

# Module 4. Data for analyzing high-press in football

## Introduction

The main goal in this module 4 is to learn about tactical defensive data that is useful for measuring the ability for exerting high-press in a team. In this module, our intention is to study how data can help coaches to make optimal decisions at the defensive level.

For that purpose, we are going to make a brief introduction about what is the big data when applied to football, which are the data providers that exist, from which of them can we extract data on defensive pressure, which data is going to be useful for us for measuring a team's ability to put high pressure and how can we interpret them.

For doing this, we are going to make comments on the following points:

1. Introduction to big data in football.
2. Main current data providers in football. From where can we obtain data on defensive pressure?
3. Which data is going to be useful for measuring a team's ability to exert high pressure?
4. Keys to a good interpretation of data



# Unit 4.1. Introduction to big data in football

Usage of big data and artificial intelligence in football has increased a lot in the last years and both are becoming very important tools for getting better results and optimizing professional teams and athletes' performance.

The application of the sports big data on football tries to generate an advantage over the opponent, not only in the tactics and strategies to use on matches, but also on the transfer market or football players' care.

In a football match, an approximate amount of eight millions of data is captured synchronized with the match time, which register the information generated by players and the ball in the game space. This information- which is not seen on the surface, since the human eye is only capable of retaining 30% of what happens during the match- is stored and treated through the big data tools, in order to be used and looked up for taking decisions of economic, tactical, physical nature.

That is why, clubs and coaches use these technologies for improving decision-making. And much more in the professional football, since they have more resources for getting access to these types of tools and they also have the possibility of hiring staff that would know how to use them and take advantage of them more efficiently. (Web-Futbol, n. d., <https://www.web-futbol.com/big-data-en-el-futbol/>).

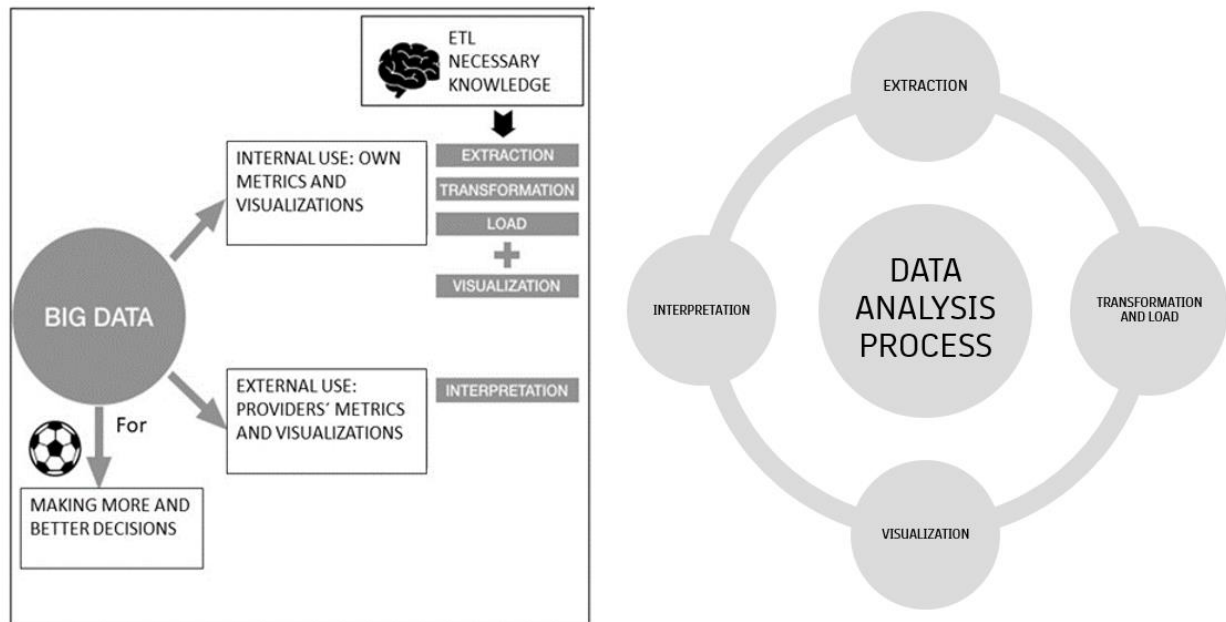
But, how can a coaching staff use big data? Coaches and analysts can use data generated by big data in football in two ways:

- 1) Outsourced usage: they can hire services from one or some data providers like StatsBomb, Opta, Instat, Wyscout, and others, which, in their analysis platforms, offer data metrics and visualizations that are closed and with their own format, which can be consulted, downloaded and used for our analyses and presentations. This way of using data does not require too much knowledge about big data.
- 2) Internal usage: staff generates its own metrics and visualizations from data collected from various providers. Providers previously mentioned also have the option for downloading raw data they generate in their API (Application Programming Interface), for later treating them with programming language programs (R or Python) and having an own visualization with programs like Power



BI or Tableau. This way of generating information and their later visualization requires a great knowledge about the processes related to big data and ETL (extraction, transformation and load).

**Figure 1: Usage of big data by a coaching staff**



Source: own creation

Below, there are some considerations to be taken into account by coaching staffs that do not have so many economic resources for hiring a provider's services:

- a) On the one hand, there are professional leagues and federations' platforms that facilitate free access to images and tracking and eventing data (raw data), as long as the club in question competes on professional leagues like the Spanish La Liga (Mediacoach), the German Bundesliga (Match Facts) or the English Premier League (DVMS).
- b) On the other hand, there are several consult websites (Understat, Fbref, WhoScored, Transfermarkt, etc.) which offer free access to eventing data we can download for working it our way through the web scraping.

In conclusion, the most important thing and what is going to make a difference for doing a good usage of big data applied to football is to have not only knowledge about football, but also knowledge about big data and the ETL data study process, to which we would add the display and the data interpretation.

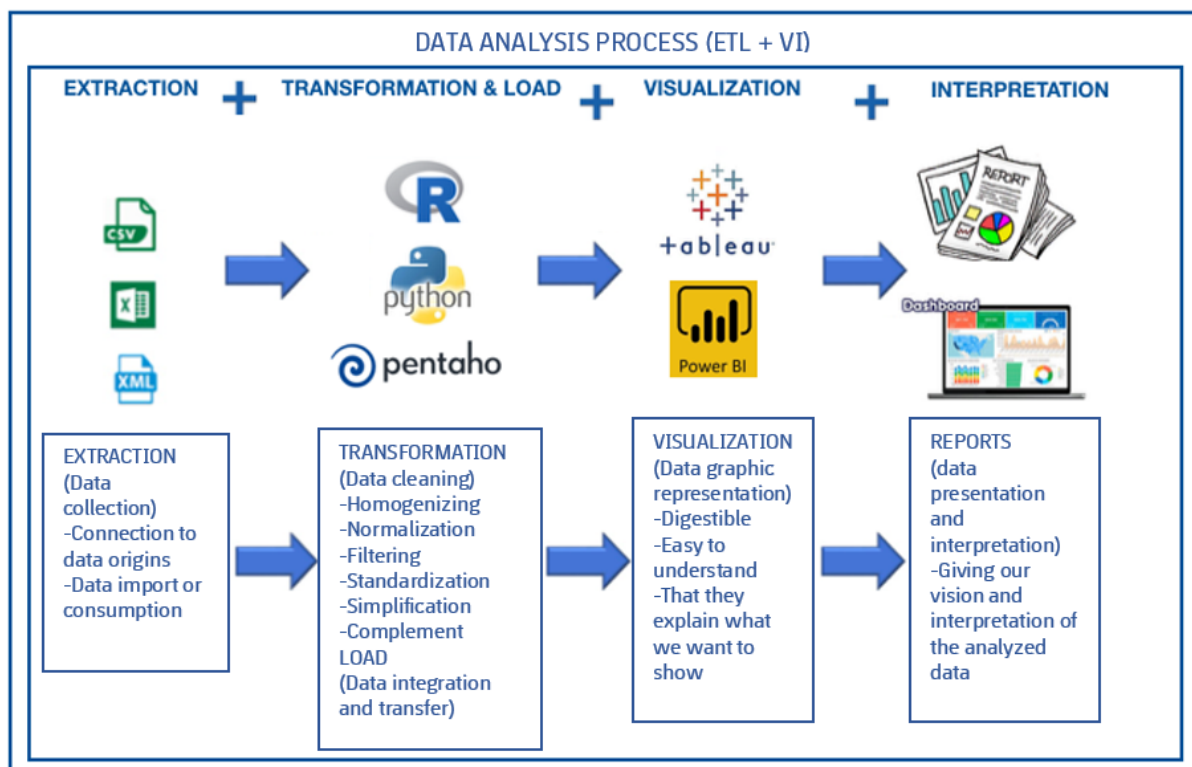


- 1) Extraction: data collection, import and consumption.
- 2) Transformation: data cleaning and homogenization.
- 3) Load: data integration and transfer.
- 4) Visualization: data graphic representation.
- 5) Interpretation: data presentation and interpretation.

In order to apply it in our daily work routine and for it to be really a useful tool for generating knowledge that can be applied to our work as coaches and/or analysts, we should learn about the data analysis process. The big data:

It makes reference to the study and analysis of big data quantities with the aim of extracting value from that information. This data management and analysis cannot be treated through conventional techniques, but through techniques such as the pattern recognition (machine learning) or predictive analyses for foretelling changes in data volumes. (Oviedo, 2020, p. 6).

**Figure 2: How to use big data?**



Source: own creation

If we want to make an internal and own use of the analysis process—by filtering information in the shape of data that derives from different providers and by giving the



visualization we want—we need to have a little knowledge about the analysis processes of big data and the necessary tools for its application.

1. Knowledge about data extraction: to learn about and know how to extract, from the different providers, the different file formats that have the information and that are importable. These file formats could be the following: .csv, .excel, .xml, and .Json. It is important to learn about providers and the data that interests us from each of them.
2. Knowledge about data transformation and load: to learn about and to know how to use the programming language tools, like R or Python, or the data integration tools like Pentaho, which are the ones that are used most currently for extracting, transforming and loading data from the format proposed by providers to the format we are most interested in:

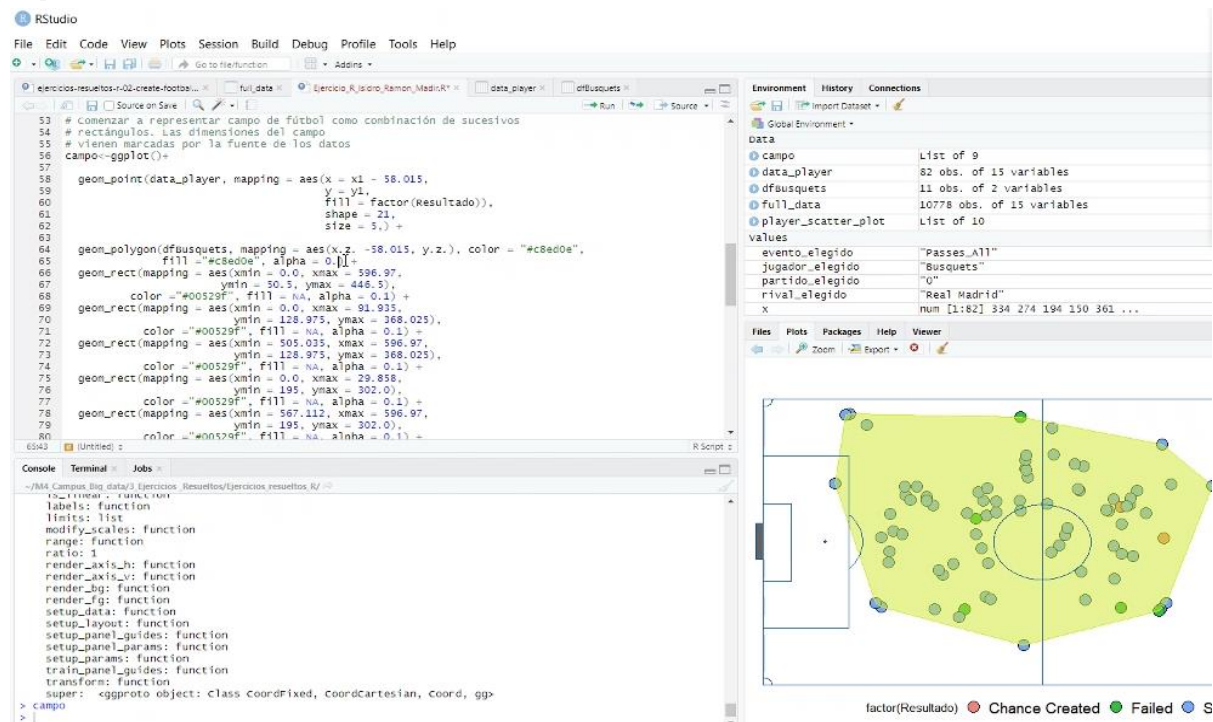
2.1. R: it is a free programming language. This environment is commonly used for statistical computing and graphics and it runs on systems like Unix, Linux, Windows and MacOS.

The R's characteristics and the different applications that make it a basic tool for data analysts. It is used in all stages in the data analysis:

- Acquisition of data from the available sources: data bases, text files, etc.
- Data preparation: duplicates elimination, incorrect data, extreme values, etc.
- Data analysis: building of predictive models, classification models, clustering models, etc.
- Communication of results: producing reports for presenting results and conclusions.
- Application of results obtained: for example, using predictive models developed for predicting certain build-up plays, in relation to a series of historical data (training and model testing data). (Unir, 2019, <https://www.unir.net/ingenieria/revista/lenguaje-r-big-data/>).



Figure 3: Example of R interface



Source: Print screen of the RStudio software.

2.2. Python: it is a high-quality free programming language, which uses the well-known “scripting language” and runs for almost everything. The great majority of computing applications and web platforms depend on its language, since it is used by most of developers, data scientists, software engineers and even hackers, due to its versatility, flexibility and characteristics oriented to objects. Python language can be executed on different operating systems like Unix, Linux, MacOS and Windows.

It is important to learn how to manage it, mainly on the emerging data sciences like data analysis (big data), artificial intelligence and machine learning.

Through the diverse libraries that connect Python with the exterior world, this language can perform the following functions:

- To interact with the exterior world, for which it will be necessary to read and write data on different formats (.csv, .excel, .Json, .txt, .dat, binary, etc.).
- Preparing data by cleaning and generating structured data from unstructured data.
- Transformation of data through the application of math operations and/or statistics to certain group of data for obtaining new data as a result (aggregation operations, segregation operations and filer operations).



- Presentation of data and results obtained by making use of static and dynamic graphics (with JavaScript), localization maps, graphos, etc.
- Automatic learning models in which data is connected with automatic learning algorithms—supervised or not—in order to obtain analytical models that could predict the future, to a greater or lesser degree. (Fombella Pombal, 2020a, p.10).

