

Course: Data Science

Group Homework #4

Anna Baghumyan
Maria Davoodian
Hermine Grigoryan
Arusyak Hakobyan
Lilia Mamikonyan
July 19, 2018

Contents

Objective	2
Part #1	3
Total Goals per Each Club	3
Top Number of Yellow Cards for Teams and Players	4
Top Number of Red Cards for Teams	6
Percentage of Goals from Shots for Each Player	7
Percentage of Shots on Goals from Shots	8
Goals and Assits per 90 Minutes	9
Percentage of Goals from Overall Goals of Each Club	10
Distribution of Goals per Each Club	11
Number of Offsides (footballers)	12
Most and Least Red and Yellow Cards	13
Most Fouls Committed (footballers)	14
Most Fouls Suffered (footballers)	15
Most Fouls Committed (clubs)	16
Most Fouls Suffered (clubs)	17
Number of Corner Kicks (clubs)	18
Number of Corner Kicks (players)	19
Number of Corner Kicks (players)	20
Penalty Kicks and Goals (players)	21
Penalty Kicks (clubs)	22
Penalty Kicks (clubs)	23
Passes (clubs)	24
Number of Saves (goalkeeper)	25
Number of Clean Sheets (goalkeeper)	26
Part #2	27
Most Common Words Used by Users (Aquaman)	27
Most Common Words Used by Users (Fantastic Beasts)	29
Frequencies of Words across Two Movies	30
Sentiment Analysis	32
Most Positive Tweets about Aquaman	32
Most Negative Tweets about Aquaman	32
Most Positive Tweets about Fantastic Beasts	32
Most Negative Tweets about Fantastic Beasts	33
Distribution of Tweet Emotions	33
Number of Positive, Negative and Neutral Tweets	34
Interactive Plots	34

Objective

In this group homework we used Web scraping techniques (Web Scraping and APIs) to scrap data from different websites and work on it. The project consists of two parts.

The first part includes web scraping, data cleaning and quantitative analysis of 2016-2017 and 2017-2018 La Liga football statistics. In the analysis, we provide some plots to visualize numbers and get interesting insights from them. The codes for data scraping and cleaning are provided in a separate R script file.

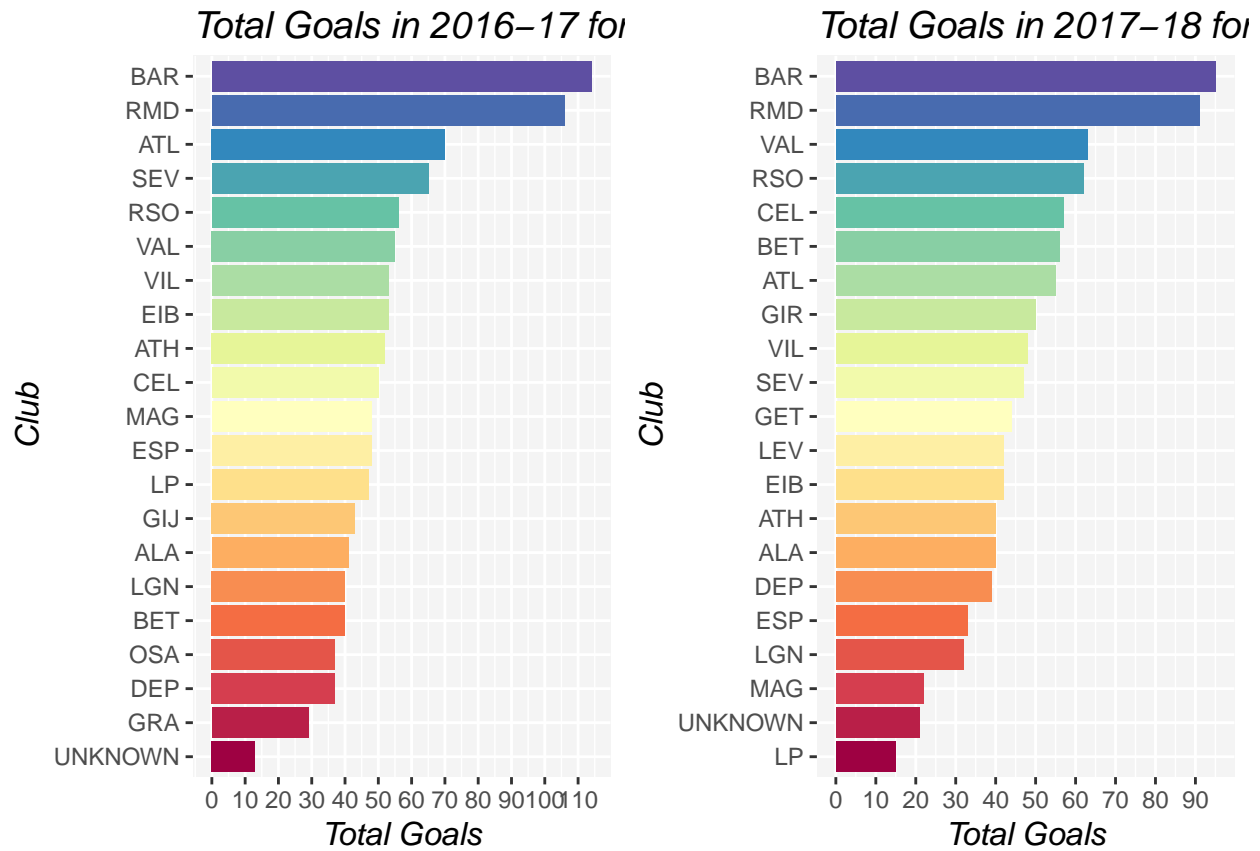
The second part of the project includes text mining. The data was scraped from Twitter. It contains information about reactions of Twitter users about trailers of movies *Aquaman* and *Fantastic Beasts*. Warner Bros. revealed the trailers of these movies on 21 of July, 2018, and this topic became very popular that day. The analysis includes some graphs and wordclouds about common words used by the users. Moreover, it includes sentiment analysis of the tweets.



Part #1

Total Goals per Each Club

We will start our analysis from simple statistics. First of all, we need to find out the number of goals per each club both for seasons 2016-17, 2017-18.

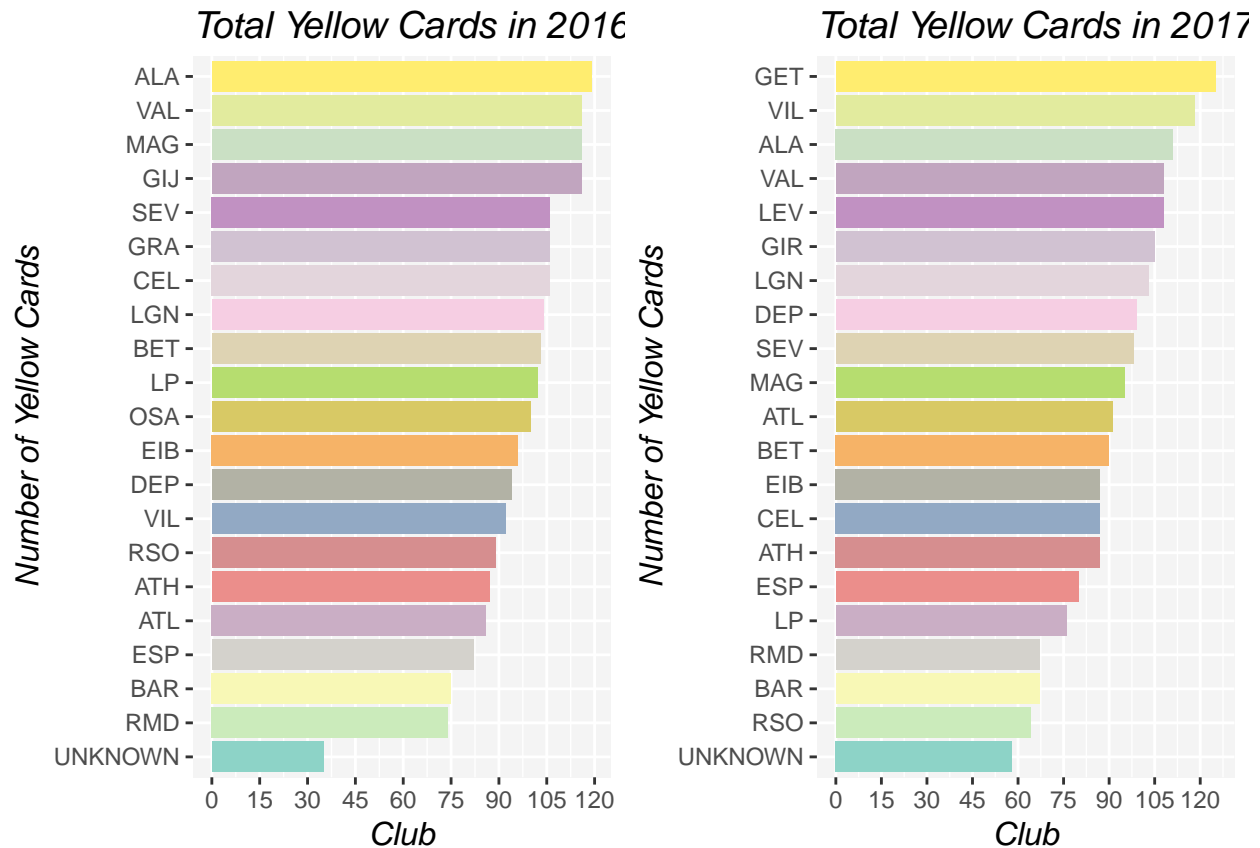


As we see from the barplot, Barcelona has scored the most goals for the both seasons. It is around 110 for the first season, and 95 for the second season.

Barcelona is followed by Real Madrid scoring around 105 and 90 goals for two seasons respectively.

The situation was different with the third team. Atletico was in the third place with 70 goals in 2016-17 season, while Valencia was in the third place in 2017-18 with around 60 goals.

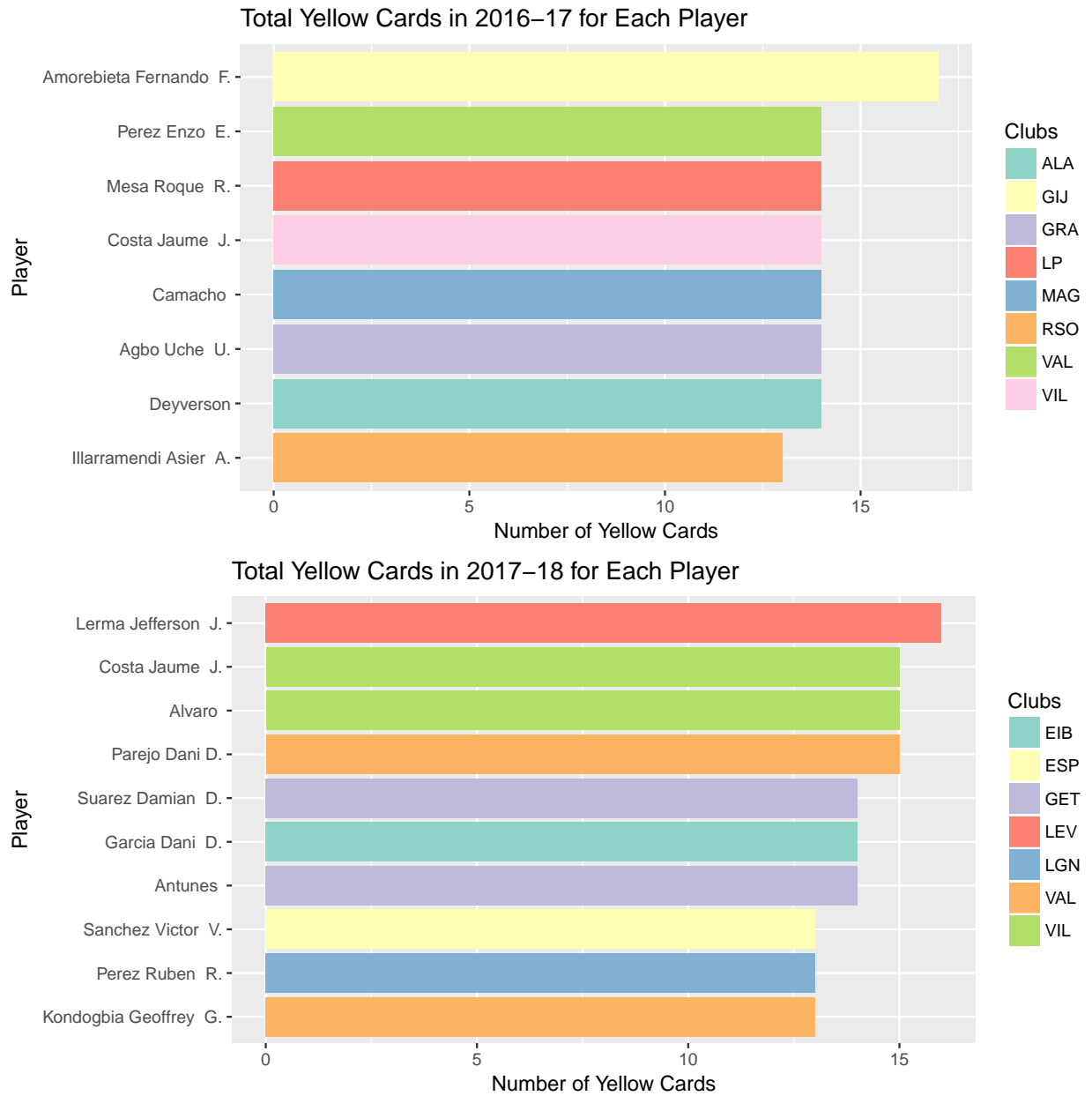
Top Number of Yellow Cards for Teams and Players



It is interesting to know which club's players got the most yellow and red cards for both seasons. As we see from the barplot, players of Alaves got almost 120 yellow cards in the first season. The trend continued for Getafe club in 2017-18. The least yellow cards were for the unknown clubs :)

- **A small note:** for every 5 yellow cards the same player gets, he misses the next match.

