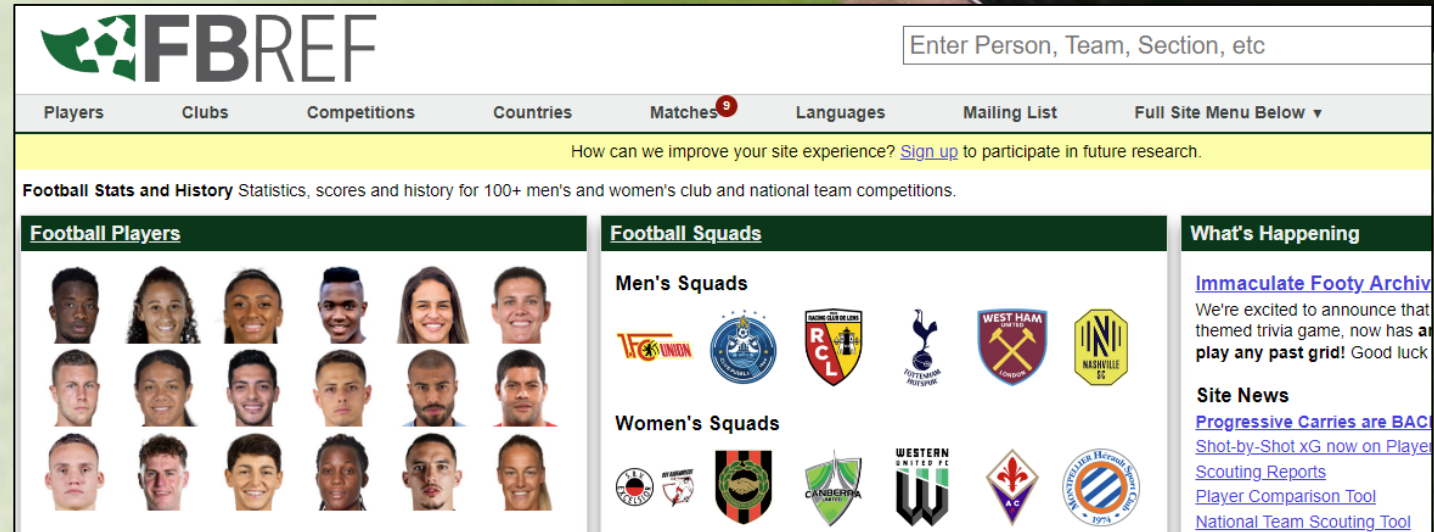




# Data Scrapping

## Sources:

- [www.fbref.com](http://www.fbref.com)  
All key team statistics for both leagues
- [www.statbunker.com](http://www.statbunker.com)  
Individual player statistics – including fantasy points



# Data Scrapping











## Challenges:

Incorrect data types

Extracting only the necessary information

Lack of season and league information

Unable to form primary and foreign keys

Overall		Home/Away								
Rk	Squad	MP	W	D	L	GF	GA	GD	Pts	
1	 <a href="#">Tottenham</a>	8	6	2	0	18	8	+10	20	
2	 <a href="#">Arsenal</a>	8	6	2	0	16	6	+10	20	
3	 <a href="#">Manchester City</a>	8	6	0	2	17	6	+11	18	
4	 <a href="#">Liverpool</a>	8	5	2	1	18	9	+9	17	
5	 <a href="#">Aston Villa</a>	8	5	1	2	19	12	+7	16	
6	 <a href="#">Brighton</a>	8	5	1	2	21	16	+5	16	
7	 <a href="#">West Ham</a>	8	4	2	2	15	12	+3	14	
8	 <a href="#">Newcastle Utd</a>	8	4	1	3	20	9	+11	13	
9	 <a href="#">Crystal Palace</a>	8	3	3	2	7	7	0	12	
10	 <a href="#">Manchester Utd</a>	8	4	0	4	9	12	-3	12	



# Data Scraping

## Collecting necessary information

Multiple currencies +  
character type



Regular Expressions



Clean data

Weekly.Wages	Annual.Wages
<chr>	<chr>
€ 6,766,923 (£ 5,674,122, \$6,896,250)	€ 351,880,000 (£ 295,054,323, \$358,605,024)
€ 6,044,038 (£ 5,067,977, \$6,159,550)	€ 314,290,000 (£ 263,534,797, \$320,296,616)

Weekly_Wages_pounds	Annual_Wages_pounds
<dbl>	<dbl>
5674122	295054323
5067977	263534797

# Data Scraping

## Extracting year and season

Fbref.com - Use information contained in link

/2022-2023-Premier-League-Stats

Statbunker.com – Extract information from drop down menu

La Liga 22/23

Season	League	Champions	Games_Won	Top_Goal_Scorer	Goals_Scored
<chr>	<chr>	<chr>	<dbl>	<chr>	<dbl>
2019-2020	La Liga	Real Madrid	26	Lionel Messi	25
2019-2020	Premier League	Liverpool	32	Jamie Vardy	23
2020-2021	La Liga	Atlético Madrid	26	Lionel Messi	30
2020-2021	Premier League	Manchester City	27	Harry Kane	23
2021-2022	La Liga	Real Madrid	26	Karim Benzema	27



# Data Scrapping

## Creating Keys:

Season ID = Season + League

Team ID = Season + Team

Player ID = Season + Player

Season_ID	Season	League	Team_ID	Champions	Player_ID	Top_Goal_Scorer
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
19/20-LL	2019-2020	La Liga	19/20-Real Madrid	Real Madrid	19/20-Lionel Messi	Lionel Messi
19/20-PL	2019-2020	Premier League	19/20-Liverpool	Liverpool	19/20-Jamie Vardy	Jamie Vardy
20/21-LL	2020-2021	La Liga	20/21-Atlético Madrid	Atlético Madrid	20/21-Lionel Messi	Lionel Messi



# Data Scraping

## The Result

---

Relational database comprised  
of seven datasets

Close to 30,000 rows of data

Over 100 distinct variables





A close-up, low-angle shot of a soccer ball hitting a goal net. The ball is white with black hexagonal panels and is caught in the white mesh of the net. The background is a clear blue sky. The lighting is bright, suggesting a sunny day.

# CHAMPIONS AND WAGES COMPARISON

We have used functions like “select”, “filter” in R and “combine”, “groupby” in Julia to target variables that we wanted to utilise for analysis and visualisation