Figure 4: Example of Python interface

```

In [110]: def extraer_datos_debut(df):
           df['club_debut']=df['Debut deportivo'].map(str) #esta funcion es para que no falle el apply, pasando la columna a str.
           df['club_debut']=df['club_debut'].apply(lambda x: x.split('(')[-1].split(')')[0])
           df['club_debut']=df['club_debut'].apply(lambda x: x.replace('2013','9999').replace('2011','9999').replace('2009','9999'))
           df['club_debut']=df['club_debut'].apply(lambda x: x.replace('2002-04','C. D. Tenerife "B"'))
           df['club_debut'].replace('nan','9999',inplace = True)
           df['club_debut'].replace('9999',np.nan,inplace = True)

           df['fecha_debut']=df['Debut'].map(str)
           df['fecha_debut']=df['fecha_debut'].apply(lambda x: x.split(' de ')[-1])
           df['fecha_debut']=df['fecha_debut'].apply(lambda x: x.replace('[1]', '').replace('[2]', '').replace('1 ','')
           .replace('4', ''))
           df['fecha_debut']=df['fecha_debut'].apply(lambda x: x.split('1/')[0])
           df['fecha_debut']=df['fecha_debut'].apply(lambda x: x.replace('2018 (República Centroafricana)', '2013'))
           df['fecha_debut']=df['fecha_debut'].apply(lambda x: x.replace('-', '9999').replace('nan', '9999'))
           df['fecha_debut'].replace('9999',np.nan,inplace = True)

           df['club_debut']=df['Debut deportivo'].map(str)
           |

           # No modifique las siguientes líneas
           assert(isinstance(df, pd.DataFrame))
           return df

In [111]: df_5=raw_data[['Debut','Debut deportivo']].copy()
           extraer_datos_debut(df_5)
           df_5

Out[111]:
           Debut                Debut deportivo  club_debut  fecha_debut
0           11 de noviembre de 2017                NaN        NaN        2017

```

Source: Print screen of the Jupyter software (Pérez, 2021).

Last Checkpoint el viernes pasado a las 20:57	Last Checkpoint last Friday 8:57 pm
La columna "club debut" tendrá valores nulos cuando "debut deportivo" tenga valores nulos	The column "debut club" will have non-existent values when "sport debut" have non-existent values
Extraer datos debut	Extract debut data
Esta función es para que no falle el apply, pasando la columna al str.	This function is for avoiding failure of the apply, going from the column to the str.
No modifique las siguientes líneas	Do not modify the following lines
Debut 11 de noviembre de 2017	Debut November 11 <sup>th</sup> 2017
Debut deportivo	Sports debut
club_debut	club_debut
fecha_debut	date_debut

2.3. Pentaho: it belongs to the Hitachi Vantara company and it is a business intelligence (BI) platform oriented and centered on processes solutions, which include the components required for implementing solutions based on processes

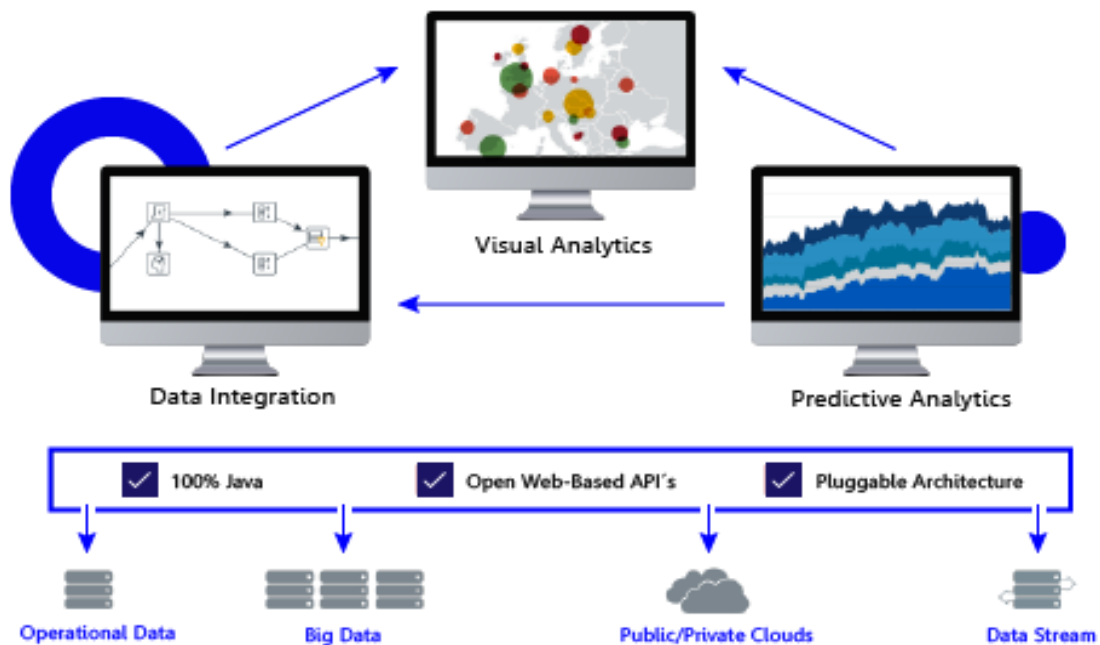


like data mining, ETL, reports' generation, etc. It is used a lot for improving the ability for analysis and decision-making, since it offers solutions that are mainly made up of a built-in tools infrastructure for analysis and reports.

Pentaho includes inside its suite the tools that cover the following areas:

- Pentaho Data Integration: data integration tool (ETL).
- Pentaho Reporting: creation of traditional reports.
- Pentaho Dashboard Designer: a component only available on the Enterprise version.
- Pentaho Community Dashboard Editor: a plugin for the creation of dashboards.
- The rest of the tools for creating reports and visualizations have been replaced, in many cases, by last generation tools like Tableau or Microsoft Power Bi, which we will see later on. (Fombella Pombal, 2020, pp. 42-44).

**Figure 5: Pentaho tools**



Source: [Untitled image about Pentaho tools]. (n. d.). Retrieved from [https://lh3.googleusercontent.com/4y-ZxvheJkzXhDva8cLQ0AccgCwF1wFHRD9PP\\_SWnRQarn\\_OIyVU5Ktwc1cgrJY71bYvxQ=s130](https://lh3.googleusercontent.com/4y-ZxvheJkzXhDva8cLQ0AccgCwF1wFHRD9PP_SWnRQarn_OIyVU5Ktwc1cgrJY71bYvxQ=s130)

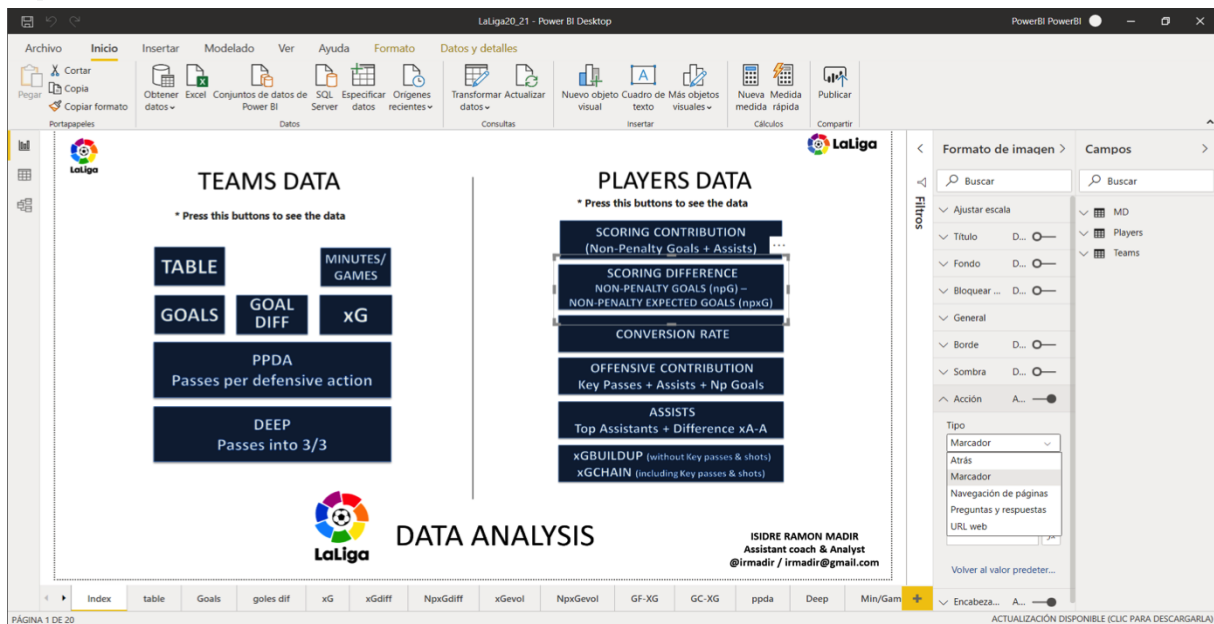
To sum up, Pentaho is a very useful tool for integrating and unifying data from different formats and providers in a unique format, which allows for the quickest and most efficient ETL, for working with visualization tools.

3. Knowledge about data visualization: to learn about and know how to use the most powerful visualization tools in the market. These are Power Bi and Tableau:

3.1. Power BI: Microsoft Power BI is the main visualization tool inside Microsoft Office 365. It is only available for Windows operating systems. Its great number of users and acceptance inside the market is because of its simple operation and low cost. We will be able to create interactive dashboards very quickly.

In any environment (companies, sports clubs, etc.) it is essential to unify all the information, whether available on the cloud or locally, around one unique platform. For that purpose, Power BI is an ideal tool for using knowledge about business intelligence (BI), about real-time data exploitation based on different sources, as well as for creating reports and boards and for later offering, in a simple visual and intuitive way, the results of all those elements in which we are interested for analyzing and sharing them among several professionals in the same club.

Figure 6: Example of visualization in Power BI



Source: Print screen of the Power BI Desktop software (Microsoft, 2021).

3.2. Tableau: it is a powerful data visualization tool, used in the business intelligence area. It simplifies raw data in a format that is very easy to understand.

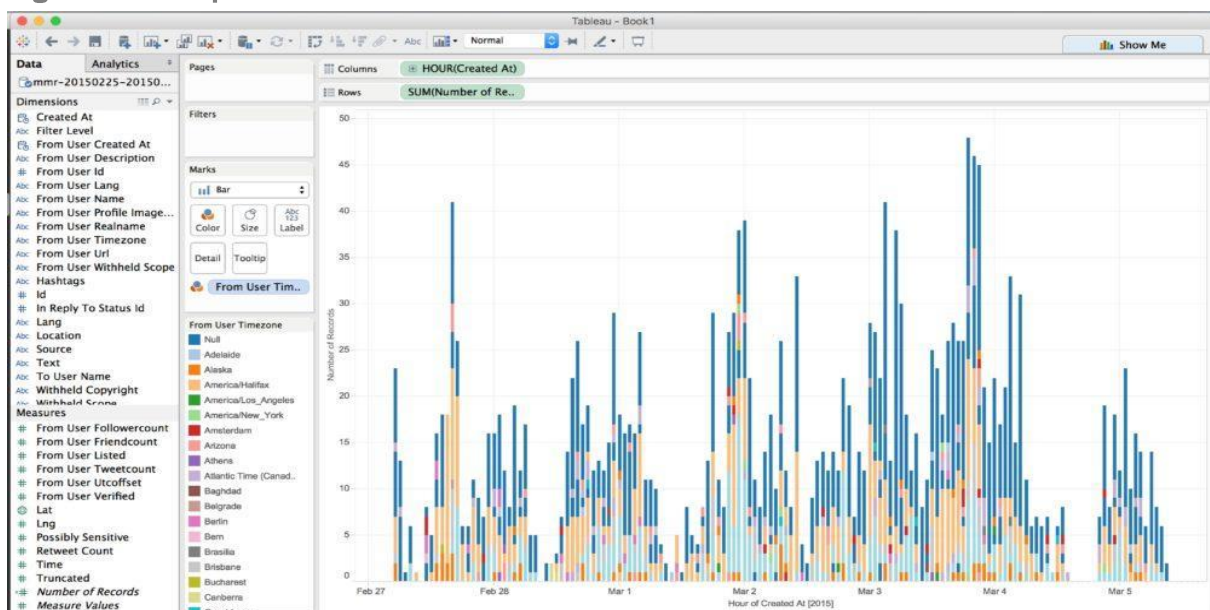


Tableau's essence is simple and, at the same time, very relevant: to help people and companies to see and understand all of their data. It achieves it by offering a selection of useful and intuitive business intelligence tools:

- Tableau Desktop: it connects and analyzes data and information.
- Tableau Server: thanks to the server, we will be able to collaborate safely and to share information based on the data we have uploaded through Tableau Desktop (the software's desk version).
- Tableau Online: it is about a version of Tableau Server stored in the cloud. This way, we will be able to have access to our data without the need of going through a tedious installation process. (SoftwarePara, n. d, <https://softwarepara.net/tableau/>).

Through simple functions like dragging and dropping, any person can have access to and analyze data in a simple way and they can even create reports and share that information with other users.

**Figure 7: Example of Tableau visualization**



Source: De Juana, 2019, <https://www.muycomputerpro.com/2019/06/11/tableau-asi-es-la-empresa-por-la-que-salesforce-ha-pagado-mas-de-13-000-millones>

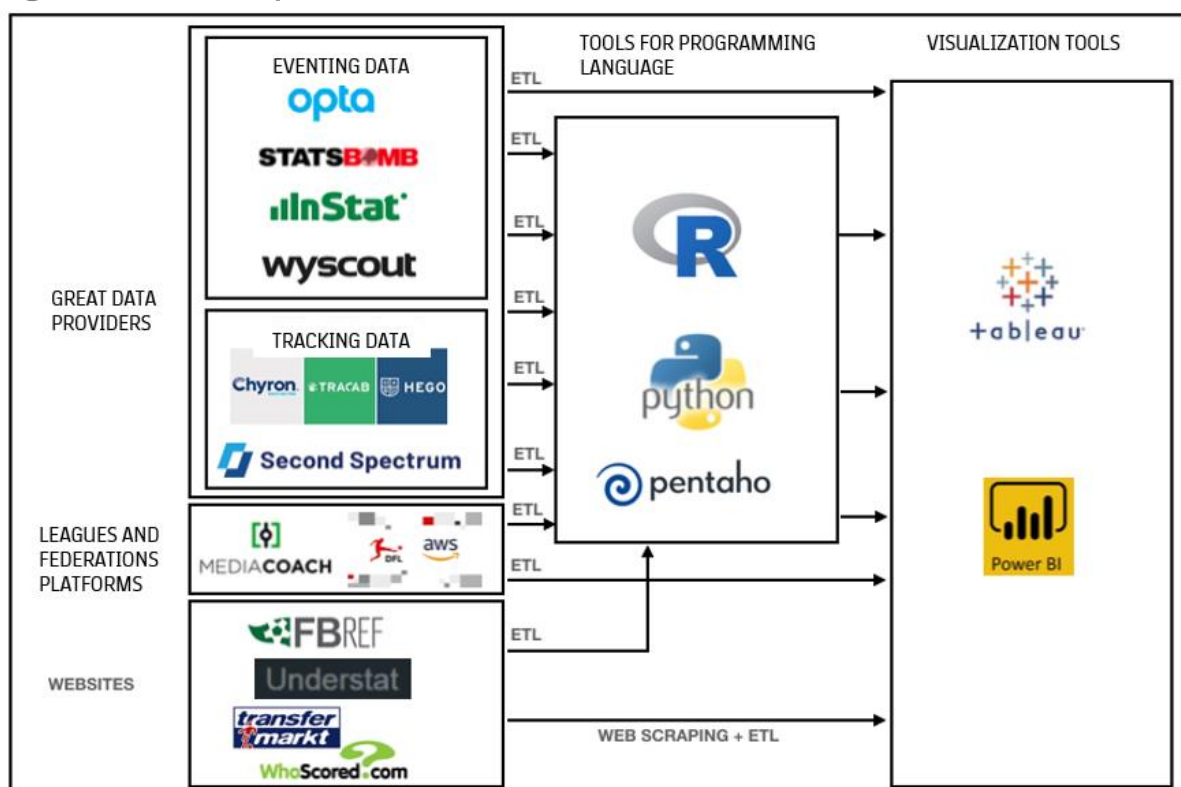


## Unit 4.2. Main data providers in football. From where can we obtain data on defensive pressure?

Following the professor master in sports big data at UCAM university, David Fombella Pomba, we are going to enumerate the main data providers and the different programming and visualization tools that analysts are using nowadays.