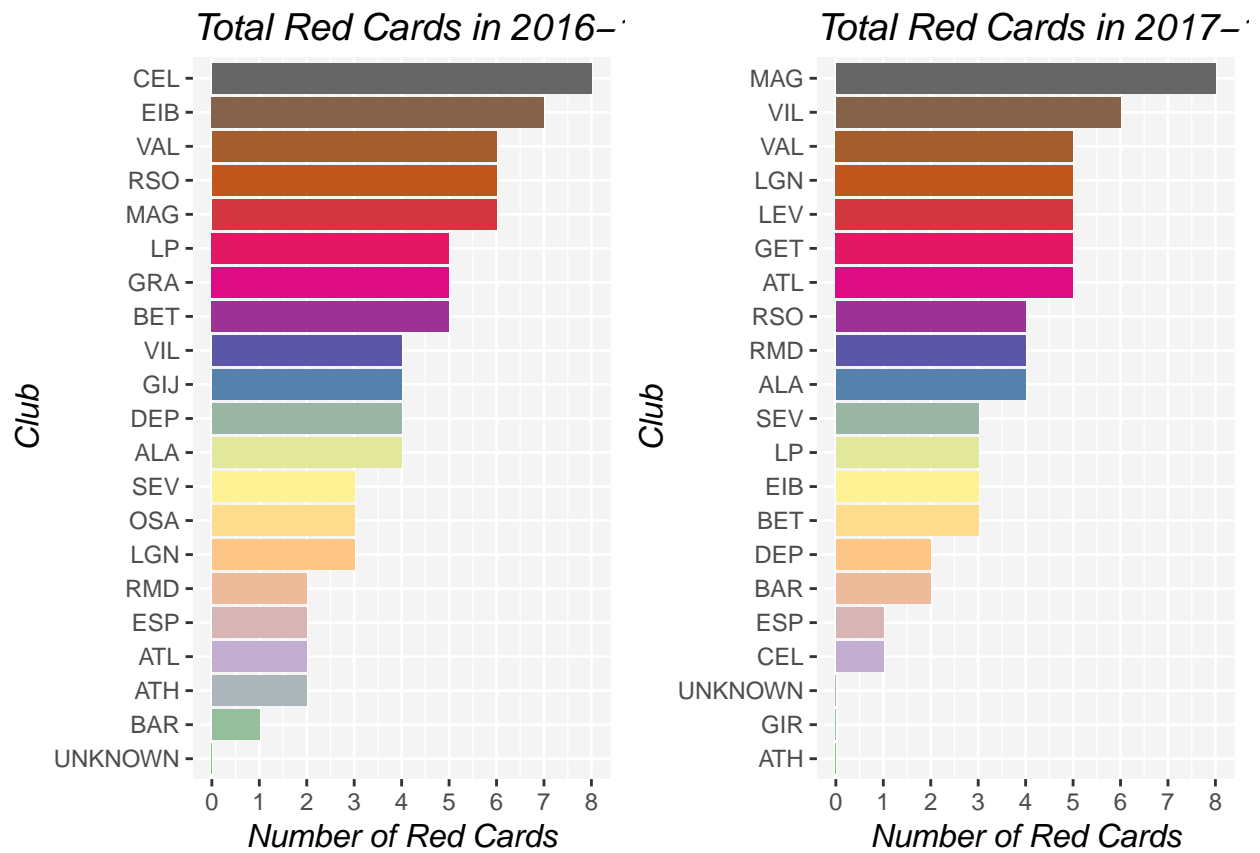
Now, let's see which players got the most yellow cards.



For season 2016-17 daredevils were Fernando Amorebieta from Gijon, and Jefferson Lerma from Levante for the next season.

Top Number of Red Cards for Teams

Now, let's see how things are going with the red cards.



On the top of the list for the first season was club Celta (8 cards), followed by Eibar with 7 red cards.

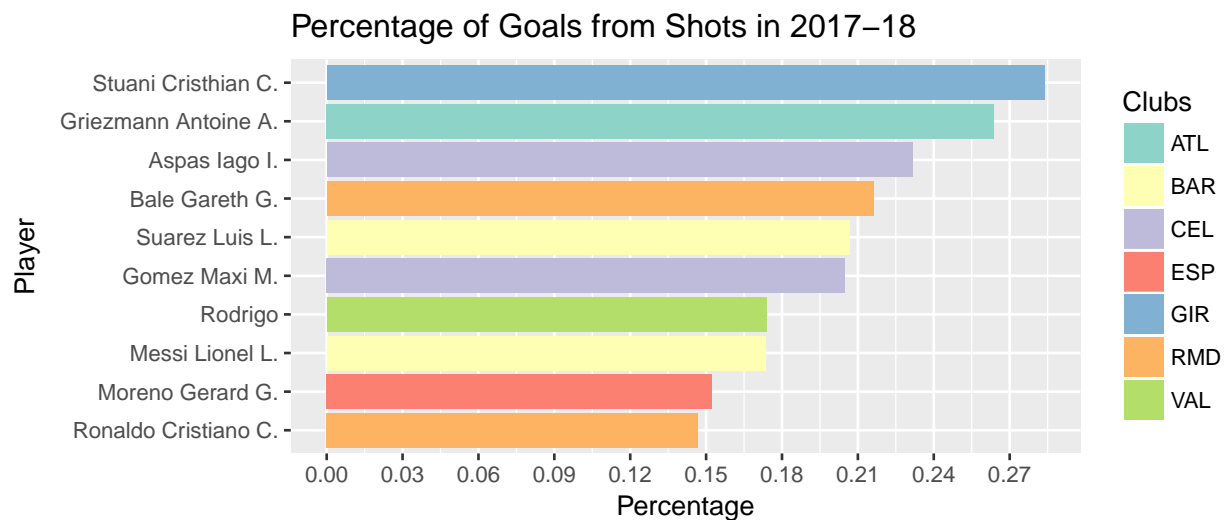
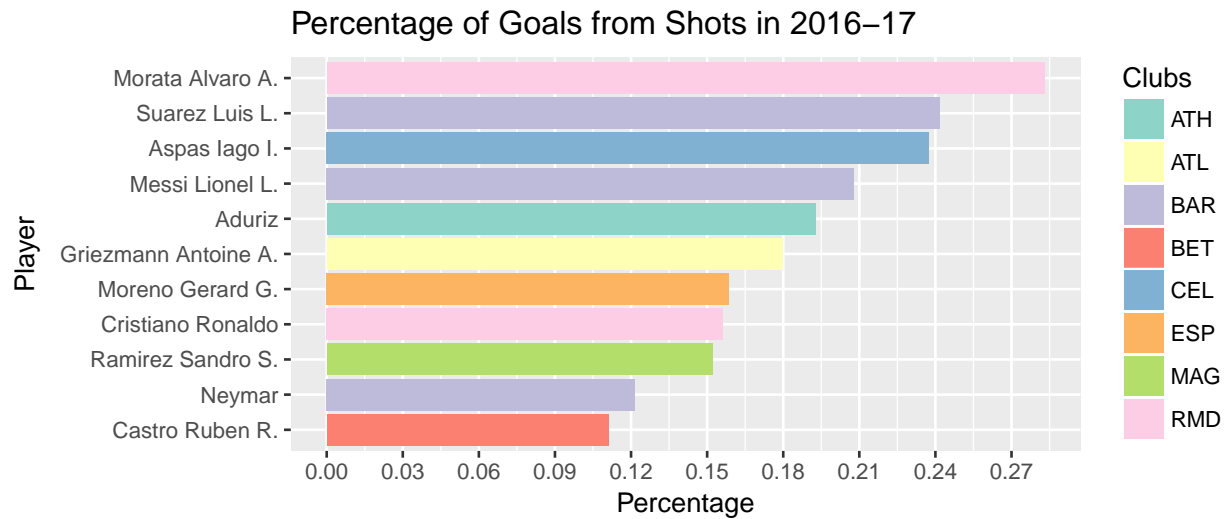
The number of the red cards for the next season was also 8. Now, on the top was Malaga.

We know that for each red card, the player will get suspended for at least one game.

Percentage of Goals from Shots for Each Player

Now, we want to understand which percentage of shots of each player were actually goals. We have limited the number of players to those who have scored at least 12 goals in 2016-2017 season and at least 15 goals in 2017-2018 season. This will eliminate most defenders, goalkeepers and some mid-fielders as they usually do not score many goals.

Secondly, we actually get that percentage by dividing total number of goals for a particular player by the shots in has made.

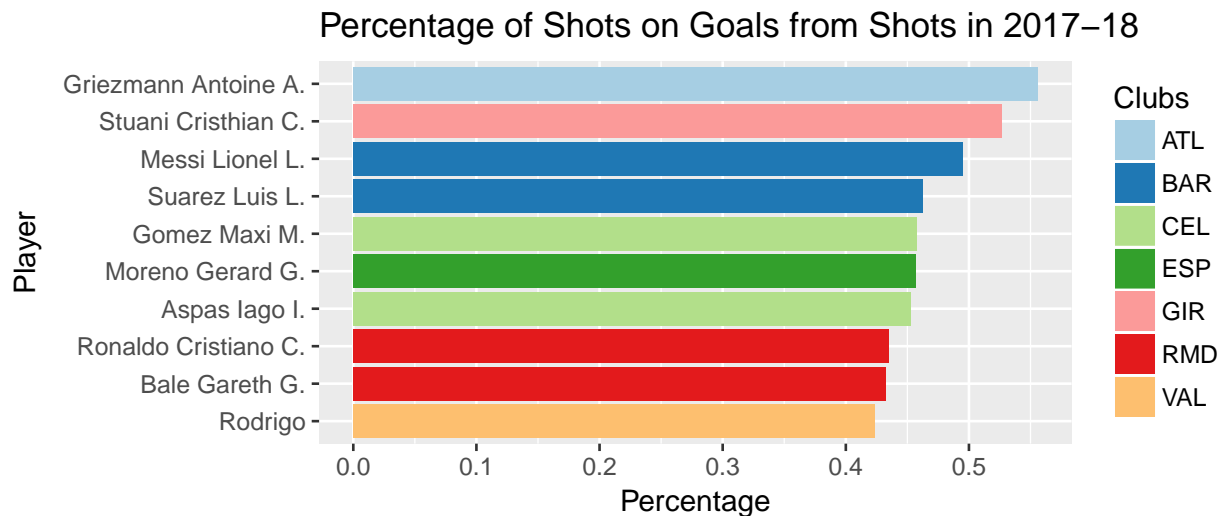
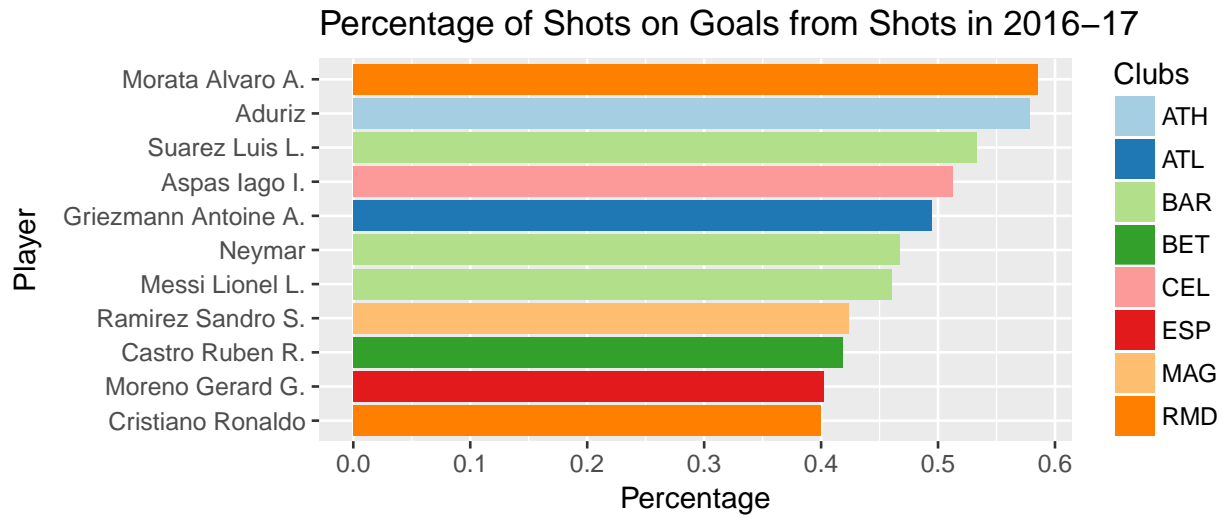


From the plots, we see that in 2016-17 Alvaro Morata's shots from Real Madrid were the most promising ones which were actually goals. Being such a deadly finisher, Alvaro Morata was not happy at Real Madrid playing as a substitute most games and probably that is why he decided to leave his club. Luis Suarez from Barcelona is next on this list.

The things are different for the second season. Here, Cristhian Stuani from Girona, who surprised everyone with their quality football, is on the top. Being in a team that was just promoted to La Liga from the second division, Stuani's 21 goals with a high accuracy percentage were crucial in Girona claiming the 10th spot on the table in 2017-2018. In second place comes Antonie Griezmann from Atletico Madrid, helping his team to secure a spot in the Champions League.

Percentage of Shots on Goals from Shots

Forwards' performance can be also measured by how accurate they are with their shots in general. Every manager expects their strikers to at least hit the target when taking a shot. For this, we found out the percentage of shots on goal (SOG) from the total number of shots (S) taken. For finding the percentage of shots on goals from the total shots for each player, we also filtered the data by leaving only players with more than 12 and 15 goals for two seasons respectively. Then, we divided SOG on S to get percentages.