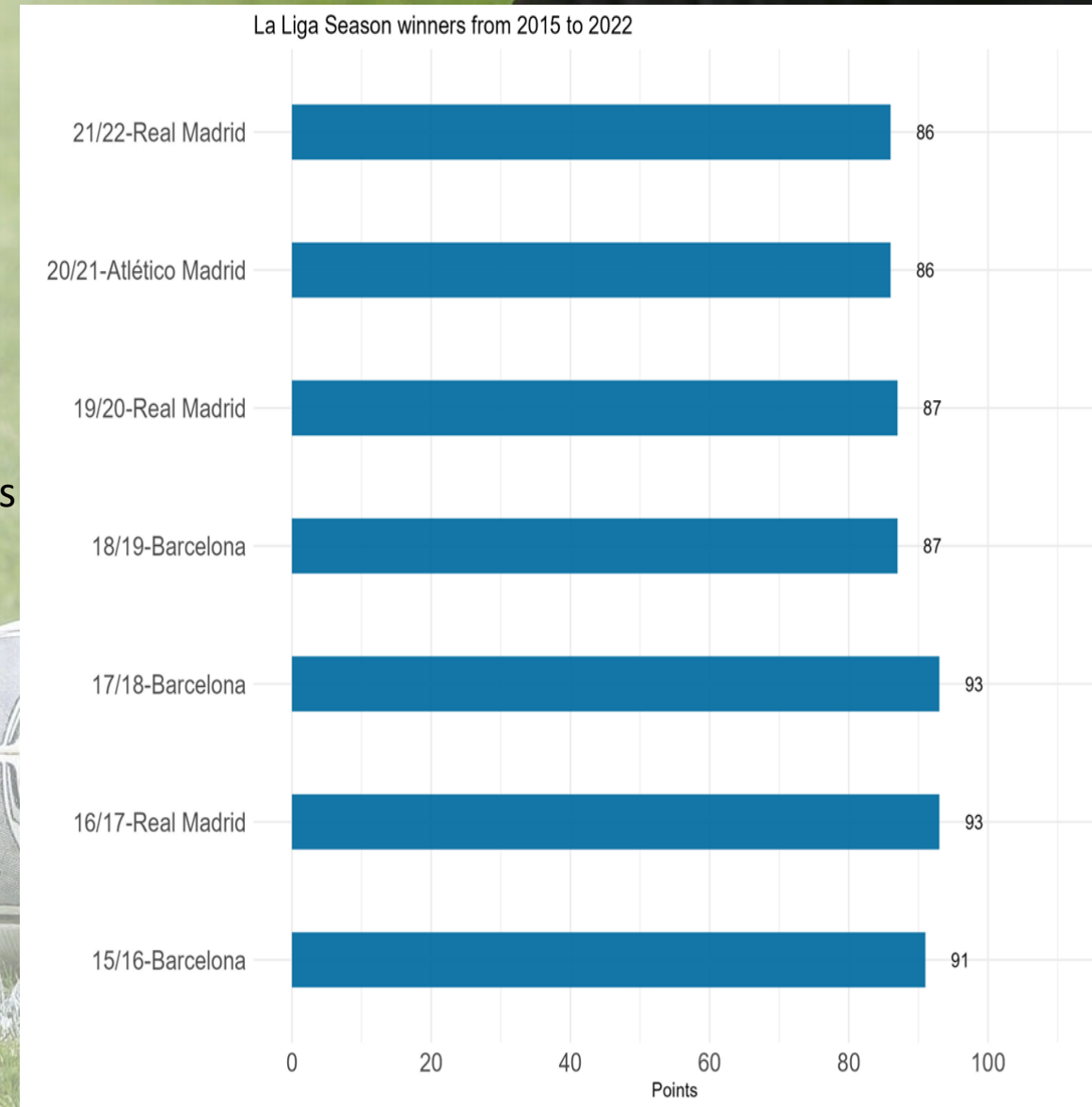
```
'data.frame': 177 obs. of 11 variables:
 $ X          : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Season_ID  : chr   "13/14-LL" "13/14-LL" "13/14-LL" "13/14-LL" ...
 $ Season     : chr   "2013-2014" "2013-2014" "2013-2014" "2013-2014" ...
 $ League     : chr   "La Liga" "La Liga" "La Liga" "La Liga" ...
 $ Rk         : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Team_ID    : chr   "13/14-Barcelona" "13/14-Real Madrid" "13/14-Atlético Madrid" "13/14-Valencia" ...
 $ Squad      : chr   "Barcelona" "Real Madrid" "Atlético Madrid" "Valencia" ...
 $ X..Pl      : int  33 35 41 43 44 32 31 34 47 31 ...
 $ Weekly_Wages_pounds: int 2874391 2188313 784925 501138 406322 299283 266936 263130 239104 235314 ...
 $ Annual_Wages_pounds: int 149468309 113792296 40816075 26059168 21128735 15562715 13880669 13682780 12433400 12236353 ...
 $ X..Estimated : chr   "100%" "100%" "100%" "100%" ...
```



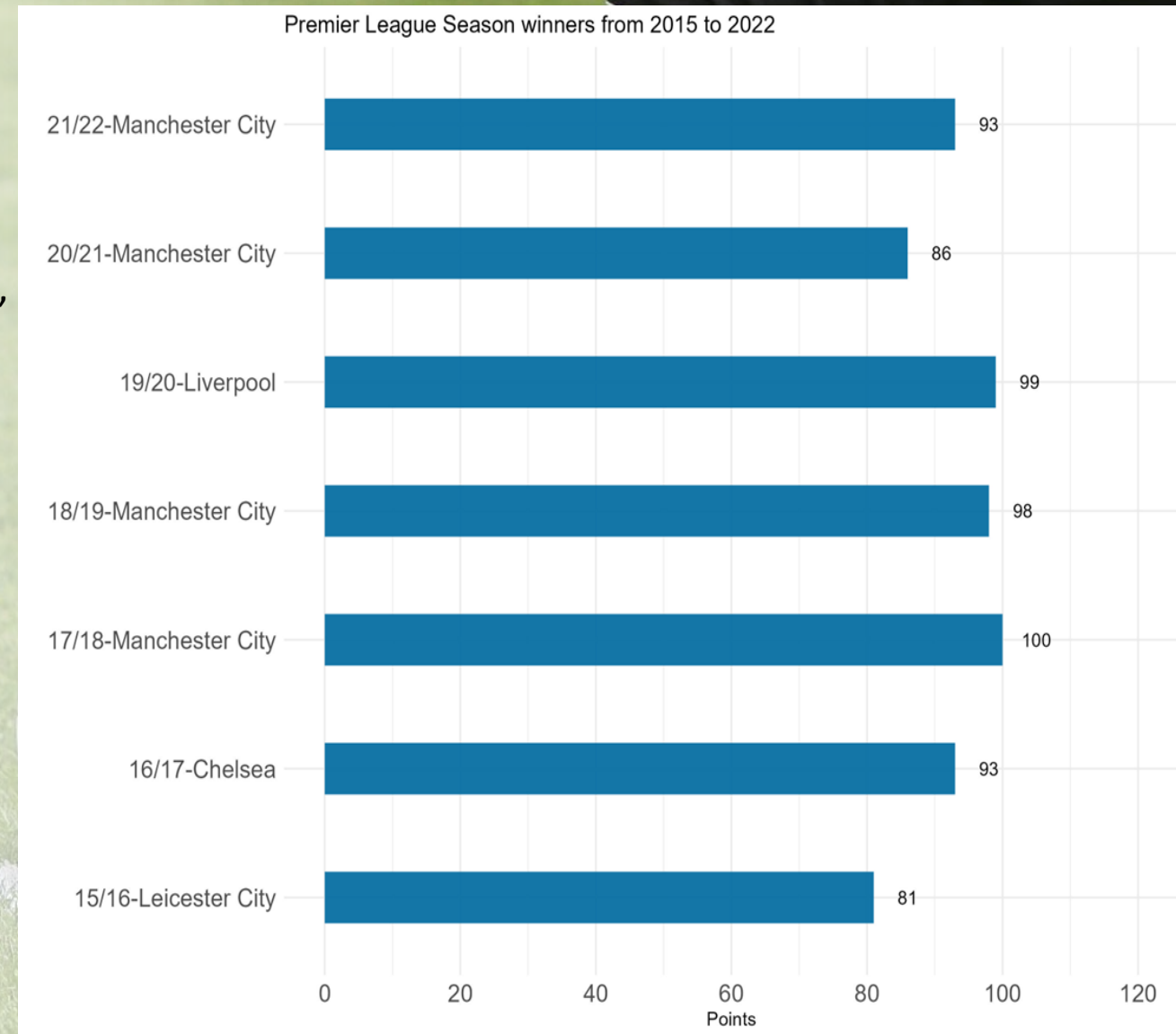
```
'data.frame': 137 obs. of 5 variables:
 $ Season     : chr   "2015-2016" "2015-2016" "2015-2016" "2015-2016" ...
 $ Rk         : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Squad      : chr   "Barcelona" "Real Madrid" "Atlético Madrid" "Valencia" ...
 $ Weekly_Wages_pounds: int 2806246 2439140 1015079 813466 552109 398662 374055 338096 316698 299219 ...
 $ Annual_Wages_pounds: int 145924771 126835291 52784100 42300237 28709687 20730442 19450878 17581005 16468305 15559362 ...
```



- The points scored by the season winners varied across the years.
- Based on the observations, it's noted that the winning teams in La Liga have earned points ranging from 86 to 93 in the past seven seasons, indicating a variation in performance among the top teams.
- Although La Liga consists of 20 teams, the teams listed above, such as Barcelona, Real Madrid, and Atlético Madrid, tend to be the frequent winners.