There are two big data providers that currently monopolize the provision of eventing data to the clubs, to the European national leagues' platforms (Mediacoach and others) and even to web sites specialized in football. We are talking about Opta and StatsBomb. There are also other providers like InStat, Wyscout, as well as Tracab or Second Spectrum (the work of these last two is focused on providing tracking data):

Figure 8: Main data providers in football



Source: own creation

1. Opta: the brand Opta belongs to the Perform group, a leading group focused on content and digital sports media. This company focuses on data collection, reception and storage for many varied sports, for distributing this information to the clubs that



work with them. Although they also work in other business areas (television, betting shops, sports journalism, etc.)

On the one hand, Opta has a group of widgets to insert in a web page. It is very common to see this type of content obtained in digital newspapers, since they include from basic concepts to results of advanced metrics and indicators. In relation to football, some of these available widgets allow for access to metric visualizations like the already known expected goals (xG), the defensive zones, the heatmap, team and player comparison, etc..

Figure 9: Opta widget of defensive zones

**Zonas defensivas incluye un polígono con la zona de las acciones defensivas de cada jugador (Faltas, Entradas, Intercepciones, Tiros Bloqueados y Despejes)**



Source: Fombella Pombal, 2020b, p. 21.

Zonas defensivas incluye un polígono con la zona de las acciones defensivas de cada jugador (Faltas, Entradas, Intercepciones, Tiros Bloqueados y Despejes)

Defensive zones that include an area with the defensive actions zone of each player (Fouls, Tackles, Interceptions, Blocked Shots and Clearances)

On the other hand, Opta's data feeds, not only for eventing but also for tracking with their downloadable statistics data from their API, are available for clubs to use them as they please, whether for sports performance analysis or for scouting or communication and press services. For example, the F24 (Event Detail Feed) "is one of the mostly used Opta packages, since it includes varied information and nuances" (Fombella Pombal, 2020b, p. 31).



Figure 10: Opta Feed F24 (Event Detail Feed)

```
<Event id="1970142897" event_id="296" type_id="16" period_id="1" min="25" sec="14" player_id="20664"
team_id="43" outcome="1" x="78.4" y="38.2"
timestamp="2011-08-21T16:25:53.299" last_modified="2011-08-22T13:51:14">
  <Q id="357476634" qualifier_id="103" value="22.2" />
  <Q id="1145873833" qualifier_id="102" value="52.1" />
  <Q id="1577156338" qualifier_id="113" />
  <Q id="902661166" qualifier_id="72" />
  <Q id="174975923" qualifier_id="55" value="295" />
  <Q id="49122906" qualifier_id="29" />
  <Q id="97008017" qualifier_id="56" value="Center" />
  <Q id="934091399" qualifier_id="77" />
  <Q id="18684958" qualifier_id="136" />
  <Q id="65794542" qualifier_id="18" />
  <Q id="924818303" qualifier_id="160" />
  <Q id="994511953" qualifier_id="118" />
</Event>
```

Ahora nos toca descifrar la información incluida dentro del evento y sus calificadores:

**Evento**

- **id** → Evento número 296 del partido
- **type\_id** → 16 Gol
- **player\_id** → 20664 D. Silva en el Feed 7 está la correspondencia entre id y nombre.
- **team\_id** → 43 (Manchester City)
- **period\_id** → 1 (1ª parte)
- **minuto y segundo** → 25:14
- **x e y del campo** → 78,4 y 38,2
- **outcome** → 1 Gol

Source: Fombella Pombal, 2020b, p. 32.

Ahora nos toca descifrar la información incluida dentro del evento y sus calificaciones	Now we have to decipher the information included inside the event and its qualifiers:
Evento	Event
Evento número 296 del partido	Event number 296 of the match
16 Gol	16 Goal
20664 D. Silva en el Feed 7 está la correspondencia entre id y nombre	20664 D. Silva on Feed 7 is the correspondence between id and name
1 (1ª parte)	1 (1st part)
minuto y segundo	minute and second
x e y del campo	x and y in the field
1 Gol	1 Goal

2. StatsBomb: it is a recently set up company in the United Kingdom, with a global team of committed and experienced professionals, with branches in Europe, Egypt and the United States. It has quickly established, together with Opta, as a leading company in the football analytics field.

StatsBomb, which covers more than forty leagues around the world and has the most precise and granular data in the market, provides the teams with the necessary tools for competing and winning in the court and in the transfer market.



Its analysis platform, StatsBombs IQ, does not only provides the analysts information about teams, players and leagues, but it also helps the different communication media through personalized data visualizations.

Inside consulting services they offer, we can find the following:

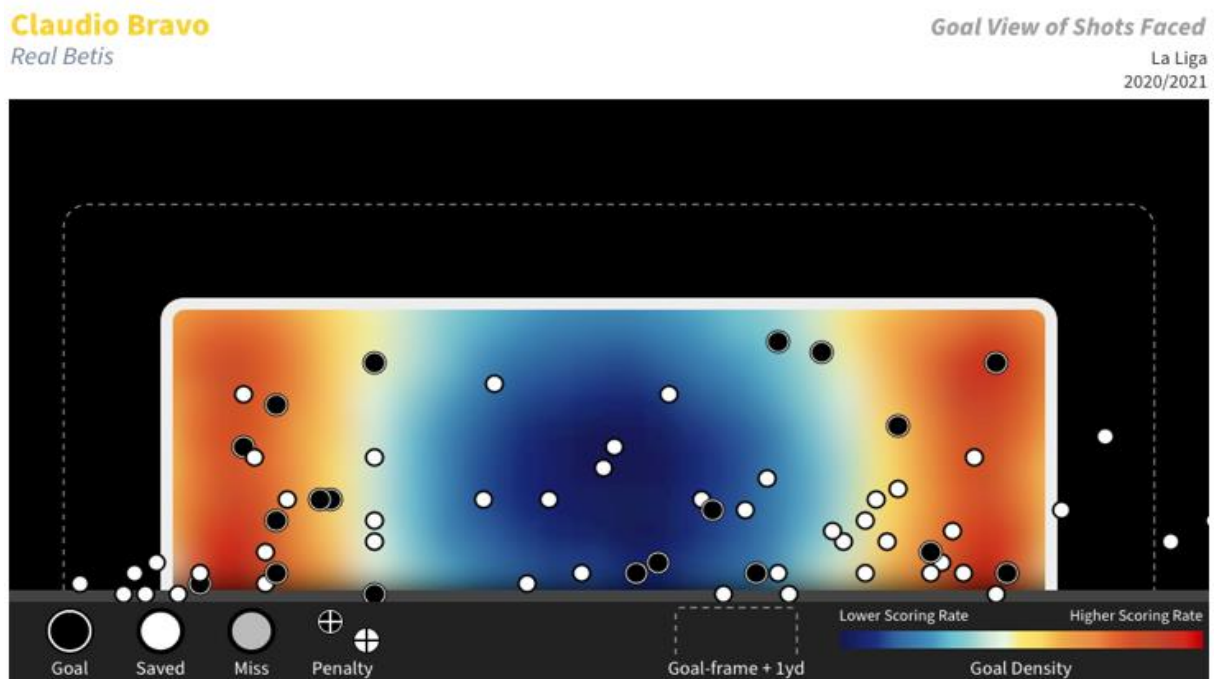
- Players' analysis (sports performance, personal information, recruiting reports, player's weaknesses and strengths, market values, etc.).
- Search for coaches according to players' game style.
- Set pieces analysis. (Fombella Pombal, 2020b, p. 138).

Above all, it stands out because of its advanced defensive metrics, like the defensive pressures, counter-pressing, recoveries after pressing, etc. and its new metrics for measuring the goalkeeper's performance.

StatsBomb includes revolutionary metrics and visualizations for the goalkeepers' analysis. Let's forget about things that are as basic as goals/attempts on goal for evaluating our goalkeeper's performance. StatsBomb includes the goalkeeper's positioning metrics on the goal's line, performance in catching crosses, contribution in attack, defensive activity, successful passes under pressure, etc. (Fombella Pombal, 2020b, p. 143).



Figure 11: Example of the goalkeeper's metrics in relation to shots received in StatsBomb



**STATSBOMB**

Source: Ovalle, 2021, <https://www.biobiochile.cl/noticias/futbol-internacional/chilenos-en-el-exterior/2021/05/27/que-dicen-los-datos-bravo-fue-el-tercer-arquero-que-mas-goles-evito-en-la-liga-la-temporada-pasada.shtml>

StatsBomb also has an Api where customers can download the whole data for processing it and for making their own visualizations, as well as an open archive in GitHub with information on some competitions in order to visualize its data format and make tests. It also includes an R library (<https://github.com/StatsBomb/StatsBombR>) that is connected to its API and obtains data. (Fombella Pombal, 2020b, p. 144).

Finally, StatsBomb never stops innovating and advancing on metrics and it offers more in-depth data and analysis than the ones we can achieve with the eventing. It's the one called StatsBomb 360.



Figure 12: StatsBomb 360



Source: StatsBomb, s. f., <https://StatsBomb.com/360-data/>

<p>Scouting</p> <ul style="list-style-type: none"> <li>-Identificar jugadores que rompen líneas con sus pases</li> <li>-Encontrar jugadores que reciben el balón entre líneas, al pie, o al espacio</li> <li>-Evaluar la toma de decisiones con balón</li> </ul>	<p>Scouting</p> <ul style="list-style-type: none"> <li>-Identifying players that break lines with their passes</li> <li>-Finding players that receive the ball among lines, into the feet, or into space</li> <li>-Evaluating decision-making with the ball</li> </ul>
<p>Análisis de equipo</p> <ul style="list-style-type: none"> <li>-Identificar debilidades de organización defensiva</li> <li>-Evaluar líneas de pase abiertas en cada situación</li> <li>-Analizar qué tipo de pases desequilibran la estructura defensiva</li> </ul>	<p>Team analysis</p> <ul style="list-style-type: none"> <li>-Identifying defensive organization weaknesses</li> <li>-Evaluating open passing lines in each situation</li> <li>-Analyzing which type of passes destabilize the defensive structure</li> </ul>
<p>Análisis de rivales</p> <ul style="list-style-type: none"> <li>-Analizar la estructura defensiva y ocupación de espacios en transiciones</li> <li>-Identificar espacios abiertos en diferentes fases de juego</li> <li>-Examinar el posicionamiento y presión en salida de balón</li> </ul>	<p>Opponents' analysis</p> <ul style="list-style-type: none"> <li>-Analyzing the defensive structure and space occupation on transitions</li> <li>-Identifying open spaces on different game stages</li> <li>-Examining positioning and pressure in build-up play</li> </ul>

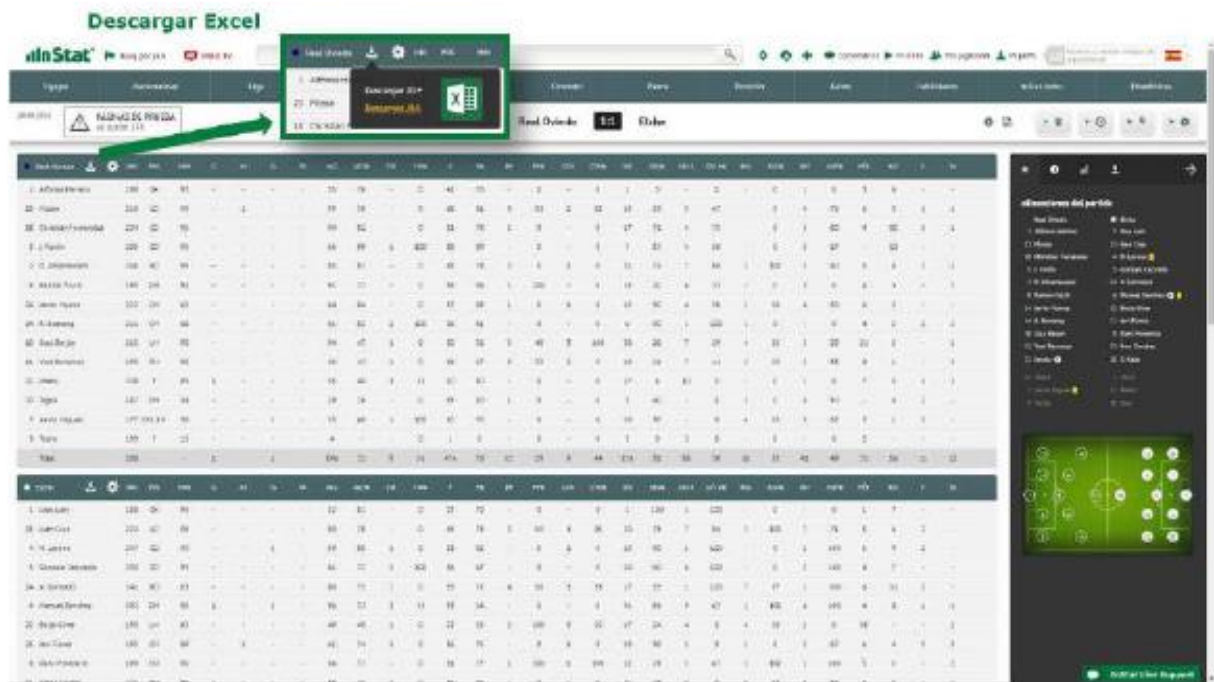
3. Instat and Wyscout: they are other data providers that are very established in clubs (90% of professional clubs have hired their services). . They are very similar and very useful tools for clubs, sports directors, coaches, analysts, goalkeepers' coaches, referees, the players themselves and several communication media companies,

They are one of the best tools for scouting and recruitment, since they not only offer images in TV formats, but also visualizations of the huge number of data they generate about teams and players. They also offer the possibility of downloading their raw data for later treatment, with the possibility of generating their own visualizations for customers:

Instat offers the match's detailed statistics in which they include the players' analyses (Instat index, goals, assistances, passing precision,

crosses and cross precision, dribbles, disputes, duels, losses of ball possession, ball recoveries, fouls and cards). From this page, we can download an Excel file with a tab for teams and another one for players, which will allow us to work with this information with visualization tools like Microsoft Power BI or Tableau. Fombella Pombal, 2020b, p. 67).