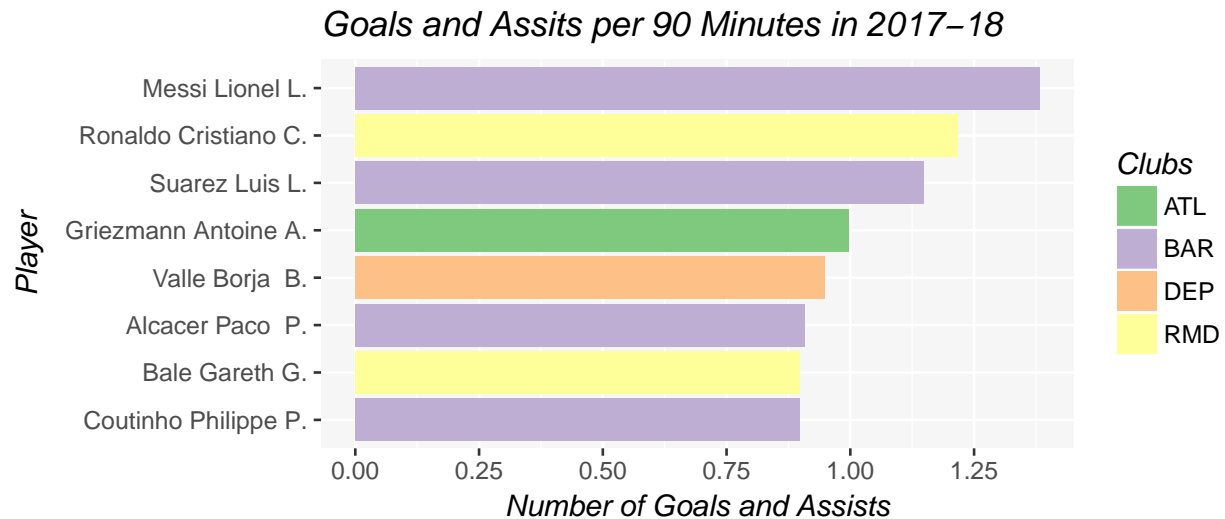
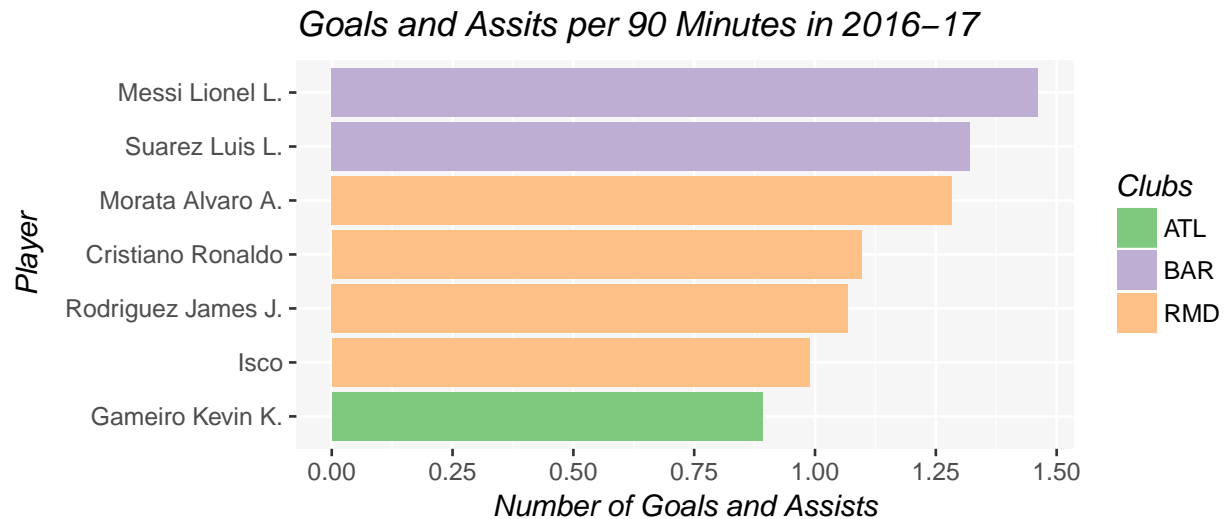
Again, we can see familiar names on the top of this list for the two seasons. Alvaro Morata again was very efficient with the shots he took in 2016-2017. Aritz Aduriz, being second in the same year, did not impress as well in the second year, though. Probably, due to this and the fact that their manager left, his team struggled in 2017-2018.

Antonie Griezmann and Christian Stuani again are leading the top charts in 2017-2018 by their accuracy of shots on goals. We could claim that these two players were the most efficient strikers in the second season in La Liga. Whenever taking a shot the players were most likely to at least hit the target. And when they hit the target, they were the most likely of the all players to score a goal.

Goals and Assits per 90 Minutes

Goal and assists are what make football fans excited and inspired to watch and play themselves. It is no secret that the biggest stars in football are either creators or scorer of goals and even both. This analysis includes those players who have scored and assisted the most per 90 minutes they played.

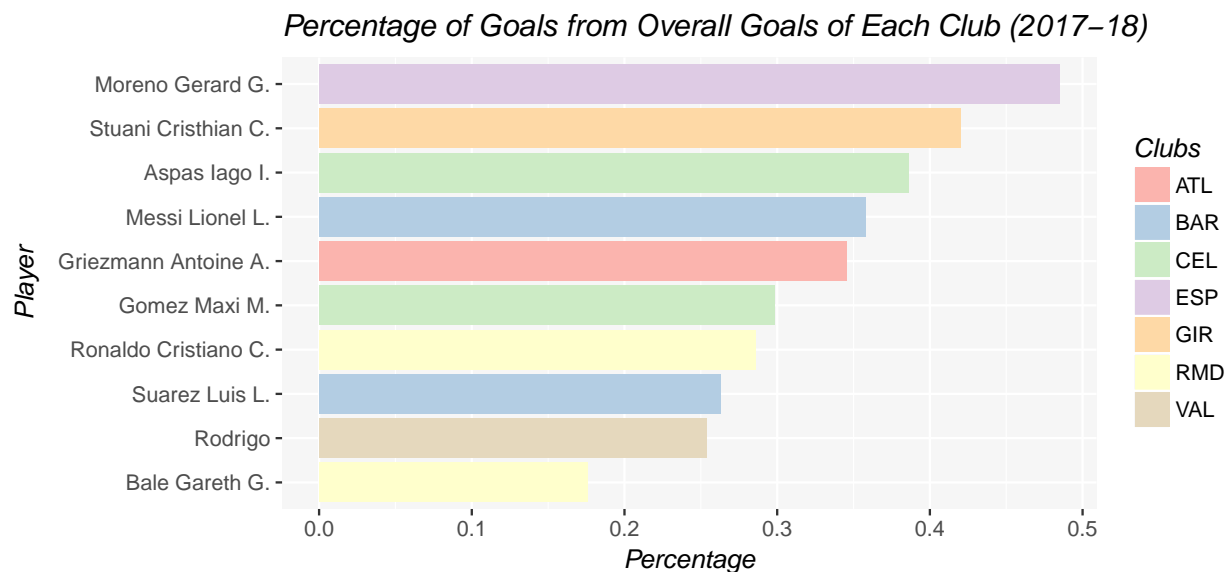
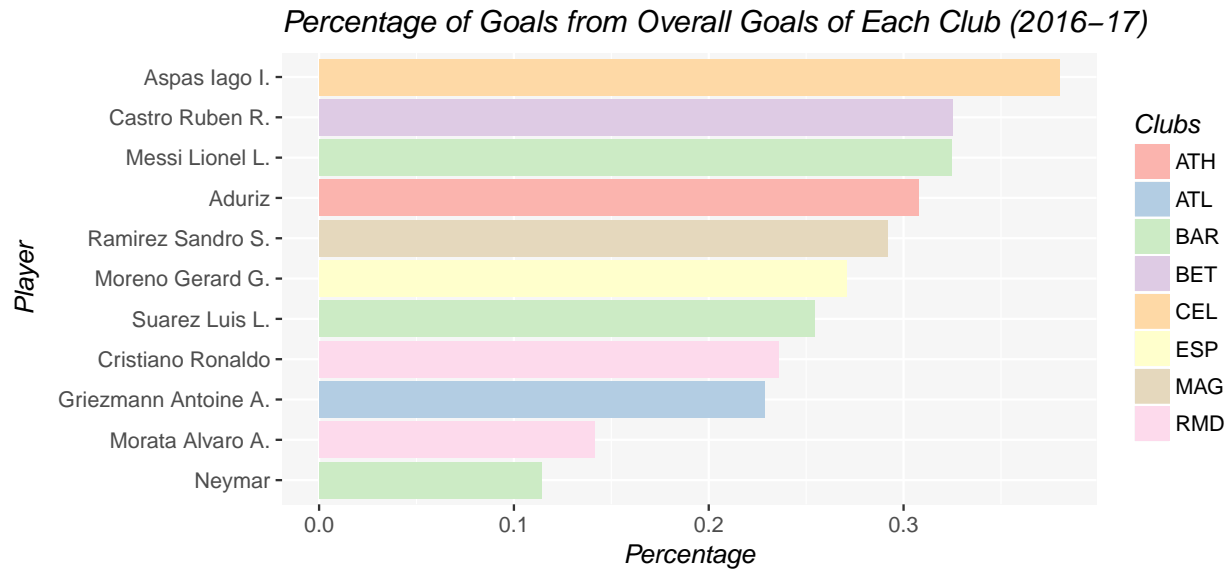
We calculated the ratio per minutes and not games they played to not discredit those who did not start the games and had little time to play but still contributed by scoring or assisting. On top of the charts is Lionel Messi for both seasons. It is interesting to see that in 2016-2017 Real Madrid had 4 players in the top 7 and they won the league that year. But last season they had only two players there and finished third.



Note: this is calculated taking into account those players who played more than 15 games per season.

Percentage of Goals from Overall Goals of Each Club

This analysis represents the percentage of goals scored by a player from the total goals of the club. The ratios were derived using stats for players who scored more than 12 goals in 2016-2017 and more than 15 in 2017-2018. These limits were set to distinguish clubs that were somewhat successful in that season. An exception from this is Ruben Castro in 2016-2017 where although he scored about 35% of the total goals for Betis, the club finished 15th.

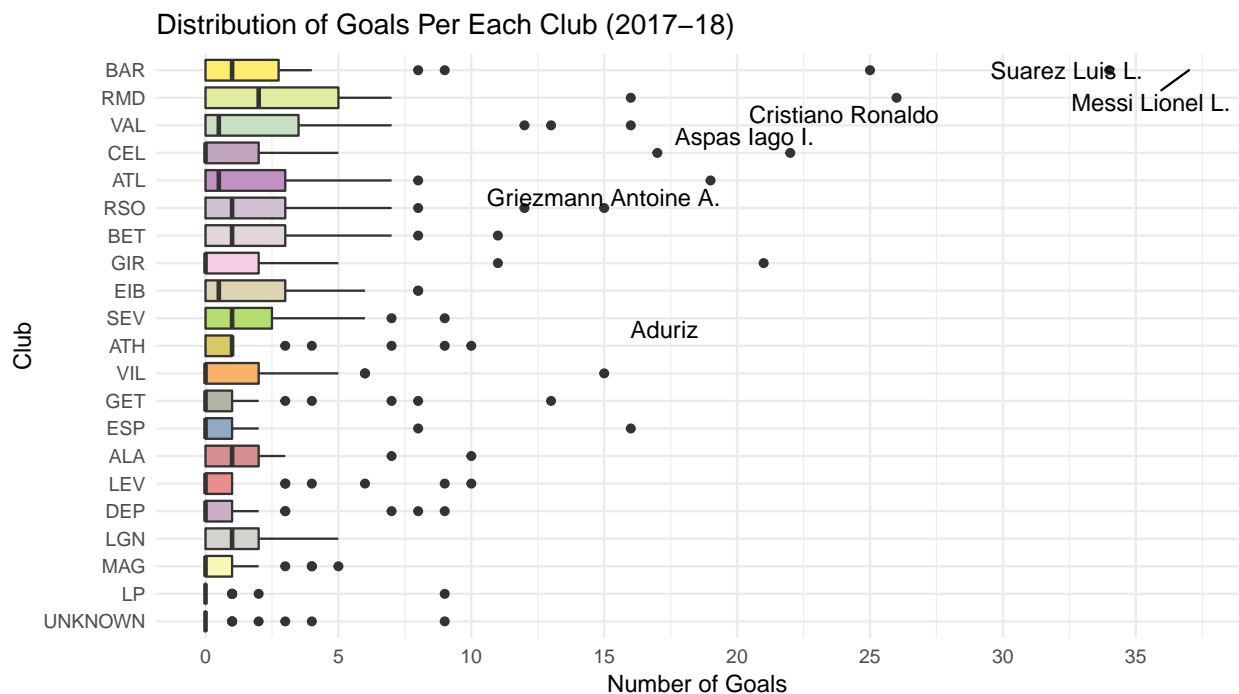
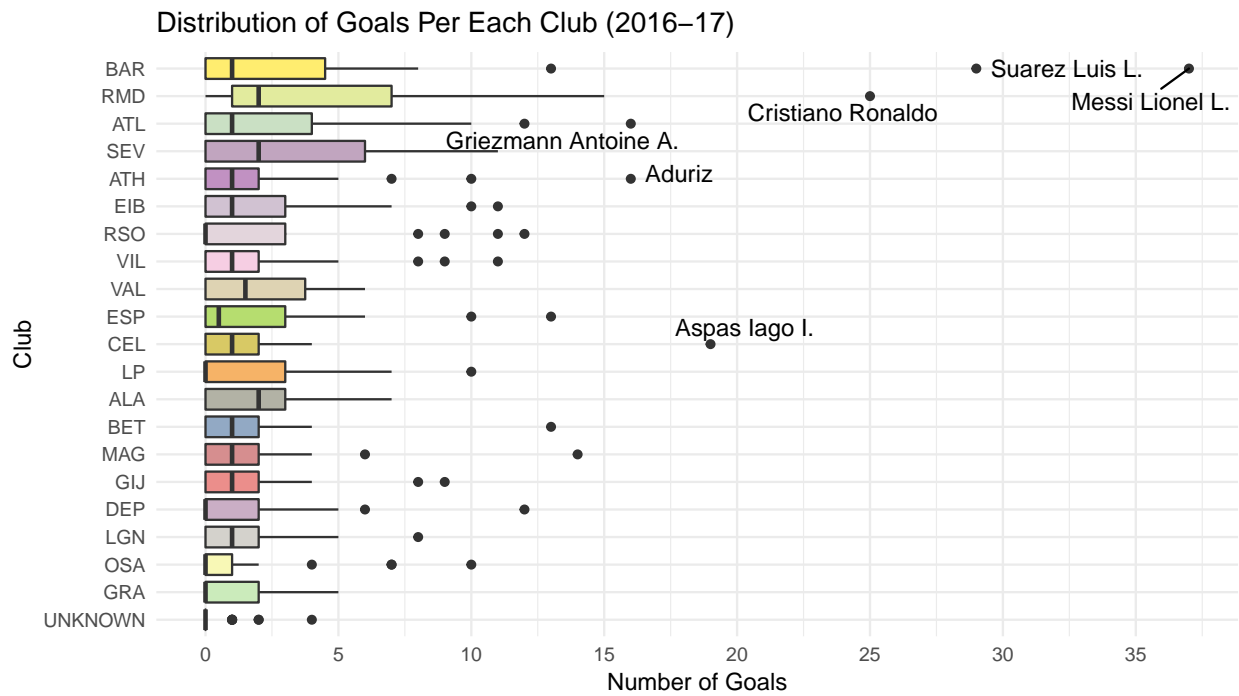


This analysis underlines the importance of the player to the club. Although it also might be a weakness for the club to depend so much on one player, this is generally the case for most teams. Iago Aspas was the person with most goals for Celta Vigo in 2016-2017 season. His 22 goals, almost 42% of total goals scored by Celta, underline his importance to the club.