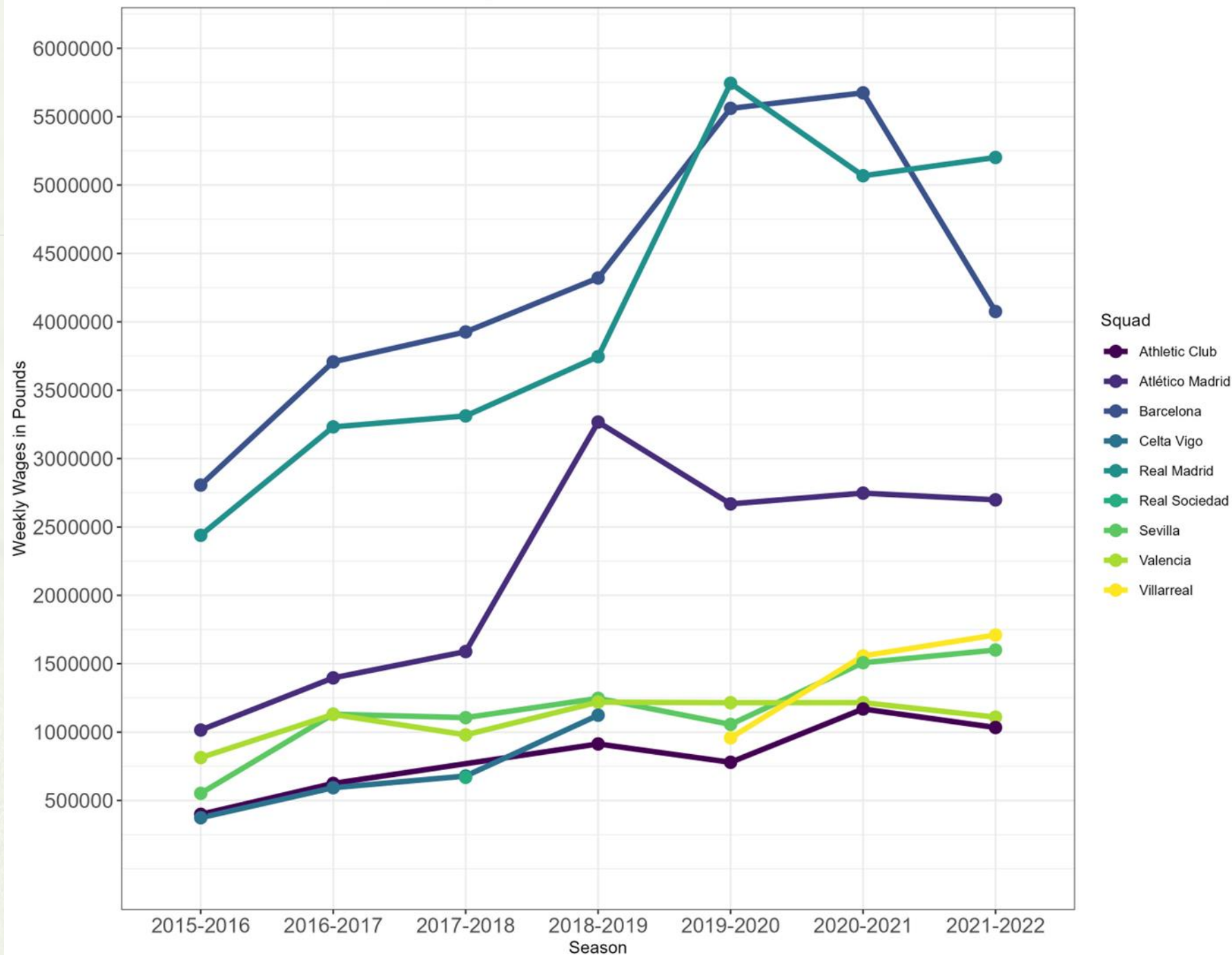


- Manchester City won the title in the 2017-2018, 2018-2019, 2020-2021, and 2021-2022 while consistently managing to earn above 90 points.
- The point totals for Premier League winners have been generally higher, with one season's winner reaching 100 points.

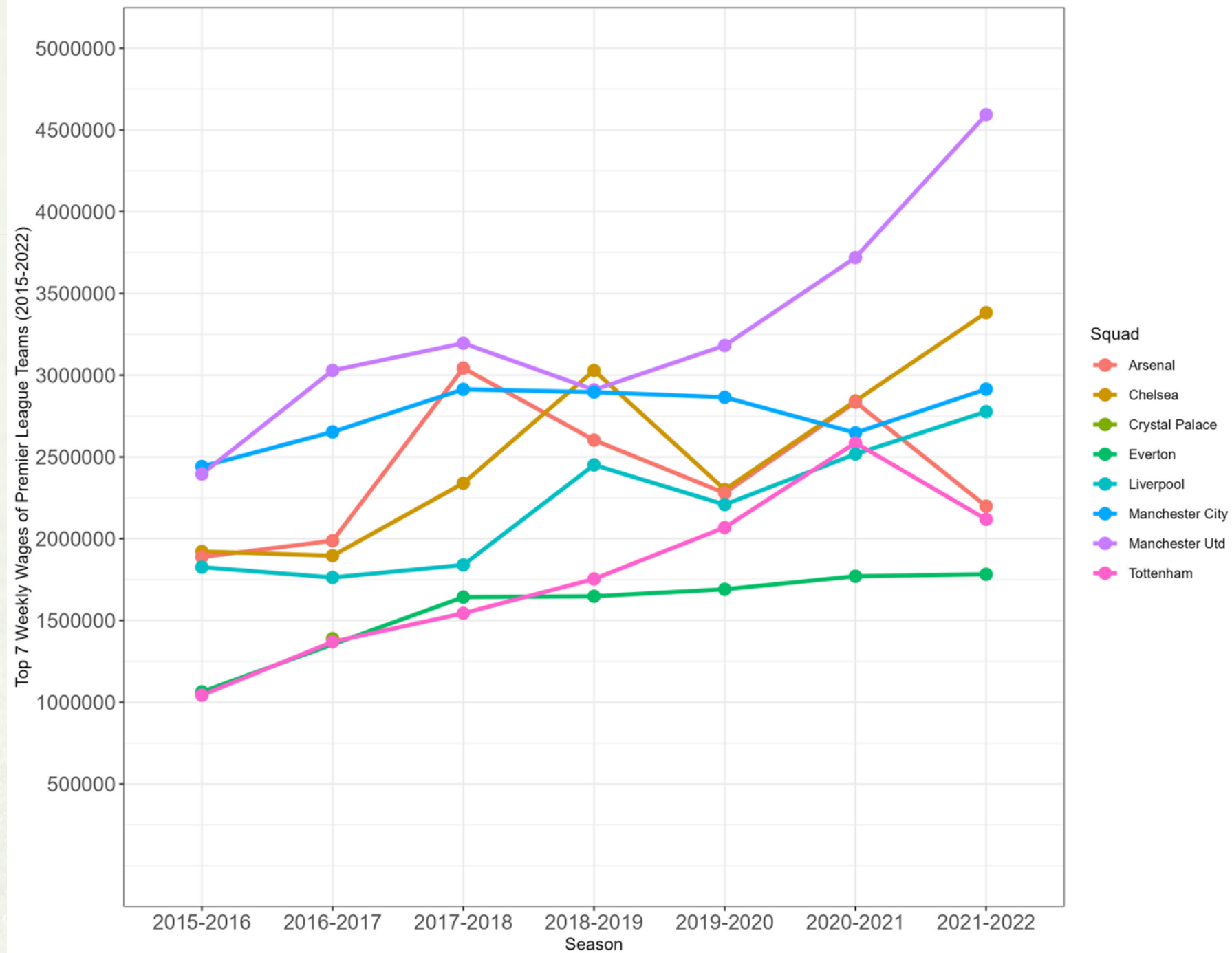




Top 7 Weekly Wages of La Liga Teams (2015-2022)