Figure 13: InStat: detailed statistics about a match



Source: Fombella Pombal, 2020b, p. 67.

Descargar Excel	Download Excel
-----------------	----------------

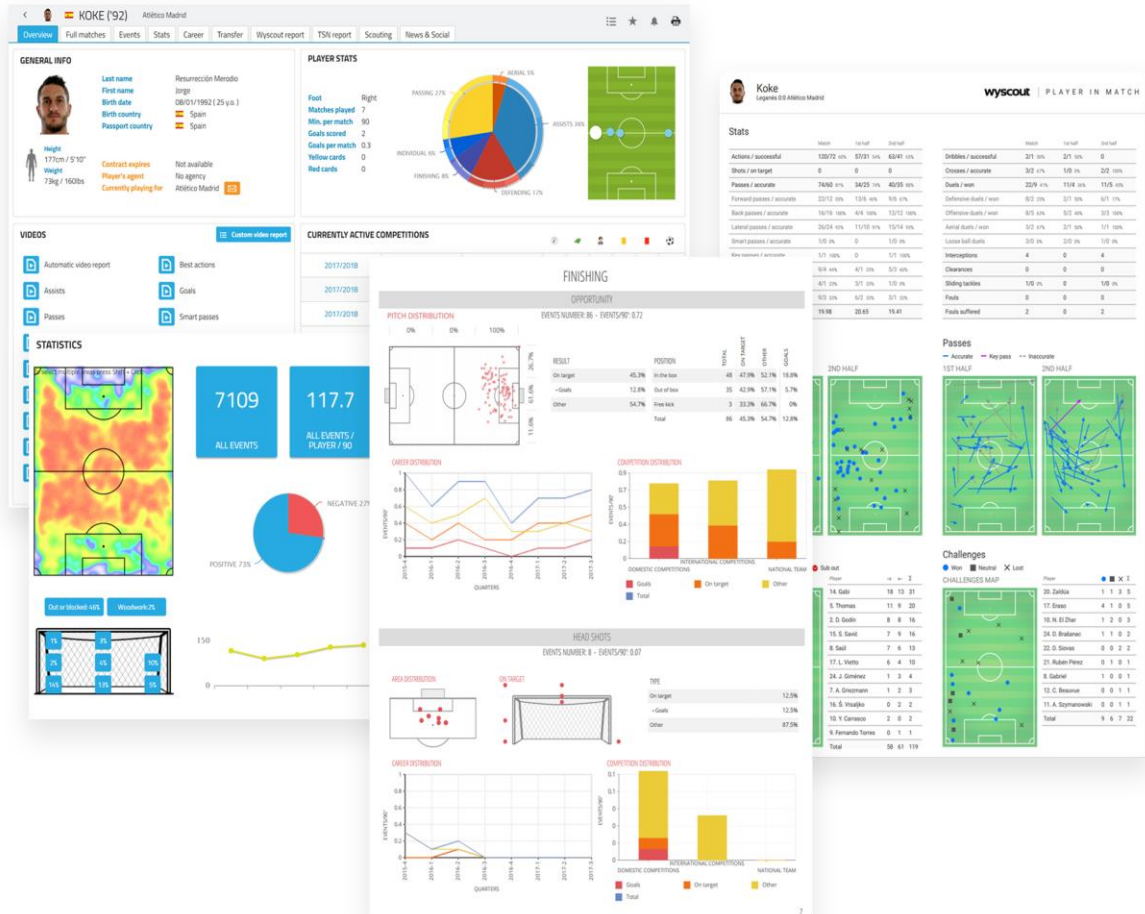
Wyscout is a tool that is very similar to InStat. In 2019, it became a Hudl associate and, together with the analysis tool SportsCode, it has two of the tools that are mostly used by professional clubs around the world.

Some of the advanced tools for statistical analysis can help in getting access to information about all the game's aspects. The functions that the Wyscout application offers are the following:

- Detailed and complex statistical reports that are downloadable (data from technical and tactical events).
- Rankings of more than 200 competitions.
- Players' analysis: by the Advanced search, we will be able to discover new talents through statistical features.

- With the Wyscout’s API we can download all of the data you may need for easily integrating them with the BI tools and the video analysis softwares like SportsCode or Nacsport. (Fombella Pombal, 2020b).

Figure 14: Wyscout: detailed statistics about a match



Source: Wyscout, n. d., <https://wyscout.com/es/plataforma-futbol/>

“The Wyscout Feed Events (events) offer information about the match with all of the events and their x/y positions in the playing field. This historical data base allows for an advanced mathematical analysis and the patterns detection” (Fombella Pombal, 2020b, p. 107).

Until now, we have talked about eventing providers, but there are other providers that are focused on generating tracking data.

Providers like Tracab and Second Spectrum capture, through intelligent robot cameras, the players and the ball’s movement, generating real time XYZ tracking raw data files.

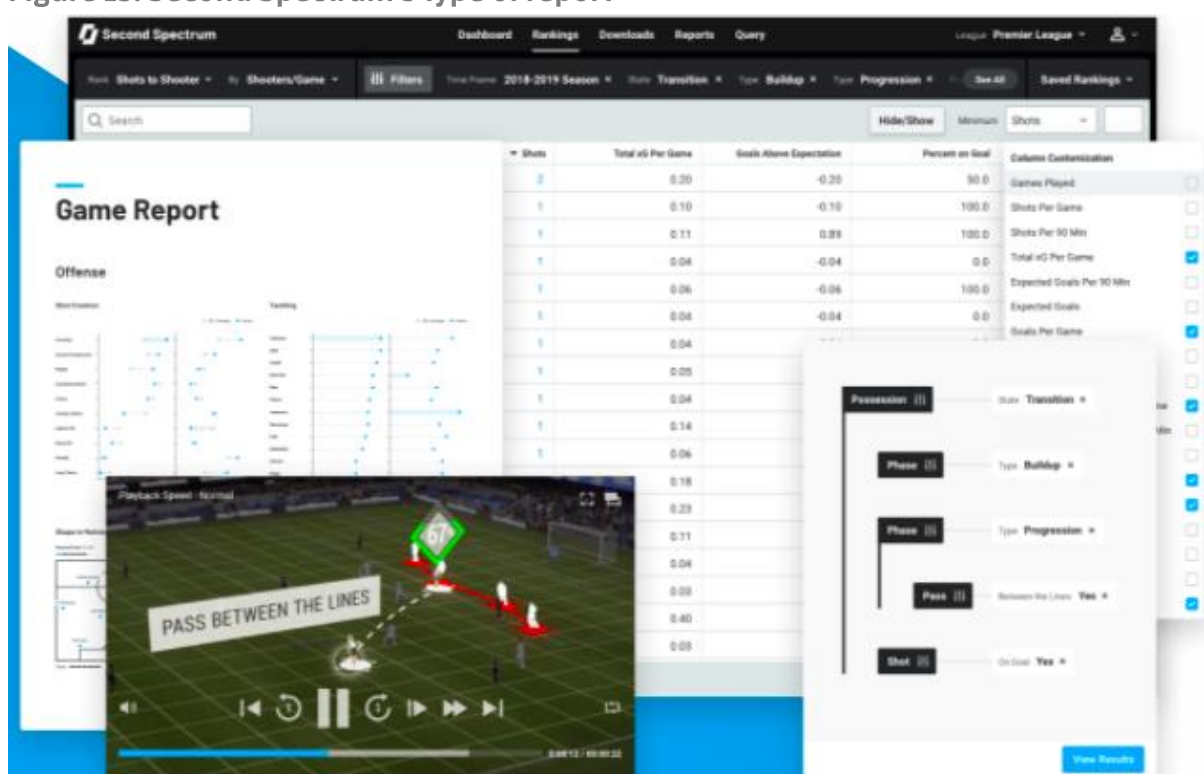


These files include XY information about the 22 players and XYZ about the ball 25 times per second (25 Hz), consequently they generate more than 3 million data per match:  
 $(90 \text{ m} \times 60 \text{ s/m} \times 25 \text{ frames/s} \times 23 \text{ (players + ball)}) = 3\,105\,000 \text{ data/match}$

4. Second Spectrum: it is one of the most advanced systems for tracking players. It applies last generation computer vision techniques and automatic learning for producing quick data, accurate for basketball, football, American football and other sports. It is the NBA or the English Premier League's official data provider.

The eventing and tracking data that it generates fosters new ways for teams and leagues to understand, evaluate, improve and create content about their game.

**Figure 15: Second Spectrum's type of report**



Source: Second Spectrum, n. d, <https://www.secondspectrum.com/ourwork/teams-leagues.html>

5. Tracab: it is one of the biggest tracking and eventing providers. It provides data to a huge number of football clubs and federations around the world (among them we can find the Spanish, German, Dutch, Mexican and Japanese leagues; the Champions League, etc.).

Tracab offers solutions in form of sports data, videos and visualizations, with very useful information for improving professional teams' performance, besides allowing an ever larger range of real-time solutions highly innovative and convincing, for getting sports fans involved more effectively.

We can see a clear example of this in the Spanish league: the Spanish professional football league offers the clubs the Mediacoach tool, which stocks up with Tracab tracking data together with Opta eventing data:

For elaborating each match's file, which is usually between 500 and 600 pages long with metrics and data, Mediacoach gets fed by two providers. On the one hand, there is Tracab, which is focused on tracking, that is to say, on the continuous tracking of every player on the field by some cams installed in every stadium of Liga Santander and La Liga 1|2|3, which are defrayed by La Liga. This system is in charge of providing the distance covered, the maximum and average speed, sprint numbers, etc. On the other hand, there is the eventing, in which three people in Opta register passes, losses of ball possession, fouls, corners, etc. Everything that is demonstrable. (Olmeda, 2020, <https://www.abc.es/contentfactory/post/eslaliga/laliga-apuesta-por-la-tecnologia-para-hacerse-mas-competitiva/>).

Figure 16: Example of a Mediacoach platform for the Spanish league

Source: López, 2019, <https://noesfutboleslaliga.elmundo.es/directos-al-futuro/posesion-pero-con-pegada-asi-es-el-adn-del-campeon-de-laliga-santander>

¿Qué es Mediacoach?	What is Mediacoach?
Suite Productos Mediacoach	Mediacoach Suite Products
Desktop	Desktop

Herramienta avanzada de video análisis profesional integrada con datos físicos y técnico-tácticos	Advanced tool for profesional video analysis integrated with with physical and technical-tactical data
Live Aplicación que permite a los cuerpos técnicos monitorizar los datos de un partido en tiempo real	Live Application that allows coaching staffs to monitor a match data in real time
Portal Repositorio de recursos online para analistas con la entrega de videos de partidos más rápida del mercado	Portal The quickest archive of online resources for analysts with delivery of match's videos
Reports Informes de rendimiento con métricas innovadoras validadas por analistas y expertos universitarios	Reports Performance reports with innovative metrics validated by university analysts and experts
Broadcast Datos de tracking y eventing para mostrar en la señal de televisión antes, durante y después del partido	Broadcast Tracking and eventing data for showing on TV before, during and after the match
Usuarios Mediacoach	Mediacoach Users

There are three Certificates on the right top that validate this software.

These platforms are the other data sources that offer this service to clubs and coaching staffs in the professional leagues.

6. Leagues and federations platforms: leagues and federations play a very important role in relation to access to information and data on the part of clubs. The creation of these information access tools or platforms, on the part of all coaching staffs (coaches, analysts and physical trainers) and all the analysis departments at the clubs in their respective leagues, has created an information access democratization, since, in the past, only the most powerful clubs could afford these providers for having access to their data.

This has fostered a lot quicker evolution in the data industry, which has implied an improvement of their features and utility and, at the same time, the access has been facilitated. By having access and using tracking data combined with eventing data (raw data), clubs can connect much more information for working it by themselves. This is done by outsourcing (consuming the own metrics and visualizations the platforms offer) or in a more internal way (working and transforming data with ETL processes that would allow them to create their own knowledge and visualizations).



By using these platforms, they can improve players' performance and reinforce fans' participation with a simple combination of tools that would allow them to do the following:

- a) To unify data and video: the data and video's capacity resulting from the tracking and eventing combination provides an incredible precision and an accelerated work flow, which allow analysts to create valuable information about planning, exploring and managing the load of players for their coaches.
- b) To improve teams' tactics: the access to raw data in these platforms provides technology that offers precise and adequate performance data through informative videos that support their analysis. This allows the coaching staff to learn about the team's performance over time, to develop specific improvement programs and to elaborate new tactics for achieving an even greater success.
- c) To make decisions during the match: these platforms offer real time content during the match for tracking the team's performance live, as well as events data with synchronized video, which allows coaches to make more and better decisions during the match, which gives them a crucial competitive advantage.
- d) To optimize the player's performance: the excellent tracking technology these providers use offers a positioning precision and an unbeatable data quality, which are useful for capturing the players' physical performance. This way, it is a key element to develop specialized training programs, to minimize injuries and to keep a combination of precise and consistent data, not only in training but also in the match.



Figure 17: Example of the German league platform Match Facts



Source: Deutsche Fussball Liga , 2020, <https://www.dfl.de/en/news/dfl-and-amazon-web-services-to-provide-new-real-time-match-analysis/>

7. Main football data web pages: Another source from where we can obtain relevant data and information about teams, players, matches and high-press defensive metrics are the web pages specialized in football.

Besides being public, free and besides even offering some kind of visualization (like the web Understat), the most interesting thing is the possibility of downloading its data (Fbref). This can also be done by web scrapping, using some of the language programming tools mentioned in this module (R or Python) or directly from the visualization tools (Power BI or Tableau).

No doubt, these web pages play an big role for clubs and coaching staffs without economic resources. We are going to make comments on this:

7.1. Fbref: it is a free web page with the most advanced data that exists currently in the world of football. The thing is that thanks to their association with StatsBomb, it offers metrics even more advanced that web pages like WhoScored.

It is closest to what is offered by Opta, StatsBomb, Wyscout and Instat, but public and free. They never stop improving their web site, adding novel data and metrics. For example, they incorporate data like xG (expected goals), pressing defensive data, advanced data about the goalkeeper, etc.

Besides, they offer the total data and its calculation x90 and they arrange it very attractively (standard metrics, metrics for shots, passes, creation of chances, possession, matches and time played, defensive actions, etc.).