Gerard Moreno of Espanyol scored 16 goals, or almost 45% of the total Espanyol goals. Although the club have scored only 36 goals for the whole season, which explains their decline in 2017-2018 season, Espanyol will think twice before selling the player if an offer arrives.

Distribution of Goals per Each Club

The following boxplots represent the distribution of total goals per each club.

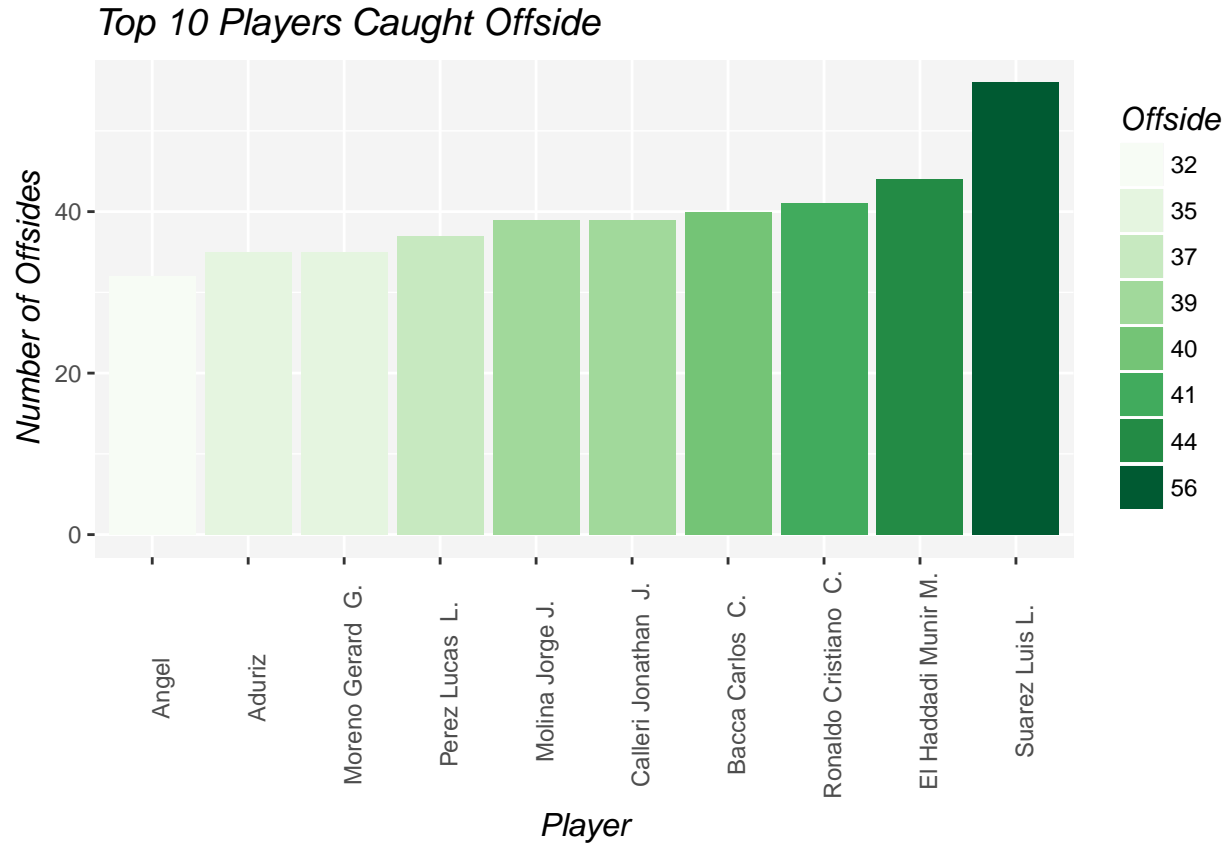


We see that the median value of goals for each player in each club is around 2 for both seasons. However, there are many outliers. For 2016-17 season, the most far outliers are from Barcelona, and they are Messi with 37 goals, followed by Suarez (29 goals). Real Madrid's outlier is Cristiano Ronaldo with 25 goals.

For 2017-18 season, Barcelona's and Real Madrid's many players are also outliers. They are the same players, with the exception that now Ronaldo is on the second place.

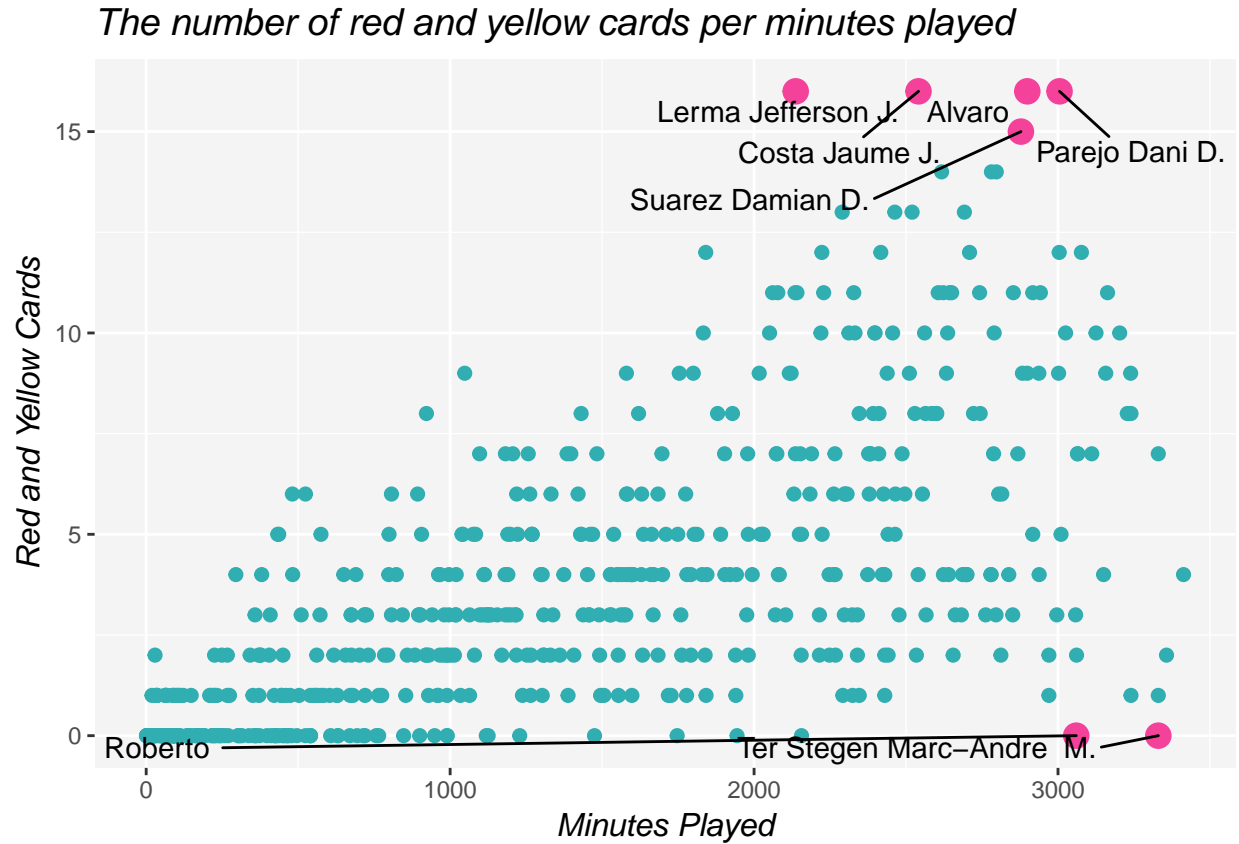
Now, we will use different sources of data (2017-2018 season) not only about the goals by players, but we will also look at the discipline, control and goalkeeping statistics.

Number of Offsides (footballers)



The plot demonstrates the Top 10 Footballers who have been recorded to be in an offside position throughout the 2017-2018 La Liga season. More than 50 offsides were recorded by Luis Suarez making him the footballer who has been caught offside the most.

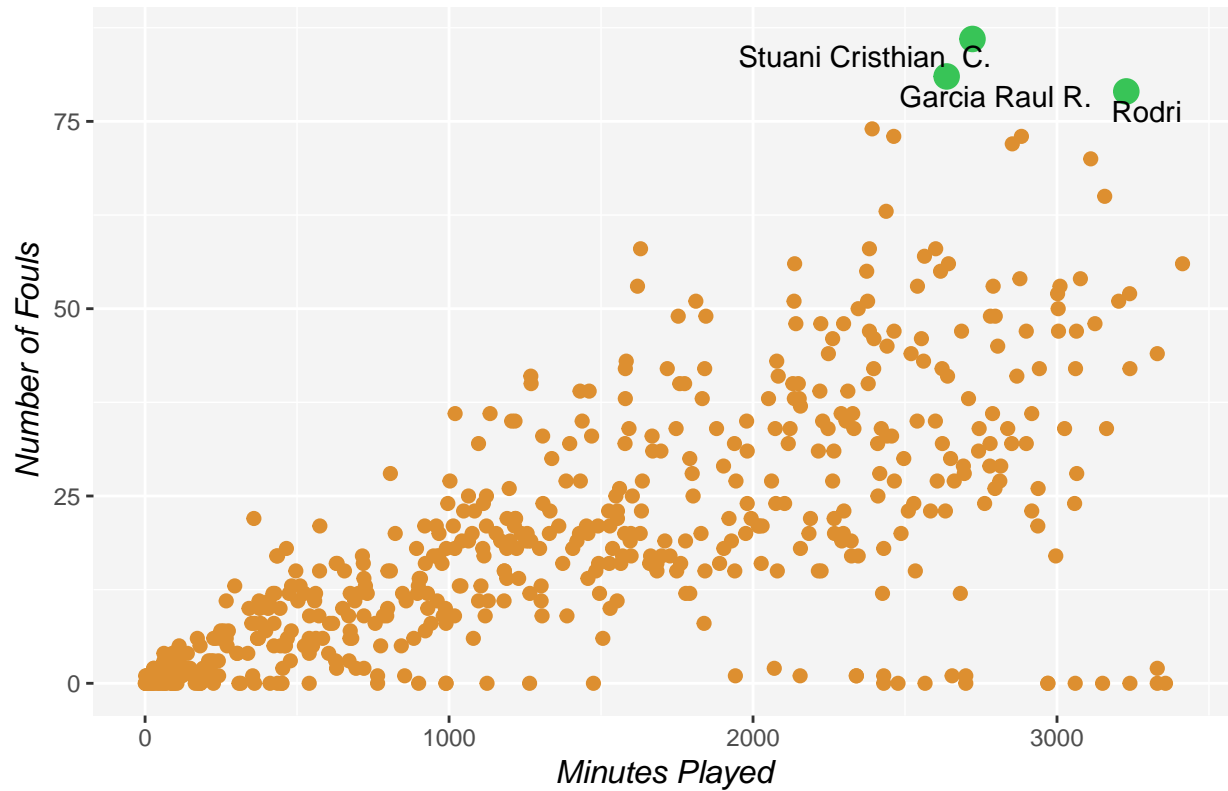
Most and Least Red and Yellow Cards



The plot indicates the number of both red and yellow cards recorded per minutes played (overall throughout La Liga). The players at the bottom of the plot are the best disciplined while moving upwards the discipline gradually decreases. Jefferson Lerma, Jaume Costa, Dani Parejo, Alvaro and Damian Suarez are recorded as being the worst disciplined players while Roberto and Marc-Andre Ter Stegen are the best ones.

Most Fouls Committed (footballers)

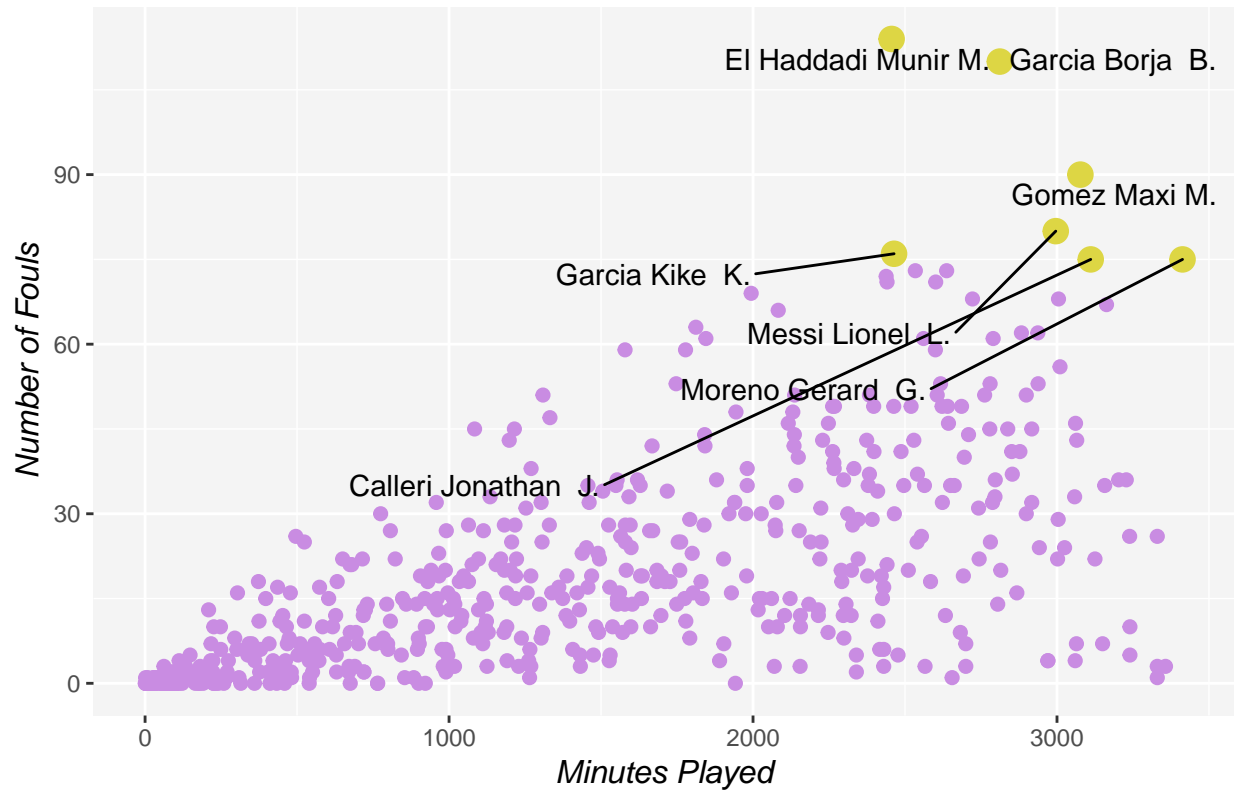
The number of fouls committed per minutes played



Similar to the plot above, players at the bottom of the plot have committed the least number of fouls, while Cristhian Stuani, Raul Garcia and Rodri have committed the most overall fouls throughout the 2017-2018 season.