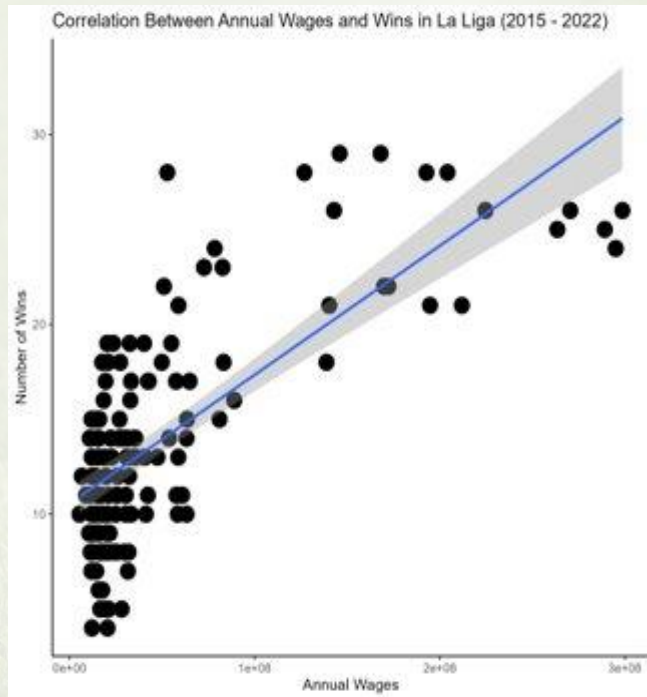
Top 7 Weekly Wages of Premier League Teams (2015-2022)



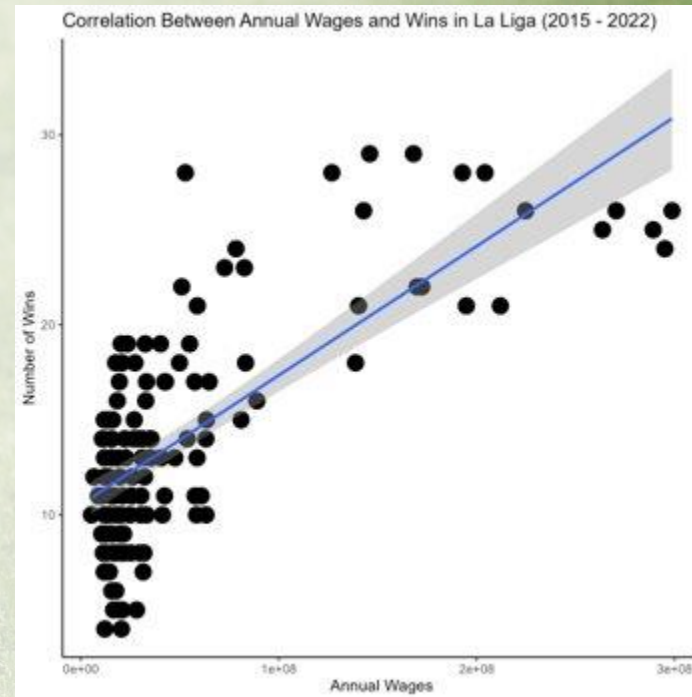


# More money, more chances of winning?

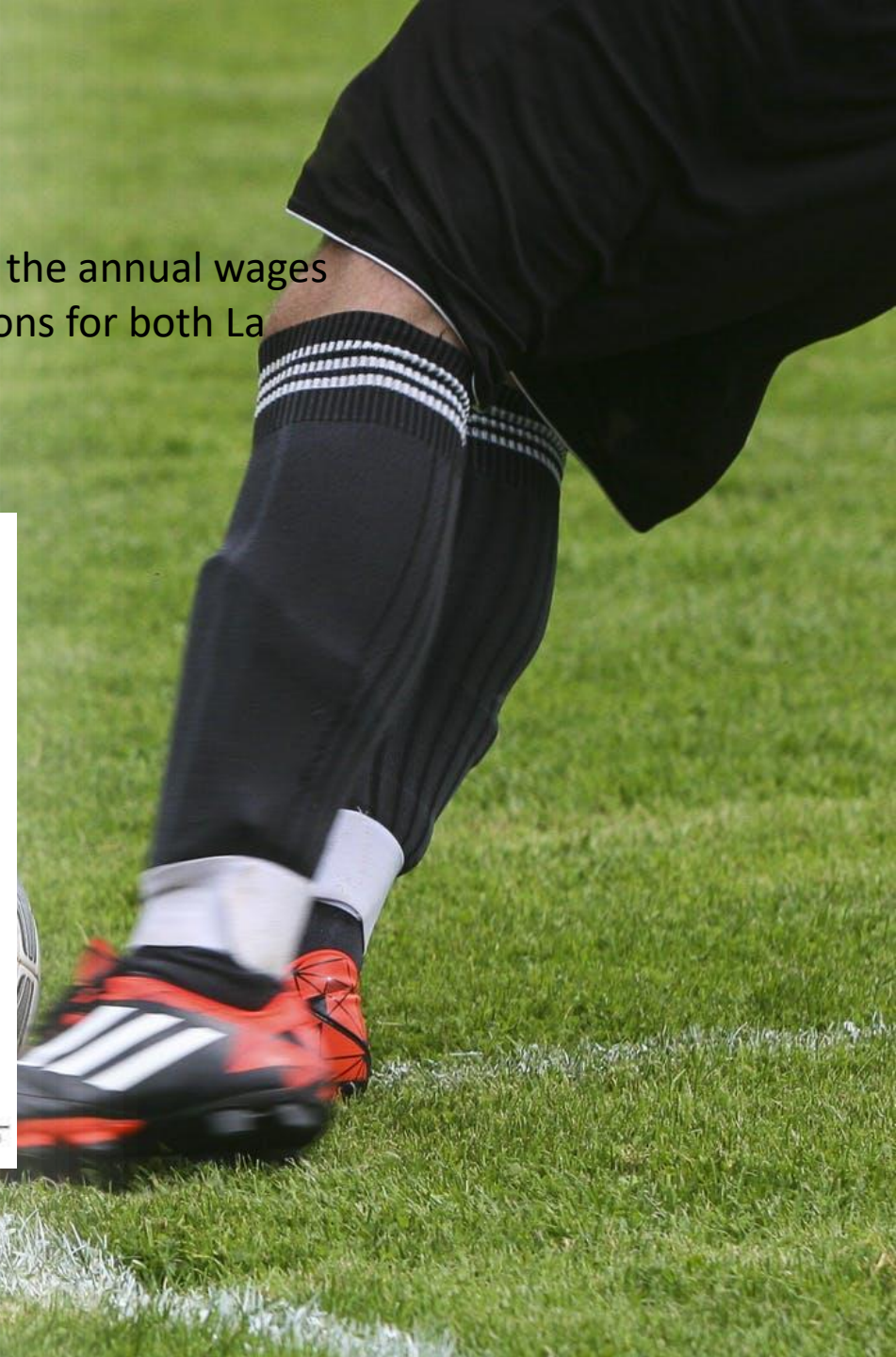
- The number of wins was plotted against the annual wages of all teams across the 2015 to 2022 seasons for both La Liga and Premier League.