Figure 18: Example of the Fbref web page

Squad Defensive Actions 2020-2021 La Liga [View Player Stats](#)

Squad Stats		Opponent Stats																							
Squad	#	PI	90s	Tackles			Vs Dribbles				Pressures			Blocks				Int	Tkl+Int	Clr	Err				
				Tkl	Tkl/W	Def 3rd	Mid 3rd	Att 3rd	Tkl	Att	Tkl%	Past	Press	Succ	%	Def 3rd	Mid 3rd					Att 3rd	Blocks	Sh	ShSv
<a href="#">Alavés</a>	30	38.0	575	356	295	216	64	185	554	33.4	369	5447	1510	27.7	1752	2541	1154	568	105	4	463	432	1007	1065	9
<a href="#">Athletic Club</a>	27	38.0	523	343	235	227	61	158	552	28.6	394	5250	1451	27.6	1494	2380	1376	553	86	2	467	348	871	897	4
<a href="#">Atlético Madrid</a>	25	38.0	704	449	346	267	91	288	671	42.9	383	5124	1540	30.1	1665	2328	1131	579	93	0	486	395	1099	797	3
<a href="#">Barcelona</a>	25	38.0	548	351	240	226	82	206	511	40.3	305	5075	1546	30.5	1404	2330	1341	519	82	0	437	354	902	474	11
<a href="#">Betis</a>	25	38.0	605	378	289	240	76	196	557	35.2	361	5263	1514	28.8	1619	2495	1149	531	84	2	447	432	1037	855	10
<a href="#">Cádiz</a>	34	38.0	611	385	337	218	56	222	623	35.6	401	5264	1433	27.2	2094	2304	866	612	118	4	494	411	1022	950	10
<a href="#">Celta Vigo</a>	30	38.0	702	419	353	279	70	222	690	32.2	468	5997	1701	28.4	2063	2850	1084	610	97	1	513	375	1077	966	9
<a href="#">Eibar</a>	30	38.0	572	362	249	236	87	189	564	33.5	375	5015	1602	31.9	1272	2251	1492	553	75	2	478	390	962	965	8
<a href="#">Elche</a>	30	38.0	598	394	346	213	39	193	608	31.7	415	5168	1455	28.2	2068	2228	872	553	111	2	442	388	986	866	8
<a href="#">Getafe</a>	31	38.0	617	383	261	278	78	199	558	35.7	359	6085	1689	27.8	1602	3055	1428	557	68	4	489	420	1037	1129	6
<a href="#">Granada</a>	34	38.0	567	374	276	226	65	188	509	36.9	321	5027	1358	27.0	1616	2345	1066	611	142	5	469	406	973	955	9
<a href="#">Huesca</a>	27	38.0	634	401	339	233	62	219	606	36.1	387	5451	1578	28.9	1951	2462	1038	551	103	1	448	348	982	775	6
<a href="#">Levante</a>	29	38.0	573	368	289	216	68	200	602	33.2	402	5363	1499	28.0	1697	2350	1316	538	108	3	430	445	1018	991	12
<a href="#">Osasuna</a>	28	38.0	569	354	255	236	78	204	536	38.1	332	5329	1550	29.1	1558	2487	1284	557	109	2	448	384	953	1027	3
<a href="#">Real Madrid</a>	30	38.0	572	353	285	211	76	184	552	33.3	368	4880	1425	29.2	1562	2175	1143	535	100	4	435	406	978	667	9
<a href="#">Real Sociedad</a>	30	38.0	560	342	251	235	74	172	529	32.5	357	5070	1487	29.3	1423	2340	1307	497	104	2	393	337	897	859	5
<a href="#">Sevilla</a>	26	38.0	571	389	253	242	76	200	546	36.6	346	4716	1477	31.3	1392	2199	1125	503	95	3	408	334	905	767	8
<a href="#">Valencia</a>	30	38.0	508	312	259	183	66	187	574	32.6	387	4855	1305	26.9	1794	2067	994	566	137	2	429	321	829	951	14
<a href="#">Valladolid</a>	32	38.0	607	376	338	202	67	220	669	32.9	449	5839	1624	27.8	2051	2600	1188	601	117	2	484	390	997	989	8
<a href="#">Villarreal</a>	29	38.0	566	340	253	242	71	177	580	30.5	403	4994	1425	28.5	1587	2226	1181	578	117	3	461	354	920	777	8

Totals may not be complete for all senior-level play, see [coverage note](#).  
 Expected Goals ([xG explained](#)) and other Advanced Data provided by [StatsBomb](#), and is available for [these competitions](#).

Source: print screen of the Fbref web page, <https://fbref.com/en/comps/12/La-Liga-Stats>

7.2. Understat: In this web page you will be able to find information with which you can analyze the teams and the players in the five big leagues in Europe, with advanced metrics like the xG (expected goal), NPxG (non-penalty expected goal), xPTS (expected points), xA (expected assist), etc. and, at the defensive level, the already known PPDA (passes per defensive action) stands out.

Just as Fbref uses data generated by StatsBomb, Understats uses Opta data, which is why you should know that there are slight differences in data, since each provider uses their own algorithms for generating it.



Figure 19: Example of the Understat web page

Table		Charts		overall	home	away	Start date	End date													Settings
Nº	Team	M	W	D	L	G	GA	PTS	xG	NPxG	xGA	NPxGA	NPxGD	PPDA	OPPDA	DC	ODC	xPTS			
1	Atletico Madrid	38	26	8	4	67	25	88	57.88	-8.34	52.31	35.87	+10.67	32.70	+18.61	10.32	12.38	310	185	68.41	+8.59
2	Real Madrid	38	25	9	4	67	28	84	65.07	-1.83	62.84	38.70	+10.70	32.75	+30.09	10.14	14.03	328	171	69.98	+4.02
3	Barcelona	38	24	7	7	85	38	79	83.98	-1.04	78.01	40.93	+2.93	37.22	+40.80	9.51	16.20	499	188	78.18	+0.82
4	Sevilla	38	24	5	9	53	33	77	55.88	+2.86	50.68	34.33	+1.33	29.84	+20.81	8.15	14.58	240	139	68.98	+0.04
5	Real Sociedad	38	17	11	10	59	38	62	63.15	+4.75	54.22	38.80	+6.80	35.08	+19.14	8.08	11.70	285	180	67.61	+5.61
6	Real Betis	38	17	10	11	50	50	61	49.77	-0.23	41.43	47.51	-2.48	40.08	+1.35	9.12	10.55	224	180	55.37	+5.63
7	Villarreal	38	15	13	10	60	44	58	57.88	-2.14	48.19	43.67	-0.33	38.98	+11.21	9.85	11.43	291	182	63.77	+5.77
8	Celta Vigo	38	14	11	13	55	57	53	51.32	-3.88	46.12	50.31	-6.69	44.37	+1.75	7.40	10.84	217	187	52.83	+0.17
10	Granada	38	13	7	18	47	65	46	41.05	-5.88	37.34	59.82	-5.18	50.81	-13.27	10.74	8.88	128	234	40.58	-5.42
9	Athletic Club	38	11	13	14	46	42	46	43.08	-2.84	39.35	38.48	-3.54	34.00	+5.34	9.75	9.80	202	197	54.77	+6.77
12	Cadiz	38	11	11	16	38	58	44	38.70	+0.70	33.73	58.91	+6.91	51.48	-17.75	16.57	5.93	77	274	39.01	-4.99
11	Osasuna	38	11	11	16	37	48	44	39.61	+2.81	35.15	49.83	+1.83	43.88	-8.74	10.91	8.73	172	243	45.10	+1.10
13	Valencia	38	10	13	15	50	53	43	41.89	-8.11	34.48	55.89	+2.88	49.74	-15.28	10.94	8.81	182	282	42.38	-8.82
14	Levante	38	9	14	15	46	57	41	40.83	-5.07	38.47	55.13	-1.87	52.18	-15.69	9.47	10.22	221	245	43.53	+2.53
16	Alaves	38	9	11	18	36	57	38	39.11	+3.11	33.17	51.55	-5.45	49.32	-18.15	12.25	7.88	119	212	43.23	-5.23
15	Getafe	38	9	11	18	28	43	38	35.82	+7.82	31.91	40.40	-2.80	34.32	-2.42	8.81	8.43	134	165	48.39	+10.39
17	Elche	38	8	12	18	34	55	38	30.40	-3.50	28.17	61.83	+5.63	54.20	-28.03	13.23	8.87	78	280	30.31	-5.69
18	SD Huesca	38	7	13	18	34	53	34	37.98	+3.88	35.00	48.84	-4.38	44.92	-9.92	11.32	8.53	98	215	45.23	+11.23
19	Real Valladolid	38	5	18	17	34	57	31	37.92	+3.82	31.97	52.58	-4.44	47.38	-15.39	10.57	7.97	134	243	41.53	+10.53
20	Eibar	38	6	12	20	29	52	30	42.45	+13.45	35.64	48.84	-3.18	45.12	-9.48	8.99	10.58	149	168	48.82	+18.82

Source: print screen of the Understat web page, [https://understat.com/league/La\\_liga](https://understat.com/league/La_liga)

7.3. Transfermarkt: it is a web page that has data about the teams and their players, but a web that is focused on transfer market and on the players' value on the market. Inside it, we can find news about transfer rumours, last transfers and renovations, assignments, players that have finished a contract and the ones who are free, the most expensive transfers, the most valuable players, etc.

They have achieved something extremely difficult: being the reference about the players' economic valuation. This valuation has transcended their own web page and it is even used, on many occasions, by clubs (small ones, in general). The search can be made by position or by detailed positions, football players' ages, etc. In this list, we will find players' name, position, nationality, age, the club to which they belong and its value in market. You can check if value keeps, if it has decreased or if, on the contrary, it has increased.



Figure 20: Example of the Transfermarkt web page

The screenshot displays the profile of player #20 Bernardo Silva. Key information includes:
 

- Player Info:** Name: Bernardo Mota Velga de Carvalho e Silva; Date of birth: 10/08/1994; Place of birth: Lisboa; Nationality: Portugal; Position: Medio campo - Mediocentro ofensivo.
- Market Value:** 70,00 mill. €.
- Performance (Premier League 20/21):** 38 partidos posibles, 26 partidos, 2 goles, 5 tarjetas amarillas, 6 asistencias.
- Player Data:** Fichado: 01/07/2017; Contrato hasta: 30/06/2025; Última renovación: 13/03/2019.

Source: print screen of the Transfermarkt web page, <https://www.transfermarkt.es/bernardo-silva/profil/spieler/241641>

Liga Primera División Fichado: Contrato hasta:	League First tier Joined: Contract expires:
Nacimiento/edad: Lugar de nacimiento: Nacionalidad: Altura: Posición: Agente: Selección: Partidos internacionales/goles:	Date of birth/age: Place of birth: Citizenship: Height: Position: Agent: Current international: Caps/goals:
PERFIL RENDIMIENTO VALOR DE MERCADO FICHAJES RUMORES SELECCIÓN NOTICIAS	PROFILE PERFORMANCE MARKET VALUE TRANSFERS RUMOURS NATIONAL TEAM NEWS
Datos de rendimiento Galería 38 partidos posibles Partidos Goles Asistencias Tarjetas amarillas Segundas tarjetas amarillas Tarjetas rojas Cuota XI inicial Minutos jugados Participaciones de gol Datos de rendimiento detallado	Performance data Gallery Statistics Matches Goals Assists Yellow cards Second yellow cards Red cards Starting eleven Minutes Goal participation Detailed performance data
Datos del jugador Nombre en país de origen: Fecha de nacimiento:	Player data Name in home country: Date of birth:



Lugar de nacimiento: Edad: Altura: Nacionalidad: Posición: Medio campo – Mediocentro ofensivo Pie: izquierdo Agente: Club actual: Fichado: Contrato hasta: Última renovación: Proveedor: Redes sociales:	Place of birth: Age: Height: Citizenship: Position: Attacking midfield Foot: left Agent: Current club: Joined: Contract expires: Date of last contract extension: Outfitter: Social networks:
Posición en detalle Posición principal: Mediocentro ofensivo Posición secundaria: Mediocentro – Extremo derecho Valor del mercado Valor del mercado actual: Última revisión: Valor más alto:	Detailed position: Main position: Attacking midfield Other position: Central midfield - Right winger Market value Current market value: Last update: Highest market value:
Ver página de valores detallados	Market value details

7.4. WhoScored: data in this football statistics portal is not so high level as the ones previously mentioned, but it has interesting data about most of football competitions and it shows them on a very visual intuitive web page, which allows playing with the level of detail you want.

It has advanced data, by team or player and a huge amount of data by match for the five big leagues. It even has its own rating for a player's performance by match. Besides, it has all kinds of statistics by countries and by team in the form of a summary (offensive and defensive characteristics, results over time, number of unbeaten matches, goal in zero, matches without winning or without marking, etc.).



Figure 21: Example of the WhoScored web page

## Top Player Statistics

View: Overall Home Away

R	Player	Team	Player Goals/Team Goals		Contribution
1	R. Lewandowski	Bayern	41	99	41%
2	L. Messi	Barcelona	30	85	35%
3	C. Ronaldo	Juventus	29	77	38%
4	A. Silva	E. Frankfurt	28	69	41%
5	K. Mbappé	PSG	27	86	31%
6	E. Haaland	B. Dortmund	27	75	36%
7	R. Lukaku	Inter	24	89	27%
8	H. Kane	Tottenham	23	68	34%
9	K. Benzema	R. Madrid	23	67	34%
10	G. Moreno	Villarreal	23	60	38%

Aggression		Shots per Game				
P. Schiattarella	Benevento	14	1	L. Messi	Barcelona	5.6
W. Faes	Reims	11	2	C. Ronaldo	Juventus	5.1
M. Cassamá	Reims	8	3	R. Lewandowski	Bayern	4.7
S. Savic	Atletico	15	0	Z. Ibrahimovic	AC Milan	4.3
D. Suárez	Getafe	12	1	L. Insigne	Napoli	4.1

Tackles per Game		Dribbles per Game				
L. Abergel	Lorient	3.8		L. Messi	Barcelona	4.5
W. Ndidi	Leicester	3.7		A. Traoré	Wolves	4.1
V. Rongier	Marseille	3.6		A. S-Maximin	Newcastle	3.8
A. Tchouameni	Monaco	3.5		J. Doku	Rennes	3.7
T. Mitchell	Crystal Palace	3.4		R. de Paul	Udinese	3.4

Source: print screen of the Whoscored web page, <https://www.whoscored.com>



## Unit 4.3. Main defensive metrics for measuring high-press

In the world of football analytics, it has historically been more difficult to measure the defensive part of the game than the offensive one. Basically, because while on the attack, the successful performance ends up in an action (to shoot, to assist, to complete a dribble, to score a goal, etc.) and it is generally better to do it more times, in defense it is not always like that. In defense, less action could be better than a lot of action:

A defense that performs more crosses or interceptions is not necessarily better than one that do less; this data measures more the opportunity than the ability. In actions performed in defense, there is much more noise than signal. How can we start reducing this noise? (Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

The defensive goal is to prevent the opponent's team from progressing and creating goal chances and, therefore, many times what is not done is so important—even more—than what is indeed done. For example, an adequate positioning or an effectively coordinated pressing limits the opponent's options and it is therefore more likely that players playing more at the back have to defend frequently in a more active way.