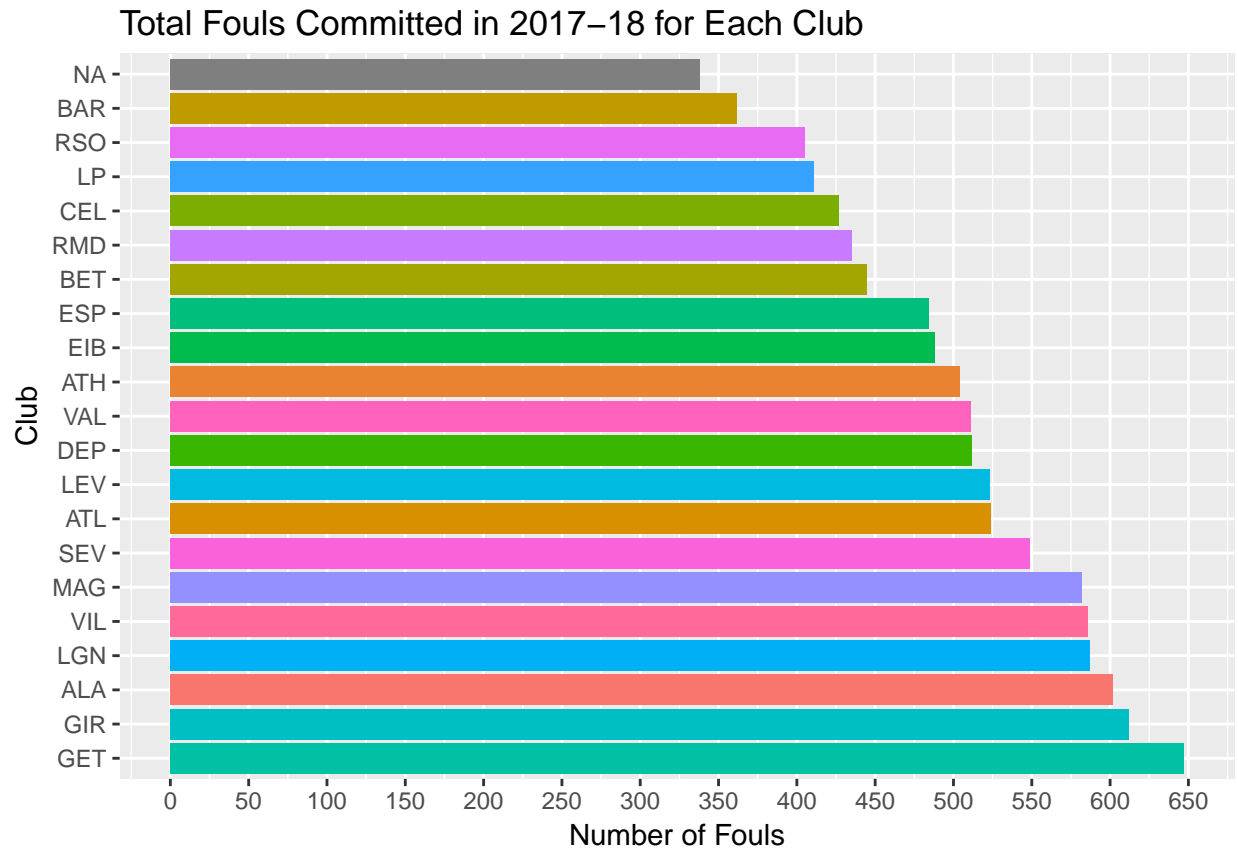
Most Fouls Suffered (footballers)

The number of fouls suffered per minutes played



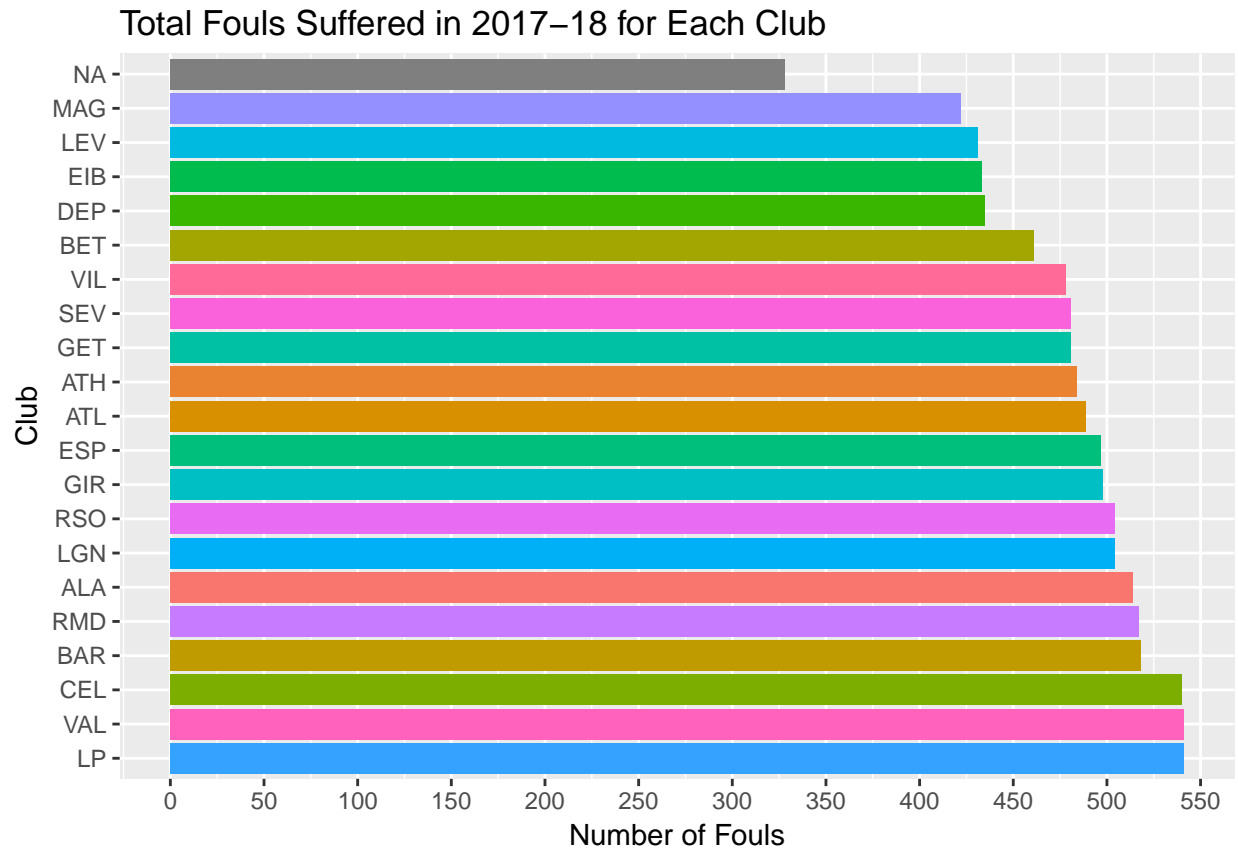
This plot clearly demonstrates the players against whom the most number of fouls have been committed. As we can see the most fouls have been committed against Munir El Haddadi, while many expected to see Lionel Messi at the top, since he is recognized as one of the best football players and stopping his actions are considered to be the most difficult. However, a possible reason for Messi not being in a higher position may be that Messi usually prefers to stay in his feet and continue playing instead of going down easily.

Most Fouls Committed (clubs)



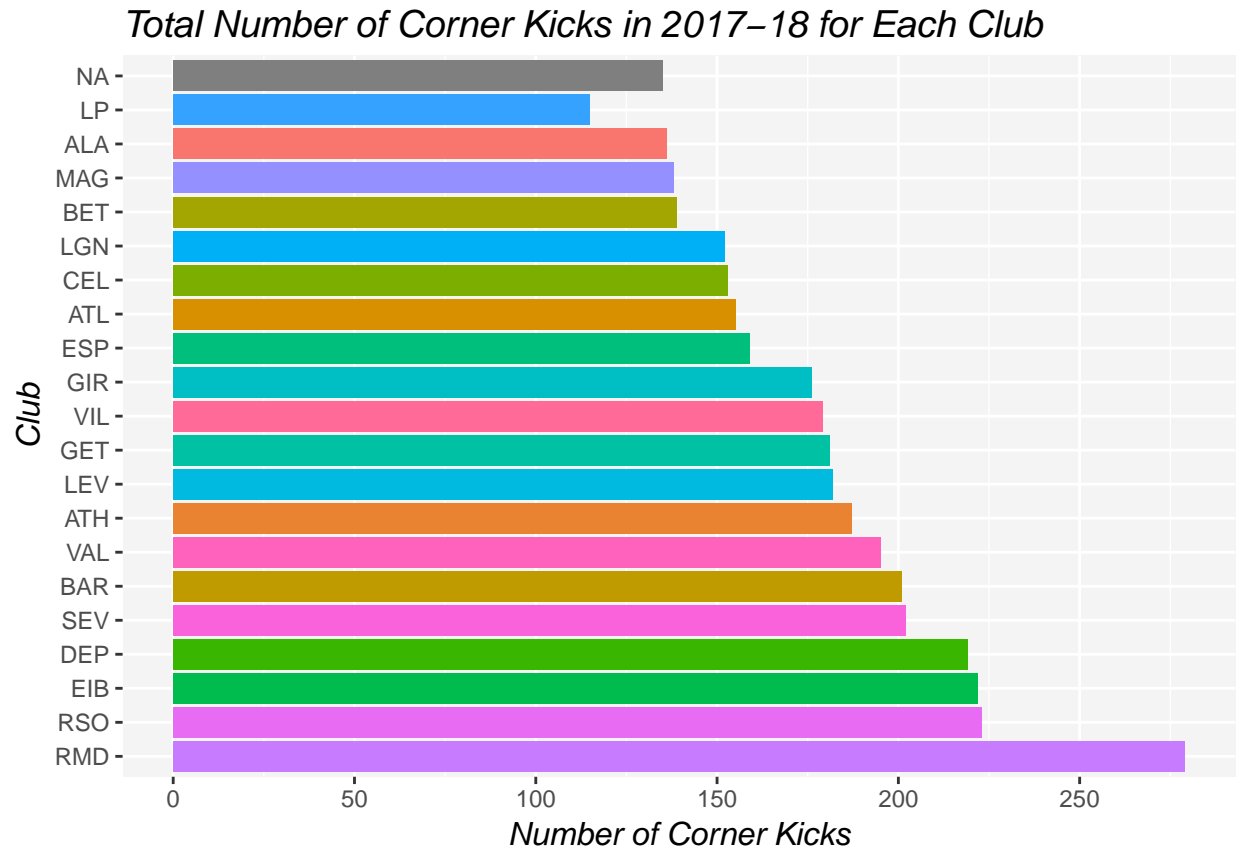
The plot above indicates the number of fouls committed by each club throughout the season. The club with the most number of committed fouls is Getafe and the club with the least - Barcelona.

Most Fouls Suffered (clubs)



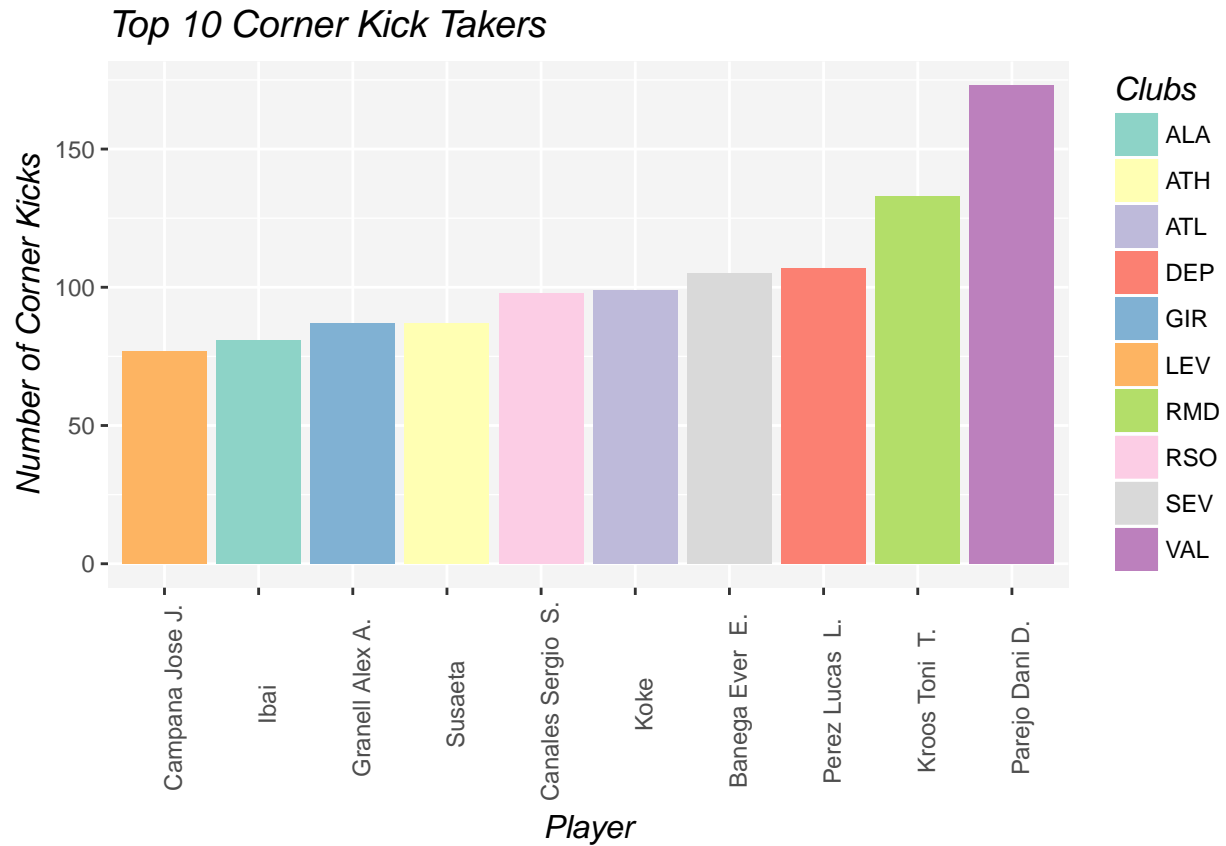
The plot above indicates the clubs against which the most number of fouls have been committed throughout the season. The club with the most number of suffered fouls is Valencia and Malaga is the least.