Correlation factor = 0.73



Correlation factor = 0.69





# FANTASY POINTS COMPARISON





## HIGHEST FANTASY POINTS PL TEAMS 2015-2022

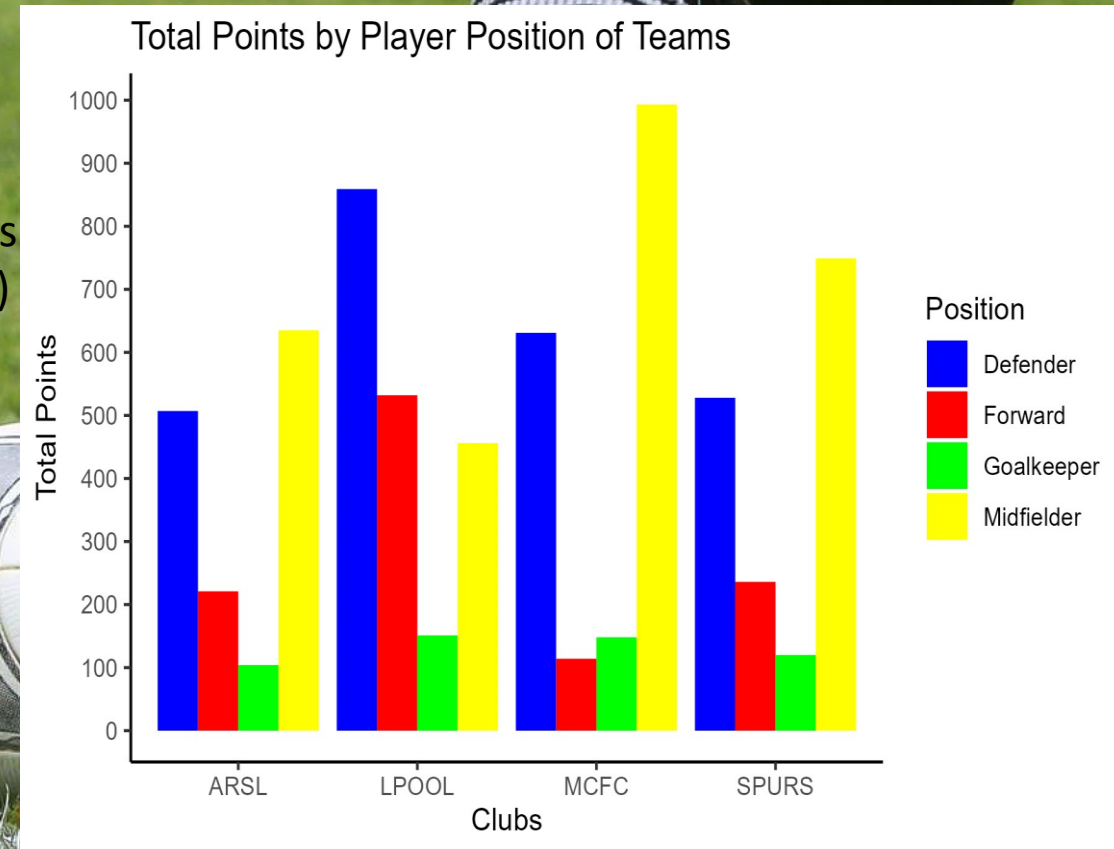
- Analyse top 7 highest scorers of Premier League from 2015-2022
- Total points of each player corresponding to a club is summed up to get the highest scored club in PL in terms of Fantasy points
- It can be inferred that players of MCFC (Manchester city) team has scored high during various years considering the time range 2015-2022
- LPOOL team members are having the highest points and this was achieved in the season 2021-2022





## TOTAL POINTS OF TEAMS BY PLAYER POSITION

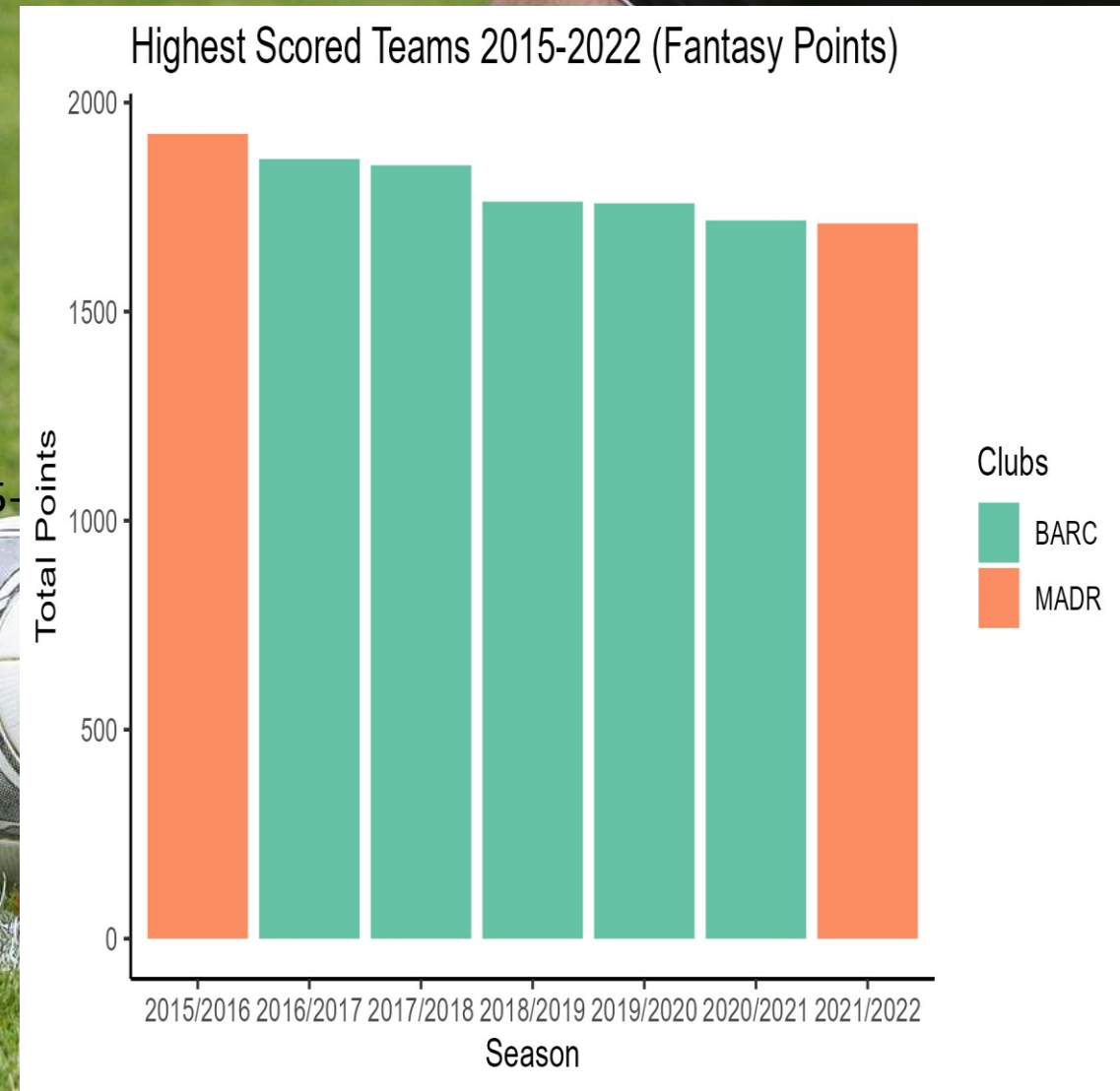
- Compare the highest scorers in terms of position of players
- It can be inferred that Midfielder of each team has scored more goals leading to highest fantasy points (goal scored by midfielder=5 points)
- Defender has contributed more to the LPOOL team than the midfielder(goal scored by goalkeeper or defender=6 points)
- ARSL midfielder scored the highest points but defender and forward was not able to score more points(goal scored by forward=4 points)