In football it has not been until later that the bases for the statistical defensive analysis have been set up. This happened due to the contribution of data providers like StatsBomb and Opta, together with the demands made by clubs in relation to the improvement in statistics and existing data, and the knowledge by their professionals in the different areas (computer experts, mathematicians, physicist, coaches, analysts, etc.); they are all making and advance that was never seen before.

For that purpose, I am going to focus on the defensive metrics StatsBomb data providers offers, since, in my point of view, it is the provider that best collects data from pressure actions, not only about teams but also about players.

Following Pablo Peña Rodríguez, Tactical Innovation and StatsBomb Business Development's boss, we are going to make comments about each one of the defensive metrics and visualizations that will allow us to measure the team's ability for high-press.

### 4.3.1. PPDA: Passes allowed per defensive action



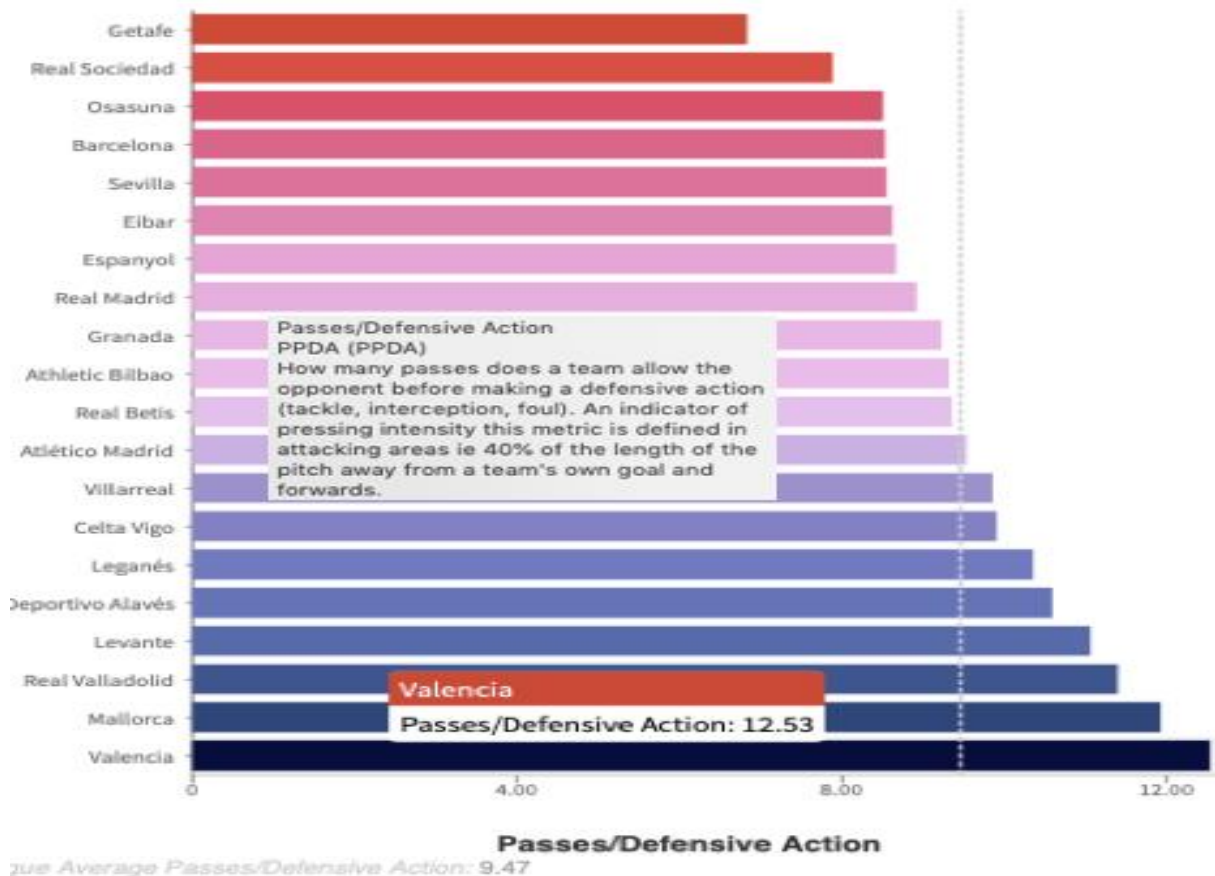
This is one metric created by Colin Trainor in 2014. He was doing work in which he was trying to measure the teams' ability for high-press, trying to answer the coaches' concern about how, when and where the teams performed the defensive actions (tackles, interceptions, fouls, duels).

For that purpose, he established a correlation between the passes allowed to opponents and the defensive actions by the team which was trying to recover the ball, with the following conclusions:

- a) The less passes the opponent plays for each defensive action by my team, the more defensive intensity my team has.
- b) On the contrary, the more passes the opponent plays for each defensive action by my team, the less defensive intensity my team has.

In figure 22, we can see the final La Liga ranking 2019/20 after the 38 matches in the championship. In here, we can see how Getafe is the team that less allows opponents passes for each defensive action on the opponent's field at the initial stage and at the progression, and Valencia is the team that allows the most passes for each defensive action performed on the opponent's half court.

**Figure 22: PPDA on a StatsBomb raking**



Source: [Untitled image about PPDA on a StatsBomb ranking]. (n. d.). Retrieved from [https://lh3.googleusercontent.com/4QRdqGWNi0\\_bJ\\_Zs392Z0mH96LDHnIenpyl145mhA9-nOsTvqJTvUzR4NS2CSroCXb0U=s85](https://lh3.googleusercontent.com/4QRdqGWNi0_bJ_Zs392Z0mH96LDHnIenpyl145mhA9-nOsTvqJTvUzR4NS2CSroCXb0U=s85)

From this visualization we can interpret that Getafe presses a lot better than Valencia in the opponent's half court, so it has a greater capacity of exerting high press for stopping the opponent's associative game.

At a first moment, this study's initial idea was to identify the teams that better pressed on the opponent's half court, since this metric was applied on the 60% of the total field that is more far away from our goal (the opponent's half court + 10 meters on our own half court).

But in StatsBomb they also use it in the whole field for establishing a defensive profile of teams in each of the six areas in the field. Besides, the defensive events are normalized in relation to the opponent's possession in each area.

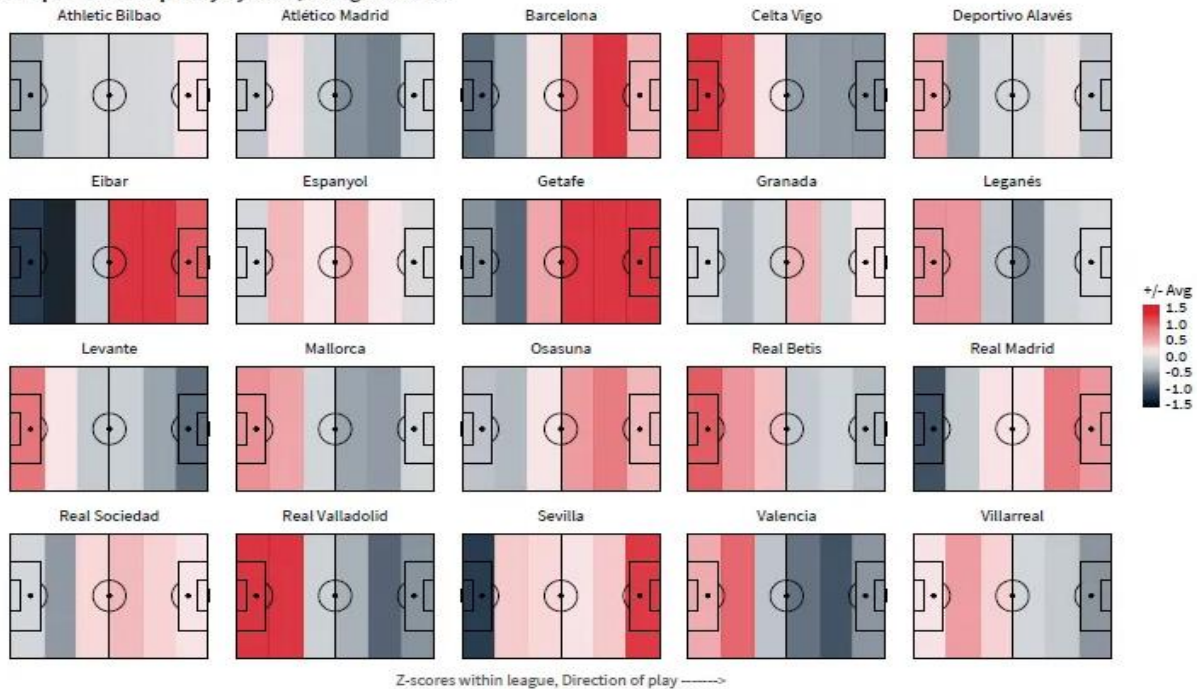
Although the initial idea was to discriminate the teams that showed a higher pressure, the metrics developed and gradually adapted to different areas until being one the mostly used descriptive metrics for showing the stylistic characteristics that define the teams' defense. In StatsBomb, we use it regularly on analyses. (Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

Figure 23: Proportion of defensive actions in StatsBomb

Proportion of Defensive Actions to Opposition Passes



Comparative Frequency by Zone, La Liga 2019-20



Source: Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>

### 4.3.2. Pressure

In my opinion, the key in all advanced metrics for measuring a team’s action and defensive intensity is taking the pressure data into account, which for StatsBomb is the action of harassing the opponent’s ball carrier:

Pressure is the action of attacking the ball carrier or a pass’ receiver in a ratio of five meters in relation to the field area, without executing a tackle, a foul or intercepting the pass (any of these actions can happen subsequently, but it is a different event). (Peña Rodríguez, P. 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

The reason behind collecting this pressure data is reinforced by the belief that many defensive actions that later allow a team to recover the ball possession or to carry out an effective collective pressure are not limited to the actions that occur on the ball (tackles, interceptions, duels, clearances, etc.). But there is a whole series of actions that lead to that, such as harassing for guiding pressure towards a certain area, closing passing lines, accelerating the player’s action, forcing uncontrolled passes or making the opponent’s ball carrier to clear the ball.



At the collective level, this allows us to have a more realistic and complete vision of the tactical mechanisms that teams use in the defensive stage and we can answer the questions that many coaches and analysts make when studying the opponent's defensive abilities.

- a) How can we have a more representative vision of the influence of certain players on defense, beyond directly recovering the ball possession?
- b) Which are the stylistic features that differentiate teams that try to exert high-press from the ones that defend on midfield or low blocks?
- c) In which areas opponents start to exert pressure? To which side do opponents go? Where are they stronger?
- d) How often do they put pressure on certain areas in the field and which players are the most involved in it?, and more. (Peña Rodríguez, 2020).

In order to recover ball possession as close to the opponent's goal as possible during the high-press stage, teams make strategical decisions that mark their defensive profiles and they are clearly identifiable on the visualizations of defensive analyzed data.

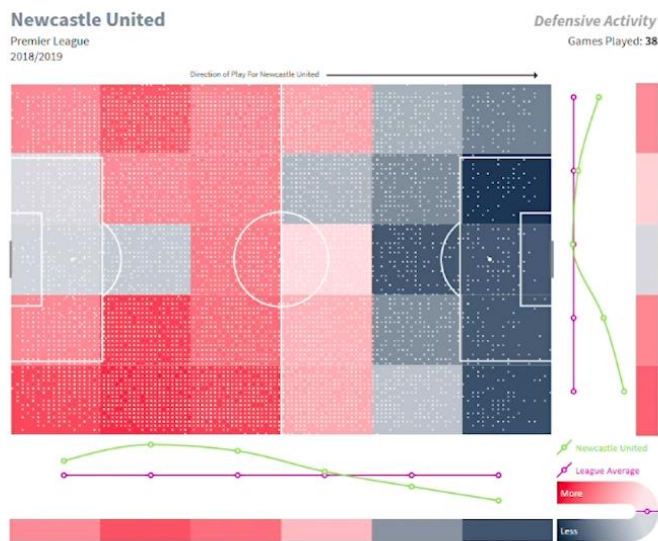
The first pressure data visualization is found on defensive activity maps.

#### Defensive activity map

Following Peña Rodríguez (2020), in the defensive activity map we can see the defensive actions performed by the team in each quadrant. StatsBomb divides the field in five vertical areas and three horizontal areas (for each half court). The red color indicates a bigger defensive activity in relation to League average and the dark gray color indicates less defensive activity in relation to the average.



**Figure 24: Defensive activity map in StatsBomb**



- Rojo= más frecuente
- Azul=Menos frecuente
- Dirección de juego = De izquierda a derecha

**STATSBOMB**

Source: Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>

Rojo=más frecuente	Red=more frequent
Azul=Menos frecuente	Blue=less frequent
Dirección de juego=De izquierda a derecha	Game direction=from left to right

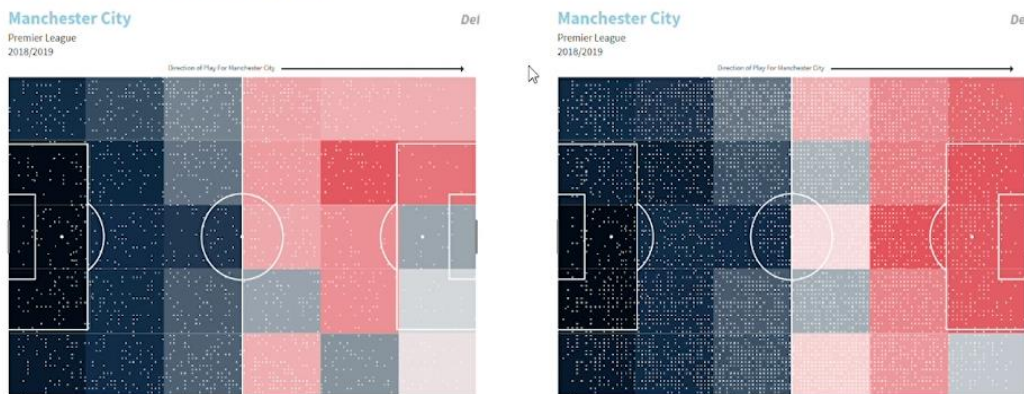
In figure 24, we can see Newcastle United’s information about the 2018/19 season, directed by Rafa Benitez. In the map, we can see how the team is focused on defending on a low block on its own half court, with more activity on the outside areas and with irregular zones in the inside and in the own penalty arch. It also has a defensive predominance in the midfield, positioning in a midfield block, but as we advance towards the opponent’s goal, we can see that it is not a team that exerts defensive high-press (it goes from a light gray to a darker gray).

In figure 25, we have an example of the opposite: the Pep Guardiola’s Manchester City is the example of a team that has a clear defensive profile, exerting high-press on the opponent’s half court.