Number of Corner Kicks (clubs)



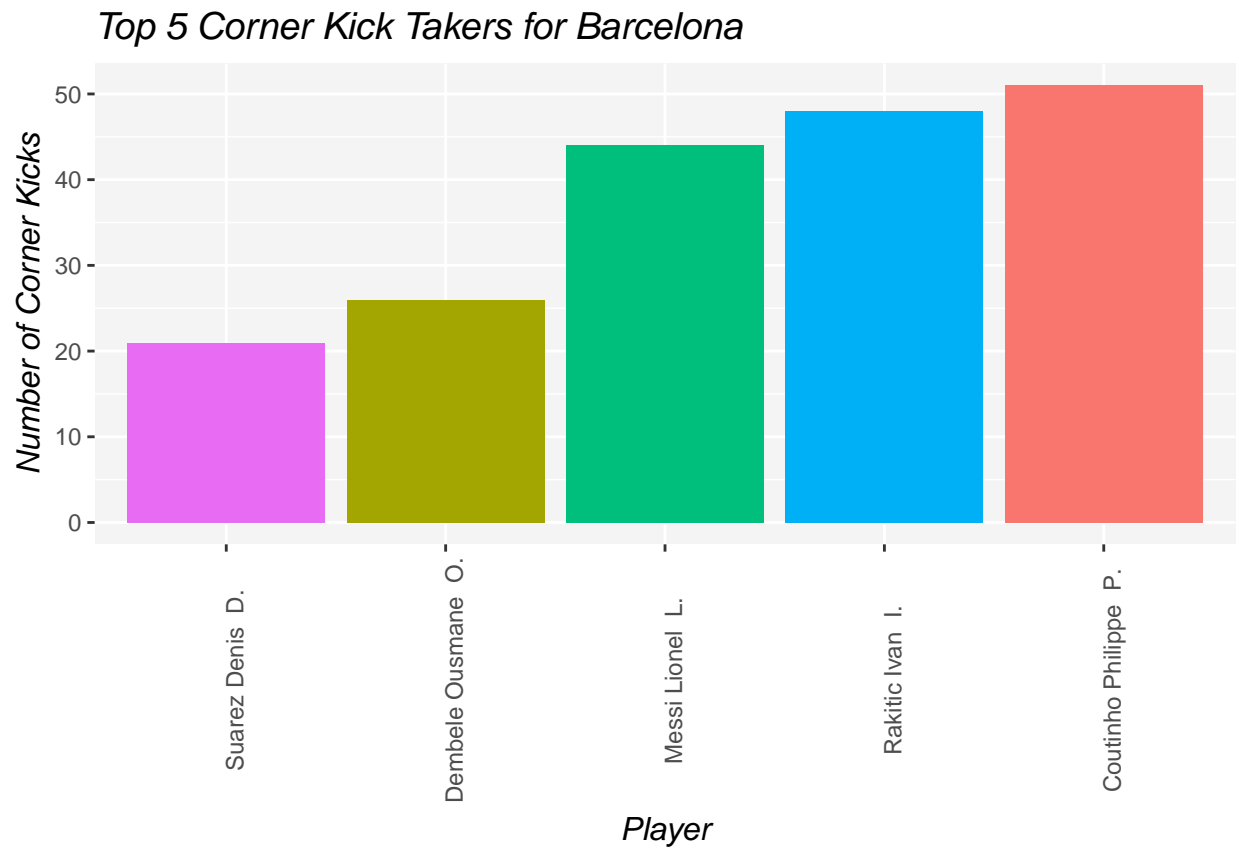
The plot demonstrates the number of corner kicks recorded by each club. Real Madrid has recorded the most number of corner kicks at the 2017-2018 La Liga, while Las Palmas has recorded the least number of corner kicks.

Number of Corner Kicks (players)



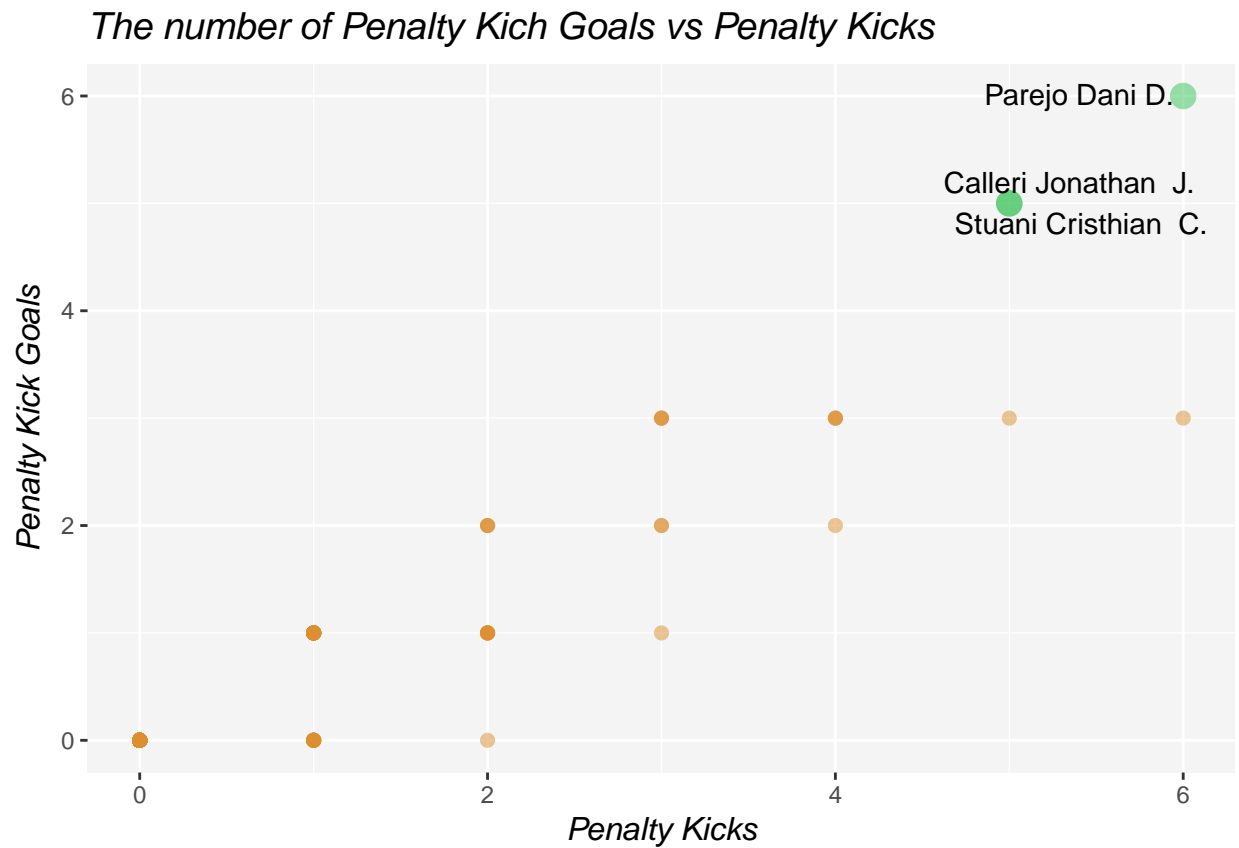
This plot names the top 10 players who have taken the corner kicks. Dani Parejo is leading with more than 150 corner kicks. It is notable that Barcelona was the 6th with the number of corner kicks but there is no player from Barcelona at this top 10 corner kick takers. That is why we decided to separately take a look at Barcelona's corner kick takers.

Number of Corner Kicks (players)



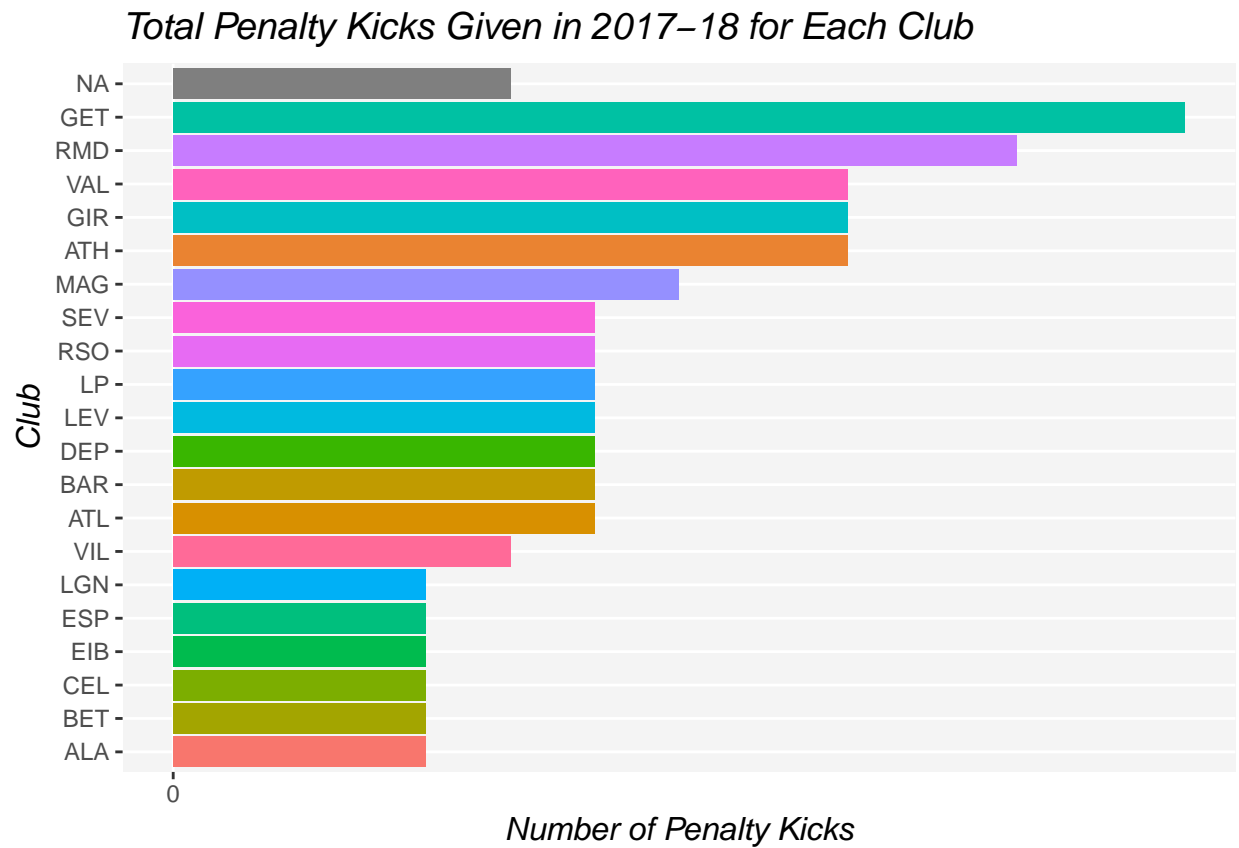
As we can see from the plot, Barcelona does not have a specific person who always takes the corner kicks, but instead has 3, which is why they were not included at the top 10 corner kick takers plot.

Penalty Kicks and Goals (players)



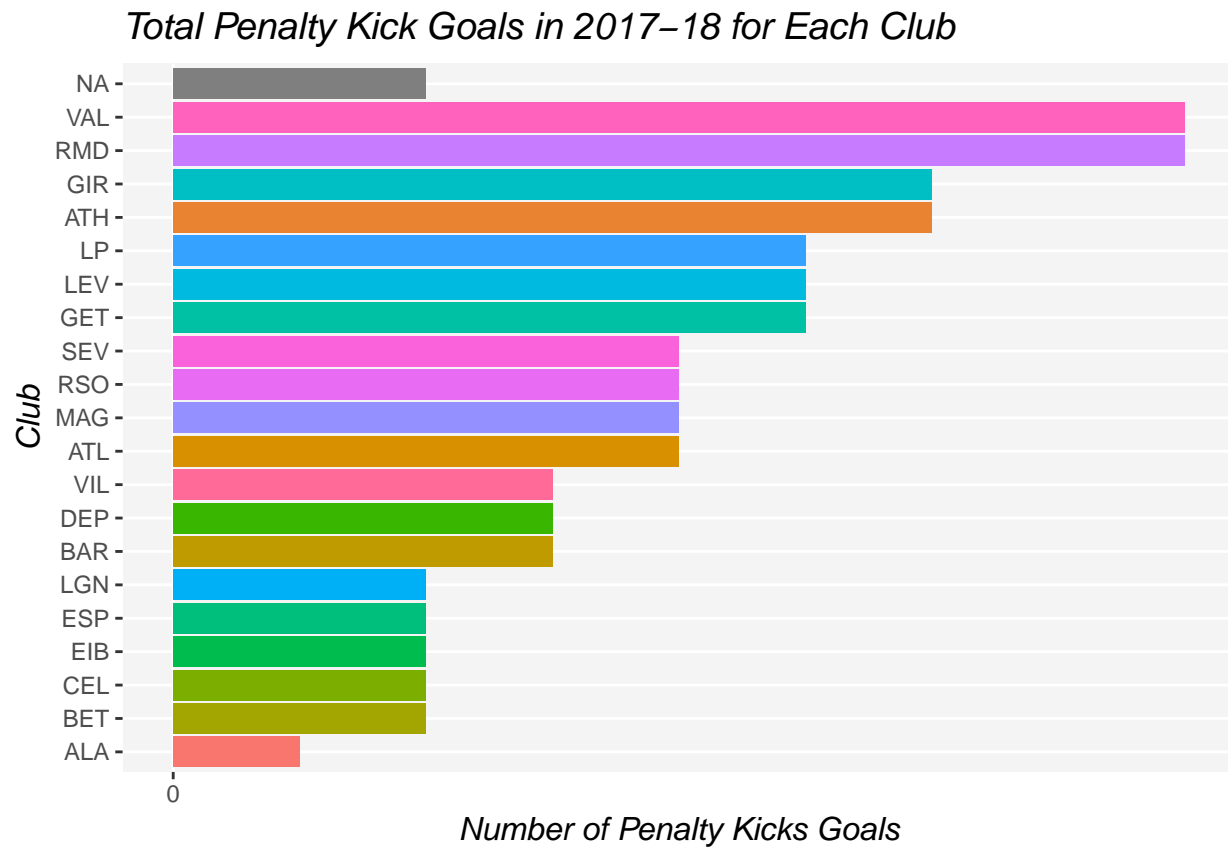
As the plot above demonstrates, Dani Parejo has taken the most penalty kicks and scored all of them. Jonathan Calleri and Cristhian Stuardi are the other two players who took more than four penalties and scored all of them.