## HIGHEST FANTASY POINTS LALIGA TEAMS 2015-2022

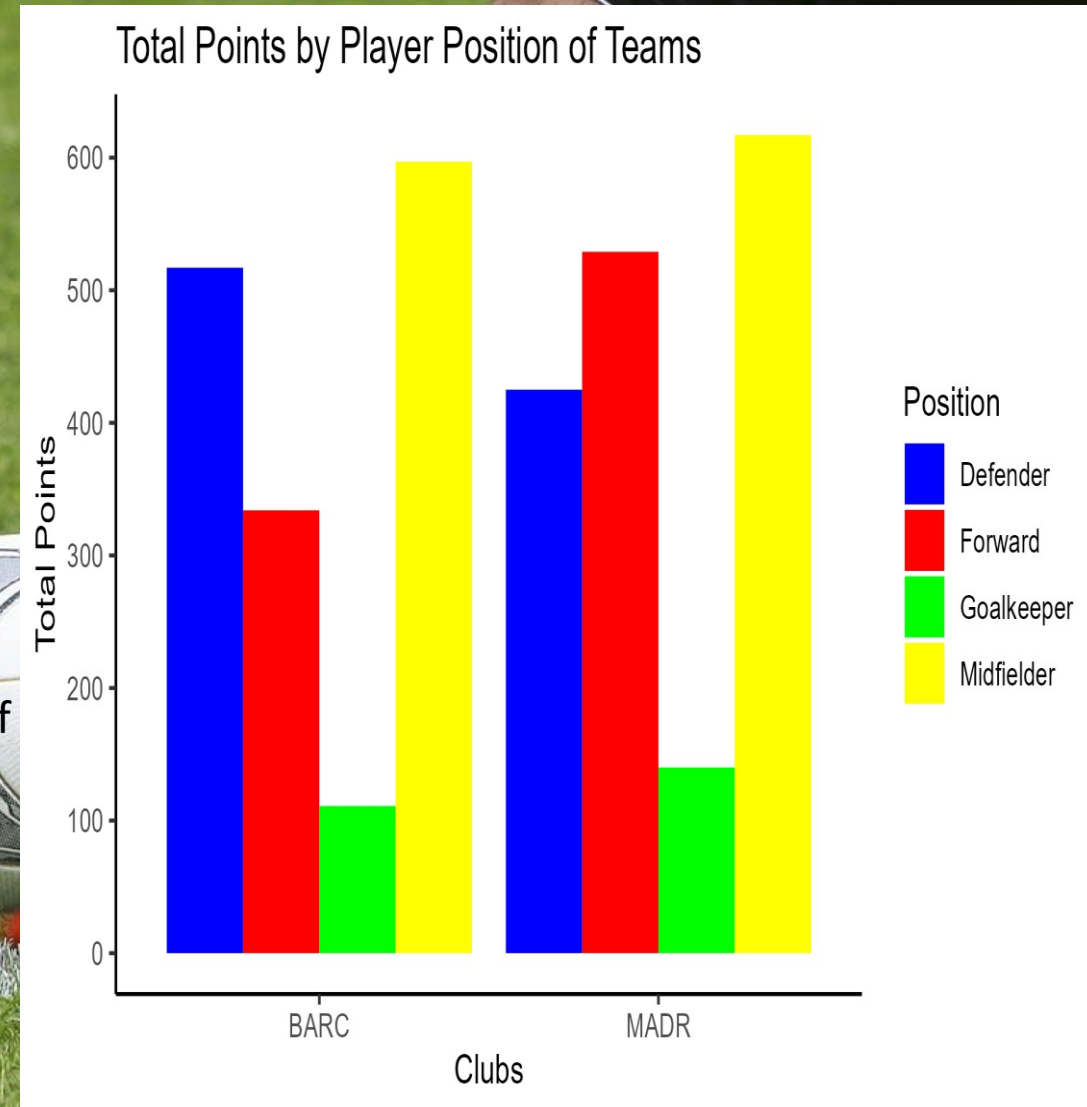
- Analyse top 7 highest scorers of Laliga from 2015-2022
- Total points of each player corresponding to a club is summed up to get the highest scored club in Laliga in terms of Fantasy points
- It can be inferred that players of MADR (Real Madrid) team has scored high almost all seasons years considering the time range 2015-2022
- MADR team members are having the highest points in this plot and this was achieved in the season 2015-2016






## TOTAL POINTS OF TEAMS BY PLAYER POSITION

- Compare the highest scorers in terms of position of players
- It can be inferred that forwarder of each team has scored more goals leading to highest fantasy points (goal scored by midfielder=5 points)
- Midfielders has the second highest points and it is clear that they scored more goals than defender(goal scored by goalkeeper or defender=6 points)
- BARC(Barcelona) and MADR (Real Madrid) is having the same trend of point distribution with respect to the position.





A close-up, low-angle shot of a soccer ball hitting a goal net. The ball is white with black hexagonal panels and is caught in the white mesh of the net. The background is a clear blue sky. The lighting is bright, suggesting a sunny day.

—

HOW MUCH  
DOES EACH  
TEAM PAY FOR  
A POINT?