Figure 25: Comparison of maps with traditional metrics and pressure metrics

Estilos: Presión alta



**Las estadísticas defensivas tradicionales no abarcan todo el espectro**  
**Izquierda:** Man City 2018-19, tackles, interceptaciones, bloqueos, etc  
**Derecha:** Lo mismo pero incluyendo acciones de presión



Source: Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>

Estilos: Presión alta	Styles: High-press
Las estadísticas defensivas tradicionales no abarcan todo el espectro	Traditional defensive statistics do not cover the whole spectrum
Izquierda: Man City 2018-19, tackles, interceptaciones, bloqueos, etc	Left: Man City 2018-19, tackles, interceptions, blocks, etc
Derecha: Lo mismo pero incluyendo acciones de presión	Right: The same but including pressure actions

In the left graphic, we can see how the map would be only with traditional defensive statistics, which only measures the defensive actions in which there is contact on the ball. Instead, on the right map, they add the team actions for pressing and harassing the opponent ball carrier. The difference is clear. The pressure data offers us a more representative image of the game defensive section and, more consciously, of the teams' high-press ability.

The defensive events on traditional data providers have been the tackles, interceptions, duels, clearances, etc. Although this data is useful to a certain extent, it is not a representative vision of the complete spectrum of actions that have an impact at the team's defensive level, delimiting, that way, the ability of extracting relevant conclusions...StatsBomb data provides a new event that changes the way of making the defensive analysis: pressure actions. (Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).



From this confirmation, we can draw the conclusion that pressure data offers us a better and wider vision about how teams press on the opponent's half court, in comparison with the PPDA metric.

Other visualizations we can use at the collective level are the teams' rankings.

### Rankings

With defensive data on partial pressure (on a certain zone in the field) or total pressure (in the whole field), we can differentiate the various defensive stages.

In figure 26, we can see a ranking taken from the Fbref web site, in which we can see, in red, the pressures exerted by each team in each third of the field.

- Def. 3.<sup>rd</sup> = low block.
- Mid. 3.<sup>rd</sup> = midfield block.
- Att. 3.<sup>rd</sup> = high block.

In yellow, in the left ranking, we can see that Sevilla is the last one in number of total pressing actions exerted during La Liga 2020/21.

**Figure 26: Rankings of the total pressing actions and by thirds in the field**

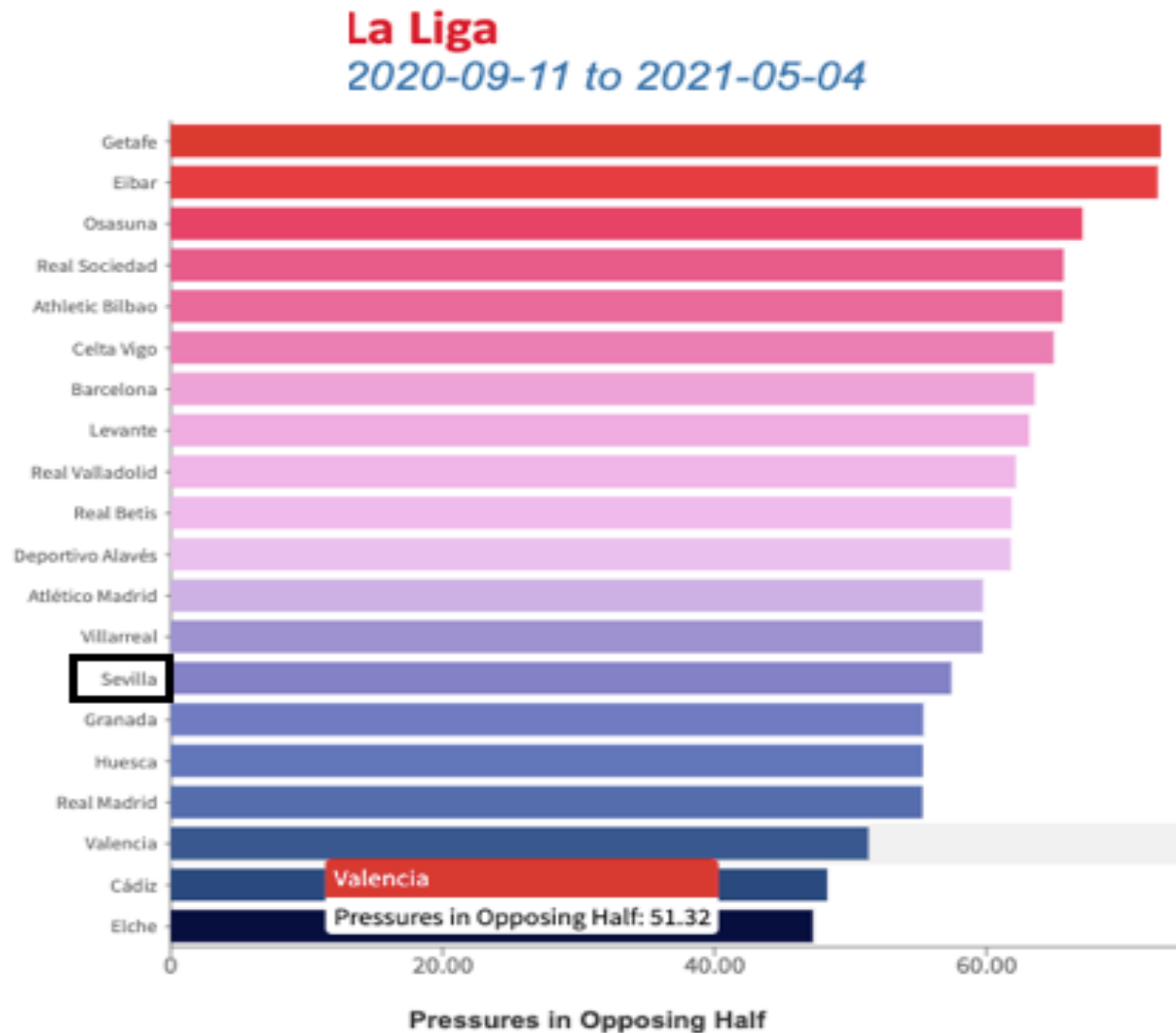
Squad Stats		Number of times applying pressure to opposing player who is receiving, carrying or releasing the ball								
Squad	Press ▼	Succ	%	Def 3rd	Mid 3rd	Att 3rd	Blocks	Sh	ShSv	Pass
Celta	6085	1689	27.8	1602	3055	1428	557	68	4	489
Colta Vigo	5997	1701	28.4	2063	2850	1084	610	97	1	513
Valledolid	5839	1624	27.8	2051	2600	1188	601	117	2	484
Huesca	5451	1578	28.9	1951	2462	1038	551	103	1	448
Alavés	5447	1510	27.7	1752	2541	1154	568	105	4	463
Levante	5363	1499	28.0	1697	2350	1316	538	108	3	430
Osasuna	5329	1550	29.1	1558	2487	1284	557	109	2	448
Cádiz	5264	1433	27.2	2094	2304	866	612	118	4	494
Betis	5263	1514	28.8	1619	2495	1149	531	84	2	447
Athletic Club	5250	1451	27.6	1494	2380	1376	553	86	2	467
Eibar	5168	1455	28.2	2068	2228	872	553	111	2	442
Atlético Madrid	5124	1540	30.1	1665	2328	1131	579	93	0	486
Barcelona	5075	1546	30.5	1404	2330	1341	519	82	0	437
Real Sociedad	5070	1487	29.3	1423	2340	1307	497	104	2	393
Granada	5027	1358	27.0	1616	2345	1066	611	142	5	469
Eibar	5015	1602	31.9	1272	2251	1492	553	75	2	478
Villarreal	4994	1425	28.5	1587	2226	1181	578	117	3	461
Real Madrid	4880	1425	29.2	1562	2175	1143	535	100	4	435
Valencia	4855	1305	26.9	1794	2067	994	566	137	2	429
Sevilla	4716	1477	31.3	1392	2199	1125	503	95	3	408

Source: own adaptation of Fbref, n. d., <https://fbref.com/en/comps/12/La-Liga-Stats>



Another visualization, in this case by StatsBomb, can be seen on figure 27. We can observe in it a ranking that measures the ability for high-press on the opponent's field of La Liga teams 2020/21. In this ranking we can see Sevilla on the 14th position.

**Figure 27: Ranking of pressure on the opponent's field by StatsBomb**



Source: [Untitled image about ranking for pressure on the opponent's field]. (n. d.). Retrieved from <https://lh3.googleusercontent.com/CFilwr8zUC-i9Gk-kyX5x-NEZRYcgj2QH4WoN63PAAtWrea3qbb1Avq4JGoepUMUvt1feuA=s85>

### Dispersion graphs

Another visualization that is mostly used is the dispersion graph, which allows us to study the relations between two game variables. Thanks to them, we can position a team or a player in relation to the rest, to see if they are above or below the average and more.

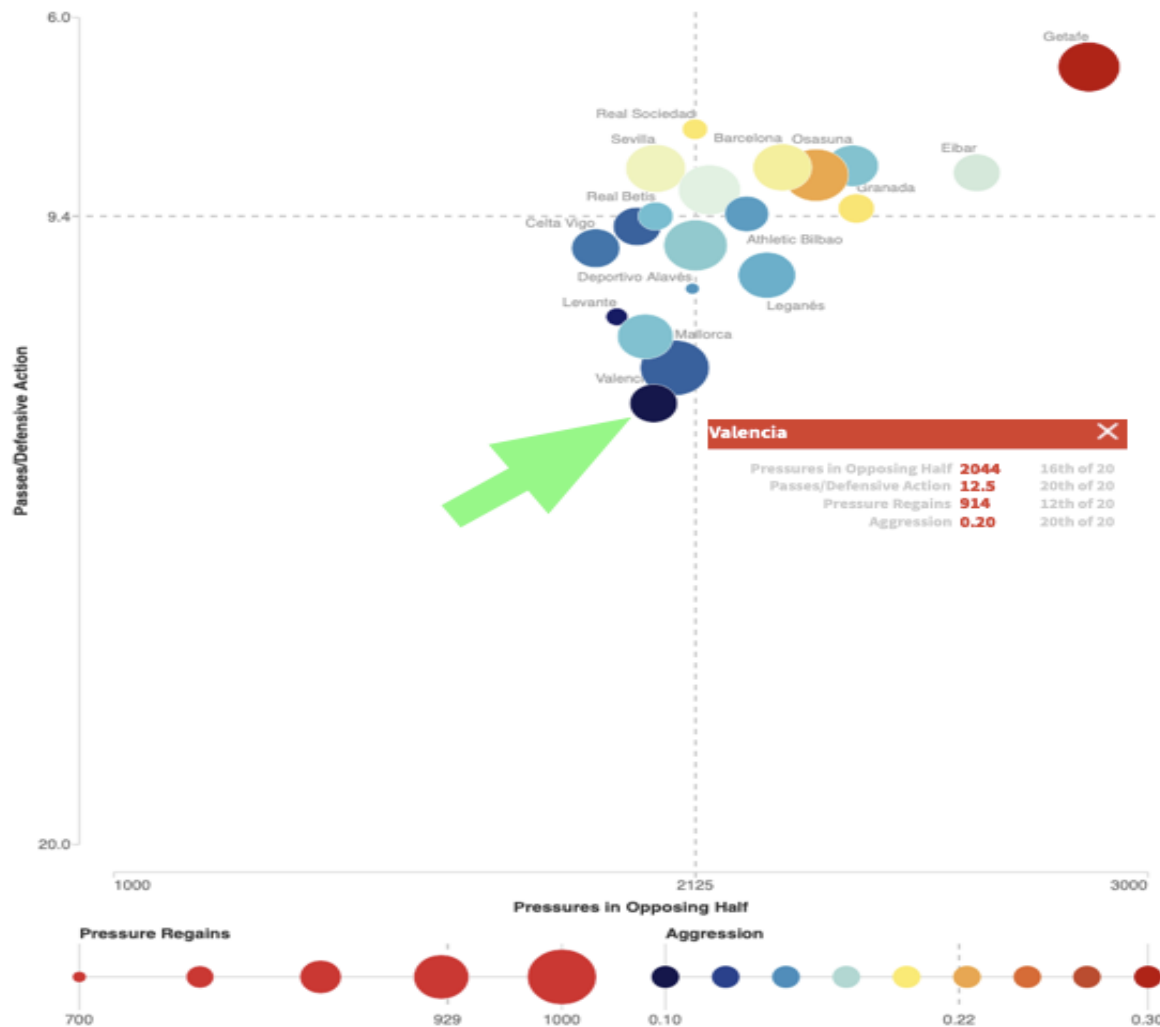
In figure 28, we can see an example of how can we measure one team's pressing ability in only one graph. We can observe in it until 4 pressing variables. The position of Valencia



in 2019/20 season was like this in relation to PPDA and pressure on the opponent's half court. And besides in the StatsBomb IQ platform they give us the possibility of adding 2 variables:

- a) The size of the bubble: it represents the pressure regains.
- b) The bubble color represents the aggression data, the team's aggression level.

**Figure 28: Team's pressure ability on a 4 variables' dispersion graph**



Source: [Untitled image about a team's pressure ability] (n. d.). Retrieved from [https://lh3.googleusercontent.com/dilrueUqUvtmDf4VyI\\_rTuVSwgSi\\_81ZkZbbaifEcEz6YIAGxBIQR04xmOKSxsmbllqUPw=s85](https://lh3.googleusercontent.com/dilrueUqUvtmDf4VyI_rTuVSwgSi_81ZkZbbaifEcEz6YIAGxBIQR04xmOKSxsmbllqUPw=s85)



### 4.3.3. The counterpressure

Another measure we should take into account for evaluating the ability for pressing on the opponent's half court is counterpressure (or gegenpressing), that is to say, pressure exerted during the five seconds after losing ball possession.