Penalty Kicks (clubs)



The plot shows the number of penalty kicks taken by each club, with Getafe leading the board. Furthermore, Barcelona is not so high in this list, despite the common belief that referees award Barcelona too many penalties.

Penalty Kicks (clubs)



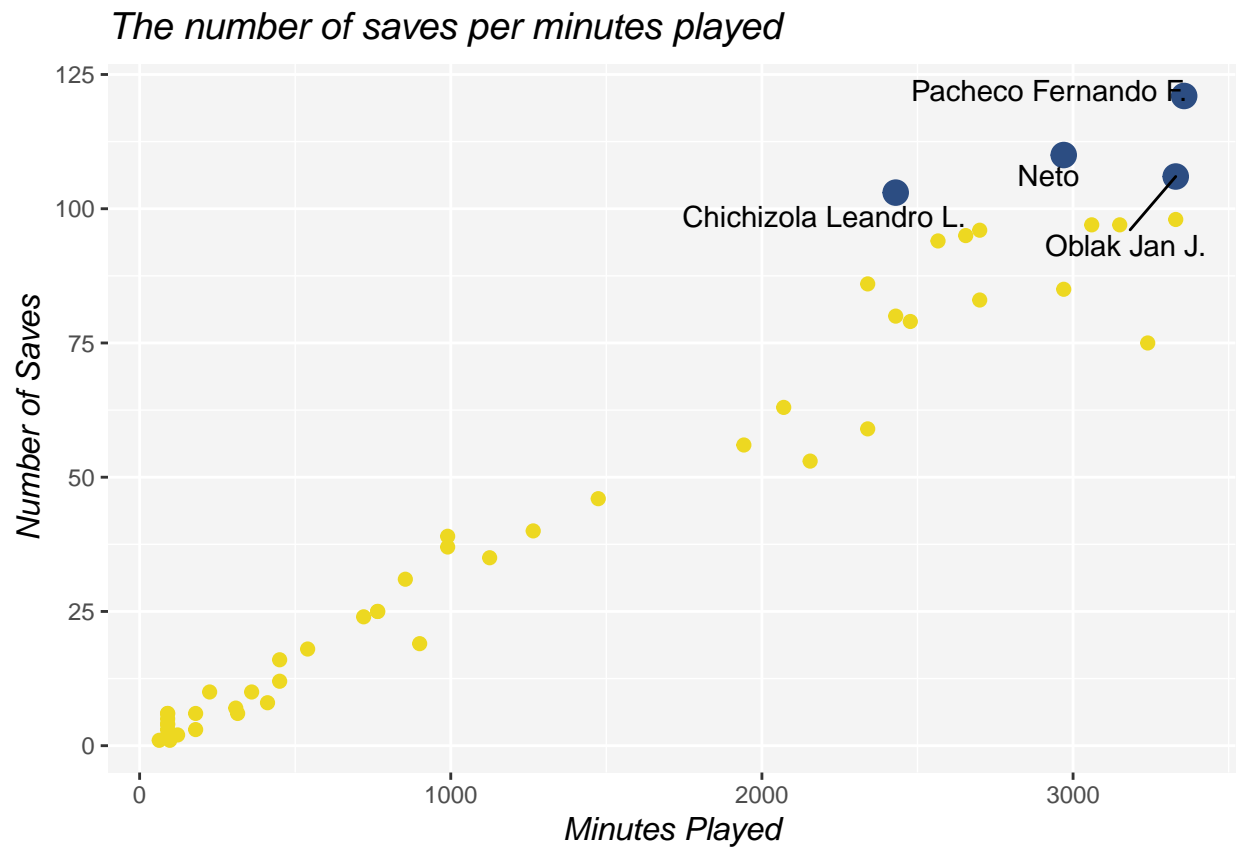
As the above plot demonstrates, Valencia and Real Madrid have scored the most penalties taken, while Alaves has scored the least.

Passes (clubs)



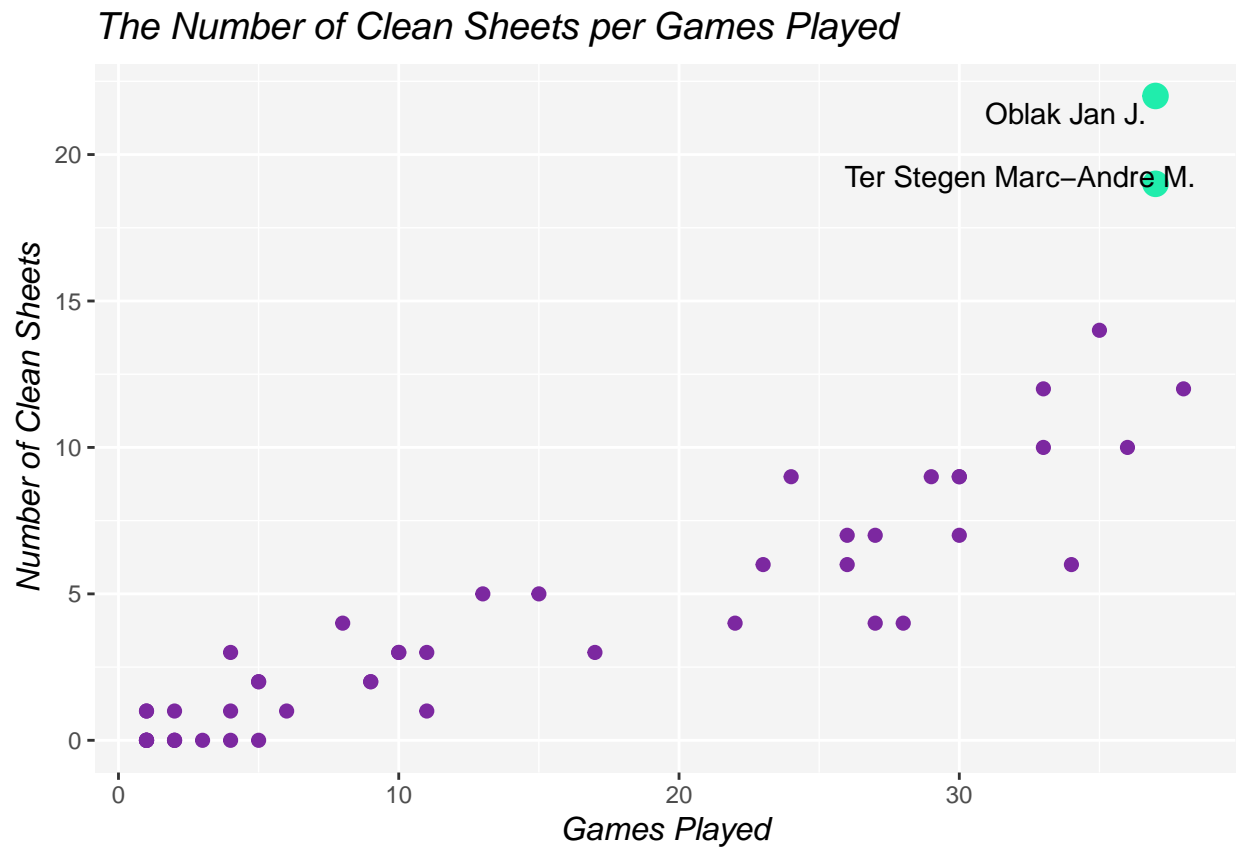
The above plot mentions the total number of passes throughout the 2017-2018 season by each club. As we can see Barcelona has the most number of passes, although right before the La Liga the manager of the club was changed, meaning that the new manager remains faithful to Barcelona's style of playing with many passes.

Number of Saves (goalkeeper)



The plot above demonstrates the number of saves per minutes played throughout the season. Goalkeepers at the top of the plot such as Fernando Pacheco, Neto, Jan Oblak and Leandro Chichizola have recorded the most saves overall throughout the season.

Number of Clean Sheets (goalkeeper)

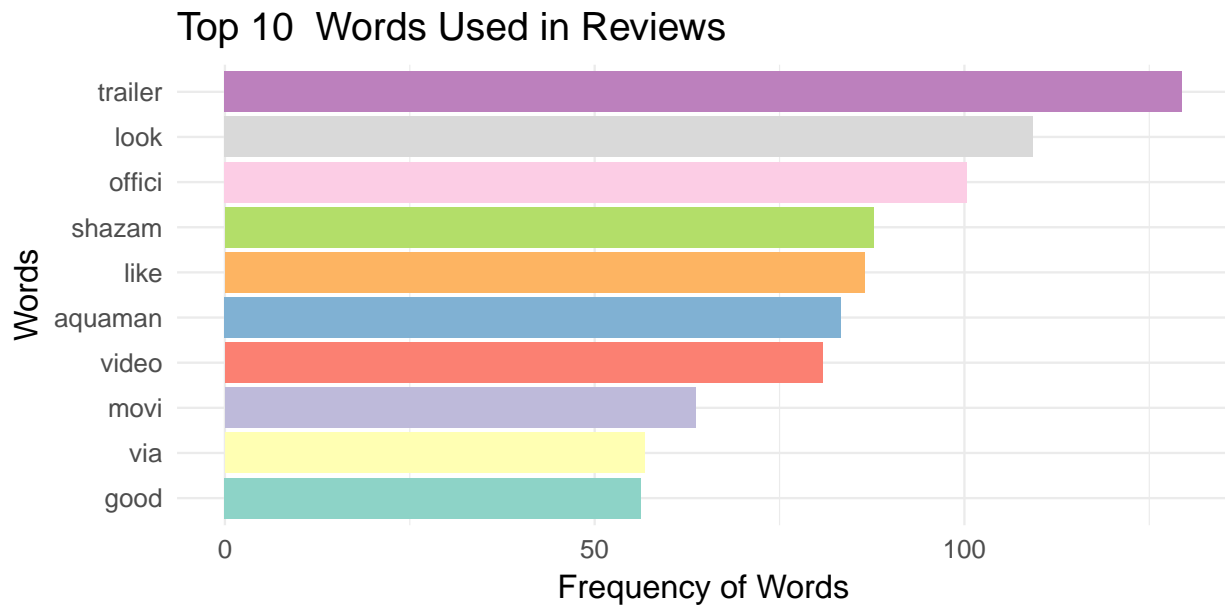


The plot above names the goalkeepers who have the most number of clean sheets recorded, As we can see from the plot above, Jan Oblak and Marc-Andre Ter Stegen are the goalkeepers with the most number of clean sheets.

Part #2

The second part of our homework represents analysis of text, particularly tweets made by different users about trailers of movies *Aquaman* and *Fantastic Beasts*. Warner Bros. revealed the trailers of these movies on 21 of July, 2018, and this topic became very popular that day. The analysis includes some graphs and wordclouds about common words used by the users. Moreover, it includes sentiment analysis of the tweets.

Most Common Words Used by Users (Aquaman)

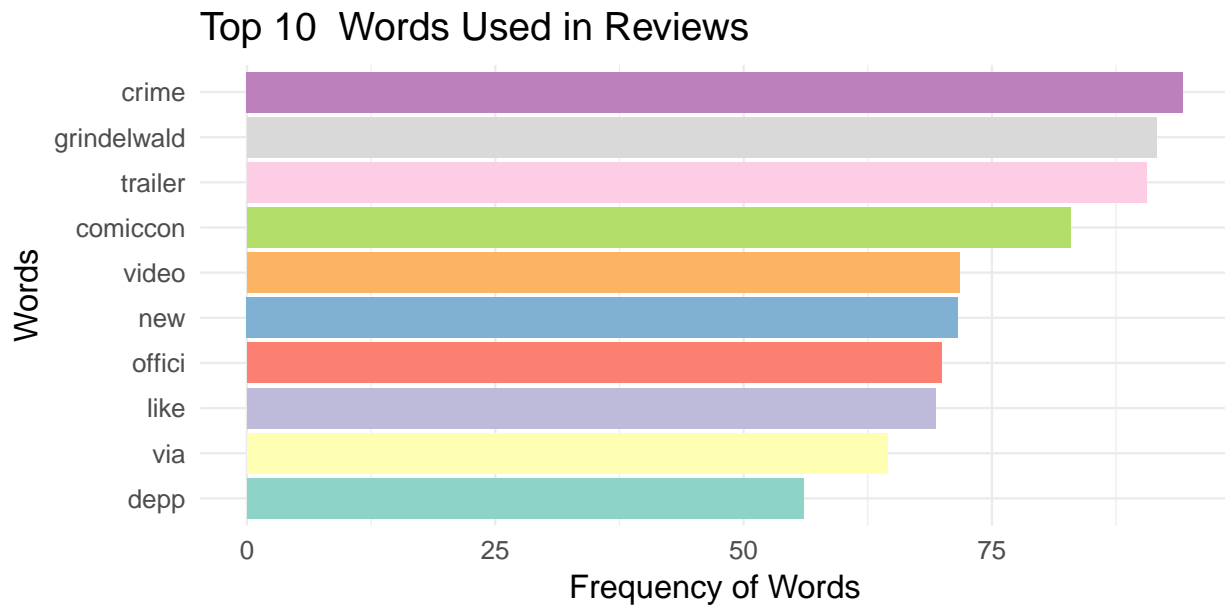


In the top 10 words list we can see that the most common word is trailer, which is logical. Moreover, we can see term “shazam”. This word is frequent because this is also a movie and its trailer was also released by that time. Probably, there were comments about comparison of those trailers.