# DATA FRAME COMBINING ALL THE RELATIONAL DATA

[158]: Laliga\_Final

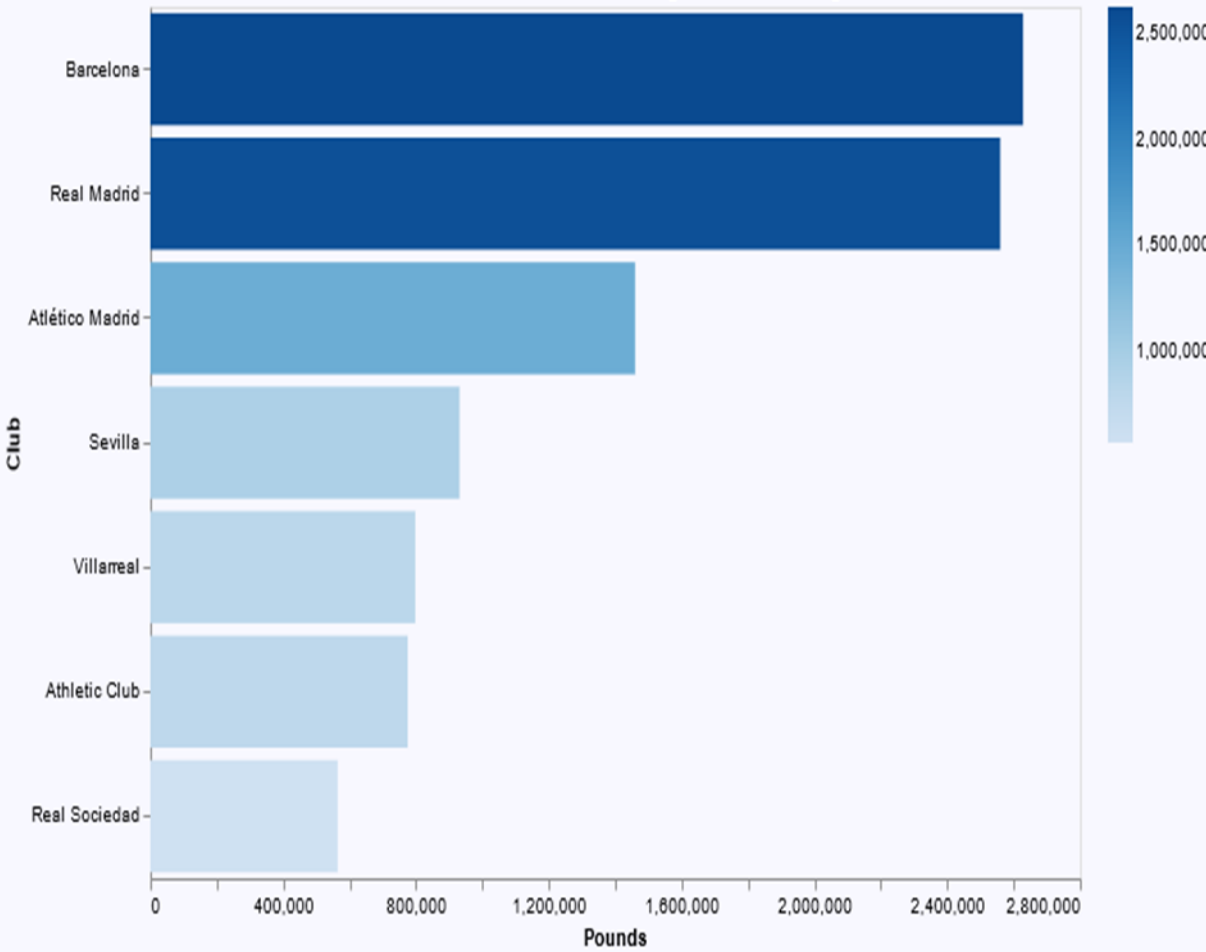
[158]: 7x8 DataFrame

Row	Clubs	Total_league_Goals_For	Total_league_Goals_Against	Total_league_Points	Total_Fantasy_Points	Total_Wages	Pounds_Per_Fantasy_Point	Pounds_Per_League_Point
	String31	Int64	Int64	Int64	Int64	Int64	Float64	Float64
1	Barcelona	656	245	595	14298	1563668007	1.09363e5	2.62801e6
2	Real Madrid	590	249	584	13490	1494577445	1.10792e5	2.55921e6
3	Atlético Madrid	429	191	548	10312	799884995	77568.4	1.45964e6
4	Sevilla	391	301	458	11481	426414202	37140.9	9.31035e5
5	Villarreal	392	300	413	11371	329384533	28967.1	7.97541e5
6	Athletic Club	323	298	373	10554	288967595	27379.9	7.74712e5
7	Real Sociedad	370	329	391	11153	220558729	19775.7	5.64089e5