We have also created a wordcloud to make the graph more appealing.



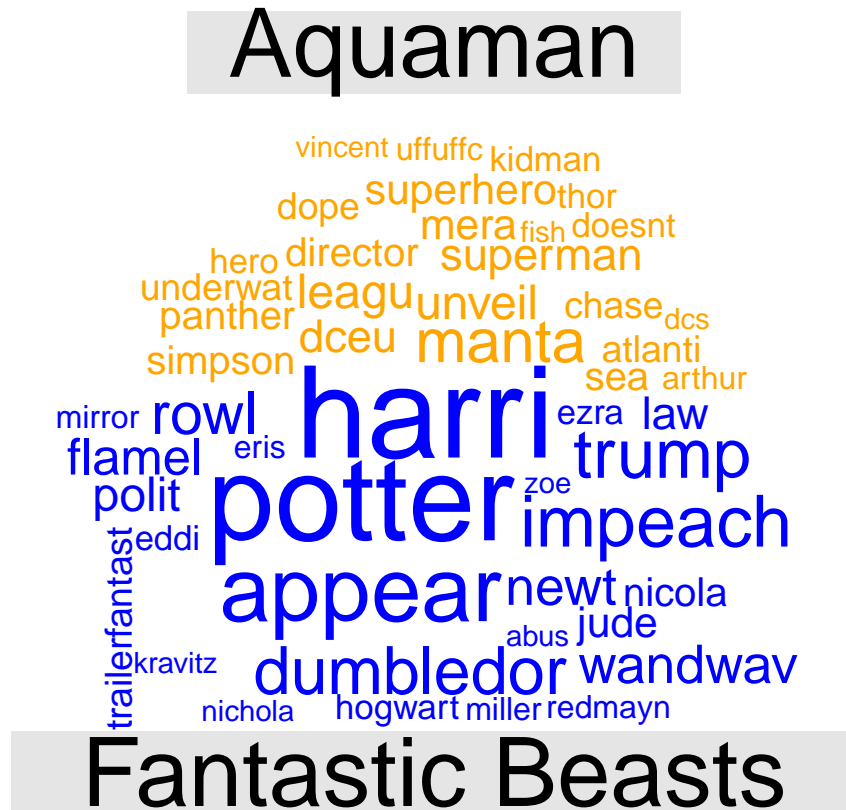
Most Common Words Used by Users (Fantastic Beasts)



As was expected, the most common words about “Fantastic Beasts” were “crime” and “Grindelwald”. Users were most interested in what is going to be the faith of one of the most dangerous Dark Wizards of all time.

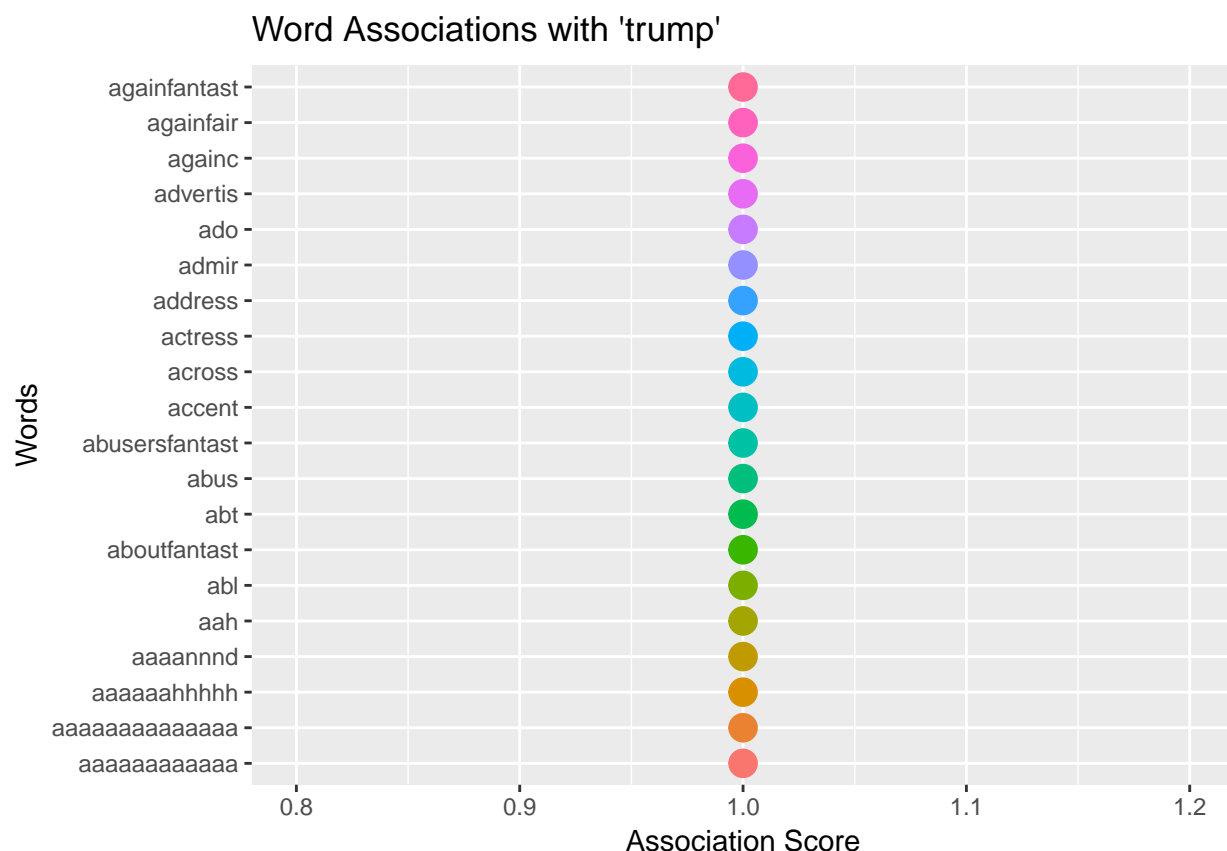
Frequencies of Words across Two Movies

Now, we want to compare the frequencies of words across two movies. For that reason, we will make a single term document matrix with two documents, one for “Aquaman”, the other for “Fantastic Beasts”.



Along with expected words, such as “Harry Potter”, “Dumbledore” and others, we see that the wordcloud contains the name of Donald Trump, which is a very interesting finding.

In order to find out why his name appeared in this topic, we decided to find word associations with his name. Here is the graph of word associations:



Actually, this graph did not help us much, except one word - “actress”. We decided to look if the name of the president was actually connected with the name of the actress.

We searched for the tweets, and found out this:

Tweets
Fantastic Beasts Comic Con Panel Explodes With Calls To Impeach Trump Not Everyone Is Thrilled
Fantastic Beasts Comic Con Panel Explodes With Calls To Impeach Trump Not Everyone Is Thrilled via
Fantastic Beasts Comic Con Panel Explodes With Calls To Impeach Trump Not Everyone Is Thrilled via
Fantastic Beasts panel gets political: 'Impeach Trump!'
Fantastic Beasts panel gets political at Comic-Con: 'Impeach Trump!' -
Zoe Kravitz Calls To Impeach Trump At 'Fantastic Beasts' Comic-Con Panel HuffPost I agree

Following the release of a new trailer for the film, cast members were asked what they would do if they could use a magic spell. Not missing a beat, Zoë Kravitz said, “Impeach Trump.” That is why Trump became popular in this topic.

Sentiment Analysis

In this section, we will introduce the emotions of Twitter users about movie trailers “Aquaman” and “Fantastic Beasts”.

Most Positive Tweets about Aquaman

Let's see, which are the most positive tweets about *Aquaman*

Reviews
Shazam looks incredible, Aquaman looks great too, but I think Im more excited for Shazam. It just looks fun, and like a friendly cartoony kinda film
WOW what a mind blowing and fascinating trailer truly great visuals unlike seen will surely brighten the Today was a great day to be a nerd. looks amazing. looks incredibly fun. seems like a great time. Plus I kinda have a Superman curl going on.
This is a much better trailer than Aquaman. This looks extremely entertaining, and hopefully clean enough for kids to also enjoy.

Here we see how users express their emotions about the film. As the table shows, most of the users express their opinions by comparing the trailer of the film with other ones.

Most Negative Tweets about Aquaman

Reviews
Aquaman looks dope as fuck.
Shazam looks mad funny, and Aquaman just looks dope! Lets go DCU!
That Aquaman movie looks fucking dope!
Aquaman is legit a fucking rip off of Black Panther its such a joke

Here are not very nice comments about the trailer.

Most Positive Tweets about Fantastic Beasts

Let's see, which are the most positive tweets about *Fantastic Beasts*

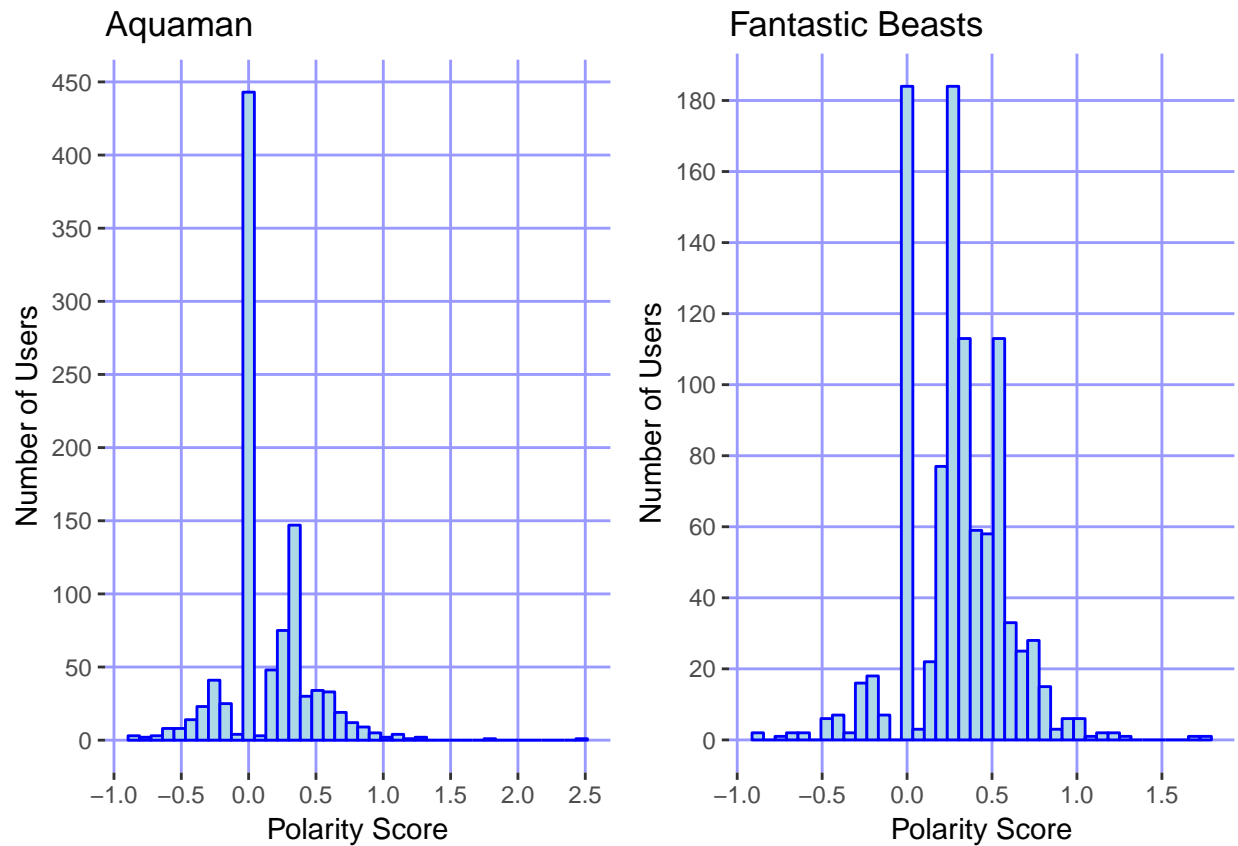
Reviews
SO MANY TRAILERS GODZILLA WILL BE GREAT AQUAMAN WILL BE GREAT SHAZAM WILL BE GREAT (YAY DC!) FANTASTIC BEASTS WILL BE GREAT GLASS WILL BE OKAY LOL
Trailer was decent, like.. I'm much more excited for the new Fantastic Beasts film, though. The trailer is incredible. ^^
Fantastic Beasts 2 Fantastic 2 Beasts Fantastic Beasts: Tokyo Drift Fantastic & Beasts Fantastic Five Fantastic & Beasts 6 Beasts 7 The Fate of the Beasts
Hands down Fantastic beasts! Amazing trailer

Most Negative Tweets about Fantastic Beasts

Reviews
Ok but if u didn't find fantastic beasts charming then I'm upset if i have to see one more head-ass post talking shit about how johnny shouldn't be in fantastic beasts or he shouldn't have showed up at sdcc imma lose my shit I am so sick and tired of every single media property out there having to be dark Loser in Real Life, gets up on stage and Acts as a Loser whilst in character.

Here we see negative feedback from users.

Distribution of Tweet Emotions



Note: Polarity scores with “minus” sign indicate negative emotions, 0 is for neutral tweets, and positive numbers indicate positive emotions.

Number of Positive, Negative and Neutral Tweets

Aquaman

Tweets	Frequency
Negative	132
Neutral	441
Posititive	427

Fantastic Beasts

Tweets	Frequency
Negative	65
Neutral	182
Posititive	753

As the histograms and the table suggest, there are many users who had neutral reaction to the trailer of Aquaman. For Fantastic Beasts, there are much more positive reactions, very few negative and neutral emotions.

Interactive Plots

Please, see interactive plots on the datasets of Aquaman and Fantastic Beasts in the section “*Interactive Plots*”.