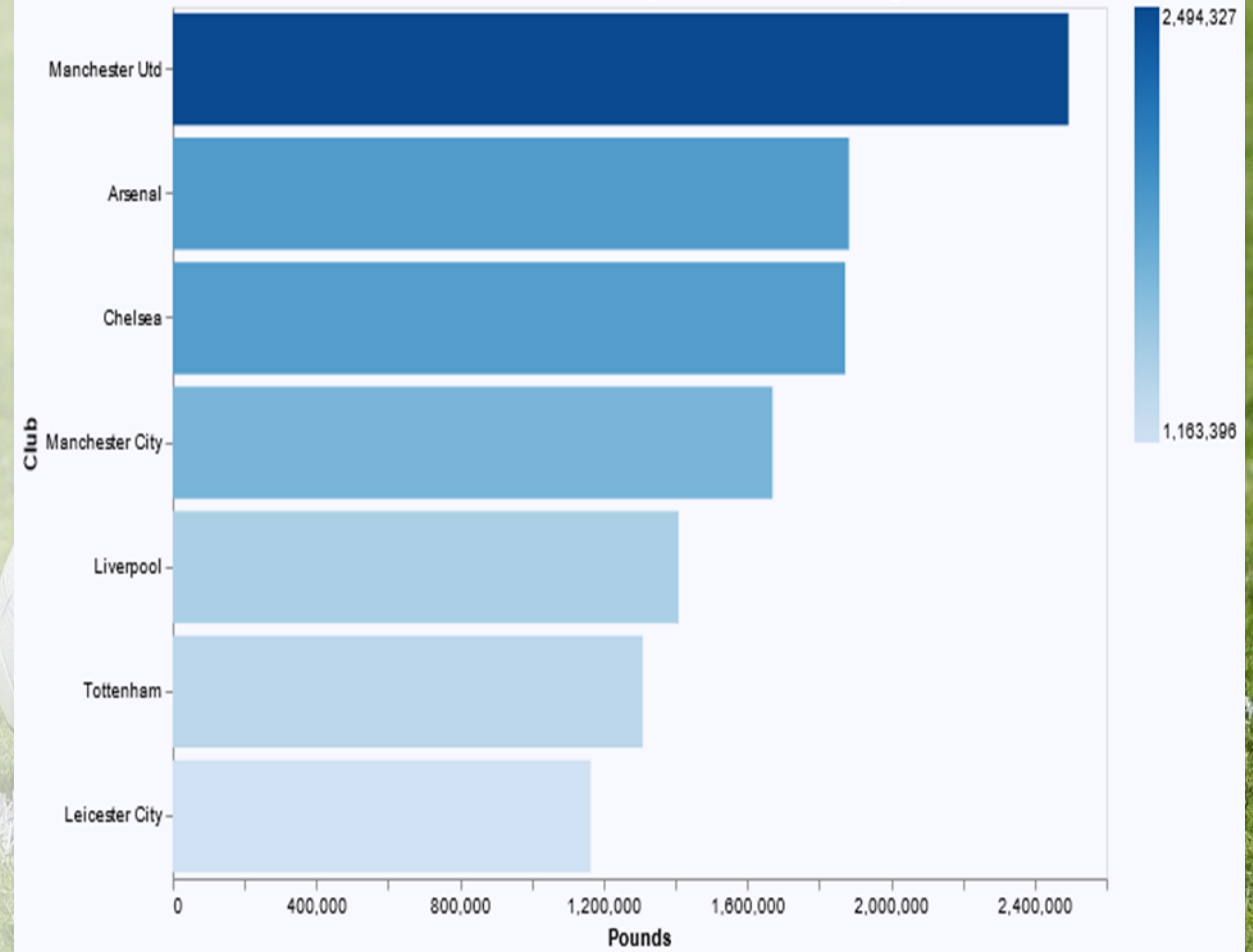


# POUNDS PER ONE LEAGUE POINT

Pounds Incurred Per One League Point in Laliga

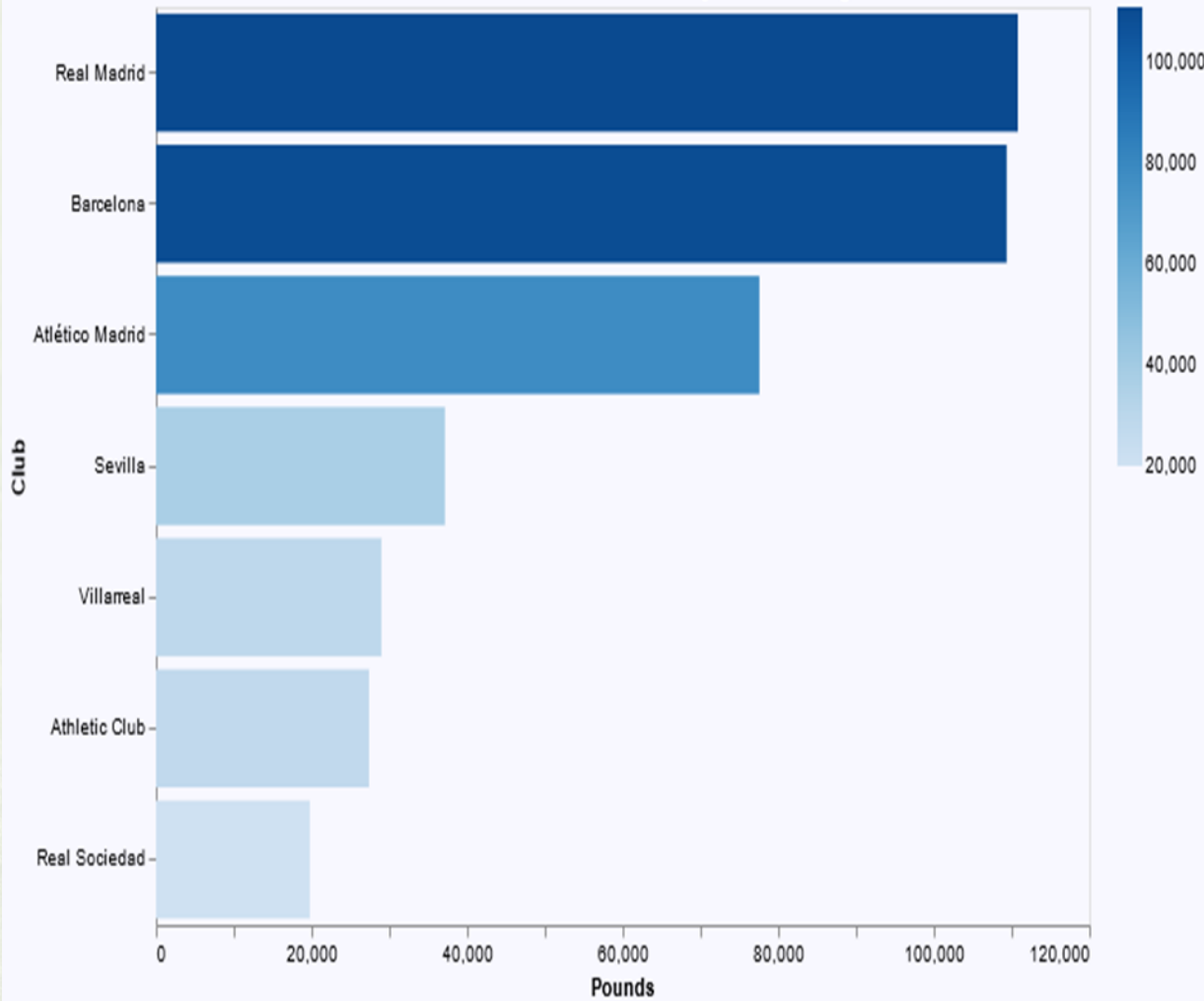


Pounds Incurred Per One League Point in PremierLeague

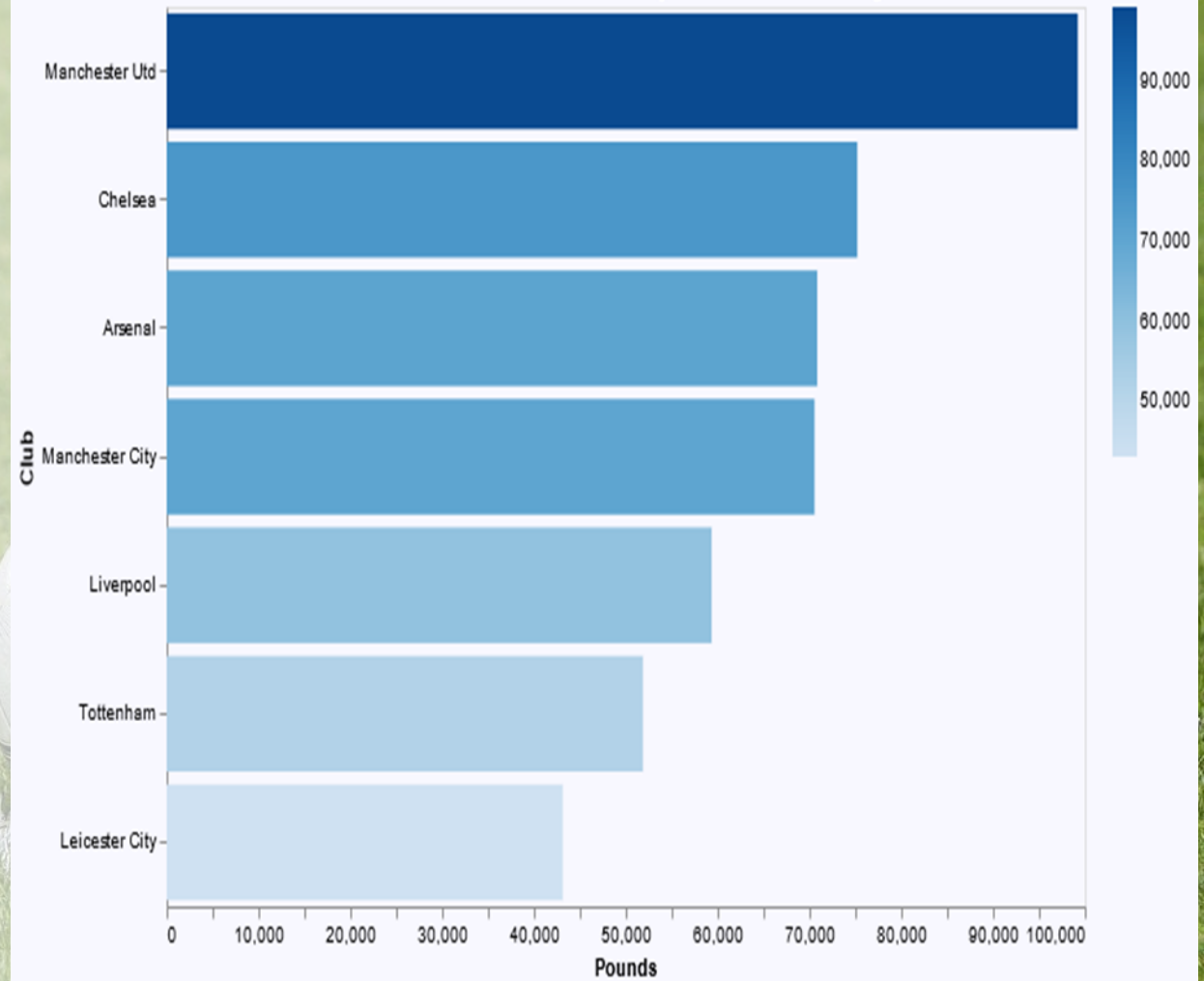


# POUNDS PER ONE FANTASY POINT

Pounds Incurred Per One Fantasy Point in Laliga



Pounds Incurred Per One Fantasy Point in PremierLeague





# Key Findings

- “Correlation does not imply causation”
- Annual wages are correlated to wins, but it does not guarantee a winning season.
- Barcelona spent the most money per league point (La Liga). They have won the league three times in the last 7 years.
- Manchester United spent the most money per league point but has not won the Premier League championship in the last 7 years.
- Higher team budget does not always result to more fantasy points.
- Real Madrid won the league championship twice in the last seven years
- Manchester United spent the most money per league point but has not won the Premier League championship in the last 7 years.