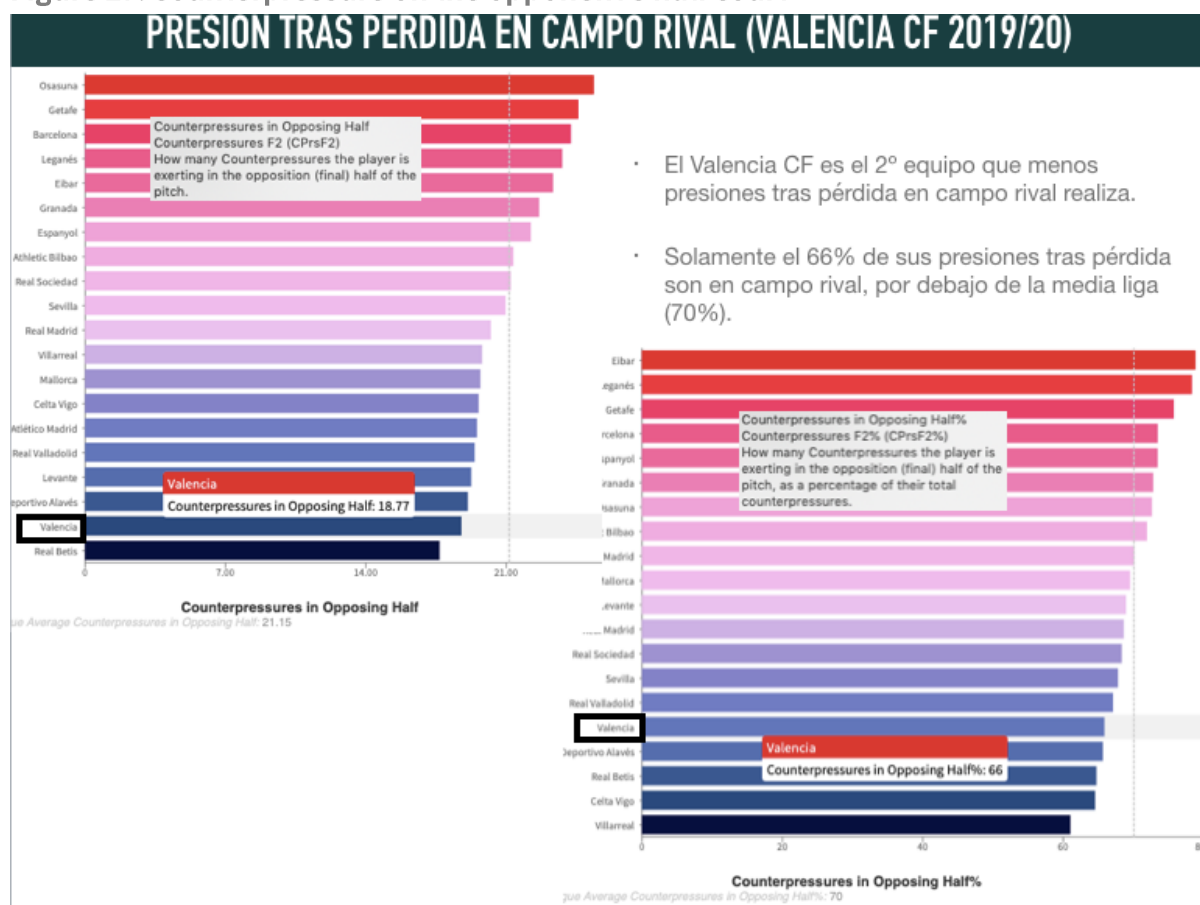
This counterpressure happens when a team presses the opponent during the five seconds after losing ball possession. It generally applies on the opponent's half court, where the loss of ball possession takes place and the team is closely enough, so that close players can perform this counterpressure.

The initial counterpressure's aim is to interrupt the opponent's counterattack during the attack-defense transition stage. It will therefore give us relevant information about which players have more influence in the attack-defense transition moments.

In figure 29, we observe how we can measure pressure after loss of ball possession on the opponents' half court, with the example of Valencia in 2019/20 season:

- a) Evaluating the average counterpressure on the opponent's half court by match.
- b) Evaluating the percentage of the total counterpressure on the opponent's half court.

Figure 29: Counterpressure on the opponent's half court



Source: [Untitled image about the counterpressure] (n. d.). Retrieved from [https://lh3.googleusercontent.com/C2rHhOr2Wm9Vqzk4xFYKKC6604tLENRDvJ7dDv1wYjnkHuYR\\_cmGkCmqNiv-TDYcXr7-HkM=s114](https://lh3.googleusercontent.com/C2rHhOr2Wm9Vqzk4xFYKKC6604tLENRDvJ7dDv1wYjnkHuYR_cmGkCmqNiv-TDYcXr7-HkM=s114)

PRESIÓN TRAS PÉRDIDA EN CAMPO RIVAL (VALENCIA CF 2019/20)	PRESSURE AFTER LOSS OF BALL POSSESSION ON THE OPPONENTS' HALF COURT (VALENCIA CF 2019/20)
-El Valencia CF es el 2º equipo que menos presiones tras pérdida en campo rival realiza.	-Valencia CF is the 2nd team that presses the less after loosing ball possession on the opponents' field.
-Solamente el 66% de sus presiones tras pérdida son en campo rival, por debajo de la media liga (70%).	-Only in 66% of their pressures after loosing ball possession are on the opponent's half court, under the average in the league (70%).

Until now, he has talked about data for measuring the teams' pressure abilities, but, how do we measure players' pressure abilities individually?

#### 4.3.4. Pressure at the individual level: pressure regain

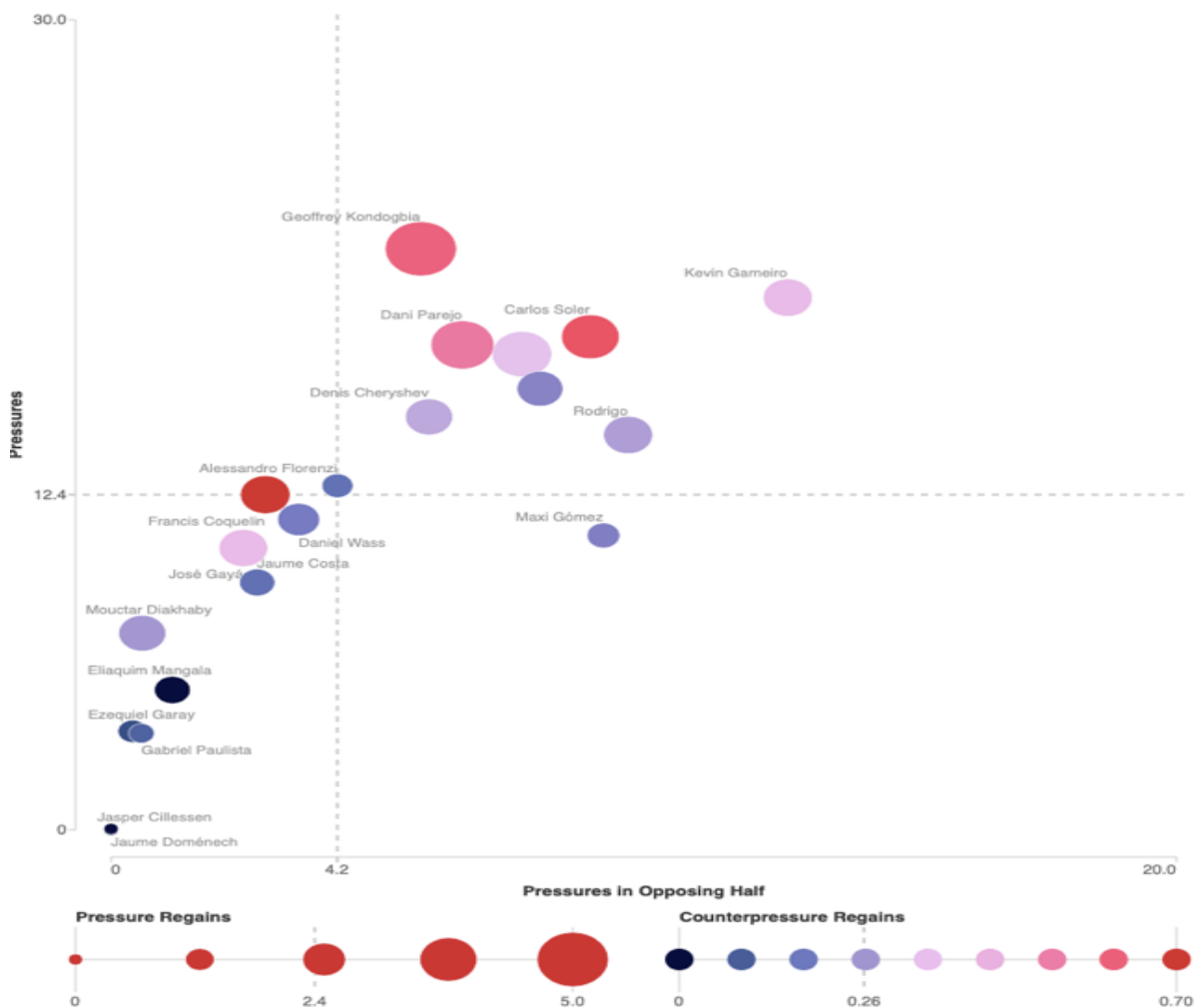
At the individual level, the pressure exerted data is especially relevant for analysts for offensive players and midfielders (mainly if we talk about high-press). Even though, it can also be applied for defensive players, when we make reference to pressure on the whole playing field.

We are not only going to collect the number of times the players press, but we are also going to collect data about the level of influence they have on their teammates recovery and, in summary, on their team.

And we are going to achieve this by using the pressure regains metric. If the team recovers the ball in the five seconds after a player's pressure action, the player will be assigned a pressure regain.



Figure 30: The 4 variables that measure a player's pressure ability



Source: [Untitled image about the 4 variables that measure a player's pressure ability (n. d.). Retrieved from <https://lh3.googleusercontent.com/D56U0KRjGbQ4IIY-E73kKoKHqs-wrtDfOtL6UhOQcERi7P8NwOIZpjYge1bGLg6QEML-t7M=s85>

With this metric, it will not only be about evaluating offensive players and midfielders defensive performance with actions on the ball (tackles or interceptions), but it will be about giving value to defensive actions they perform for the team to recover the ball.

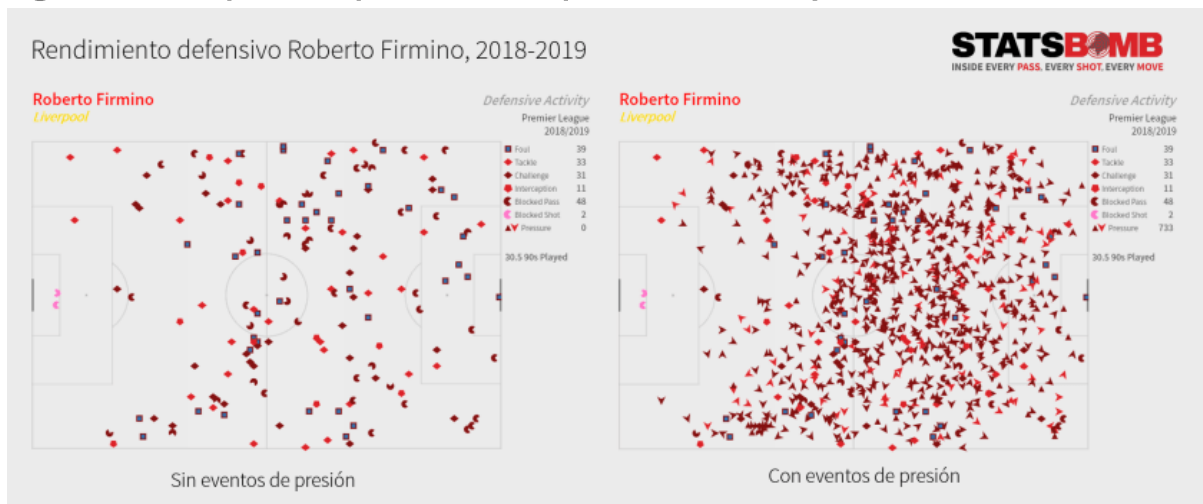
That is to say, in general, an offensive player does not press on his own for recovering the ball, but for blocking the concrete passing lines, orienting the opponent's build-up play towards concrete zones (on the outside, in general), forcing the opponents to perform uncontrolled actions, to make mistakes or to clear the ball. (Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).



In other type of visualizations, like in pressure maps, we can see each pressing action performed by the player in question and, in a more visual way, we can observe the pressing field that is covered by his/her defensive performance.

That way, in figure 31, with the aim of seeing the importance of these pressure events, we can see the comparison of the defensive performance vision by Roberto Firmino (2018-2019), excluding pressure actions (left graphic) and including them (right graphic).

**Figure 31: Example of map of individual pressure events by StatsBomb**



Source: Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>

Rendimiento defensivo Roberto Firmino, 2018-2019	Defensive performance Roberto Firmino, 2018-2019
--	--

#### 4.3.5. Advantages and additional usage for pressure events

Besides the strictly defensive analysis about high-press, the pressing events allow us to have information about the team's offensive stage like, for example, knowing how the pressure exerted by an opponent affects a certain team or player. That is to say, a player's ability to perform a pass can be measured when is under the opponent's pressure.

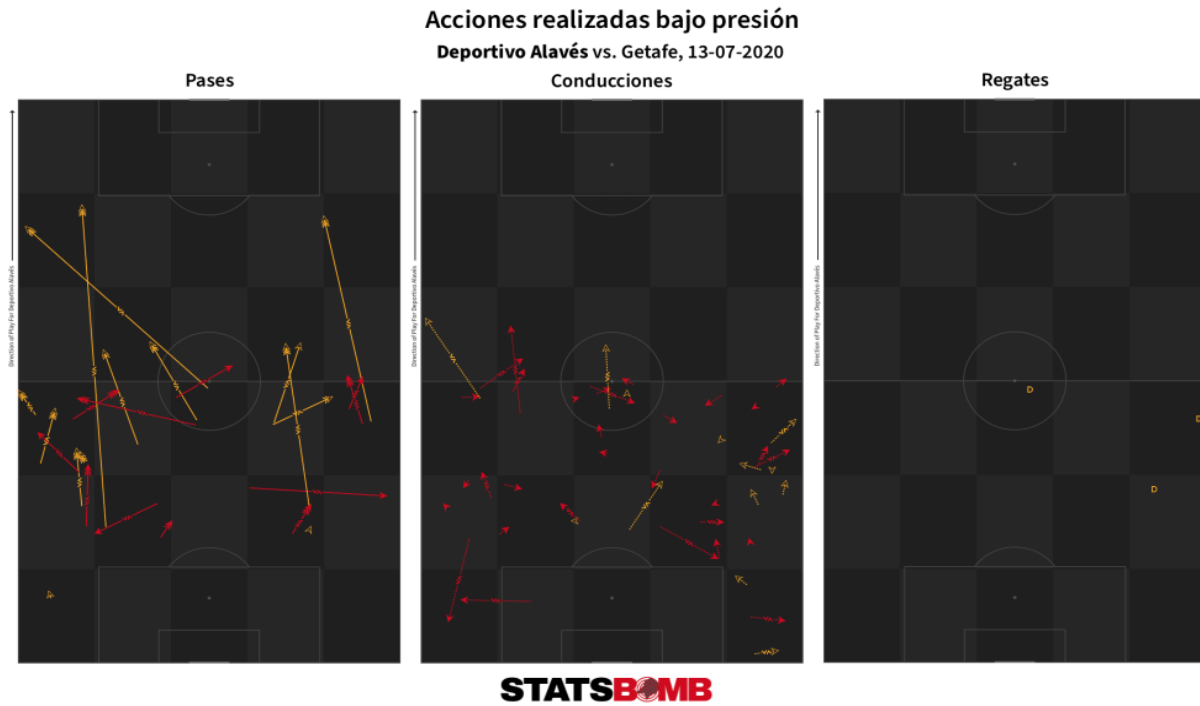
We can achieve this with the AUP or "Actions under pressure" or the "Passes under pressure" metric. For example, on figures 32 and 33, we can see that when facing a strong and suffocating high-press from Getafe, teams like Alavés and the FC Barcelona had a completely different reaction in the opponent's half court.

We can see how the Alavés had a lot of problems for combining and connecting passes on the opponent's half court.



It suffered a significant fall in relation to its average in the season, not only in the pass success percentage, but also in the pass success percentage under pressure. They were not successful in the three regains they tried (red=successful action; yellow=failed action). (StatsBomb, February 11th, 2021).

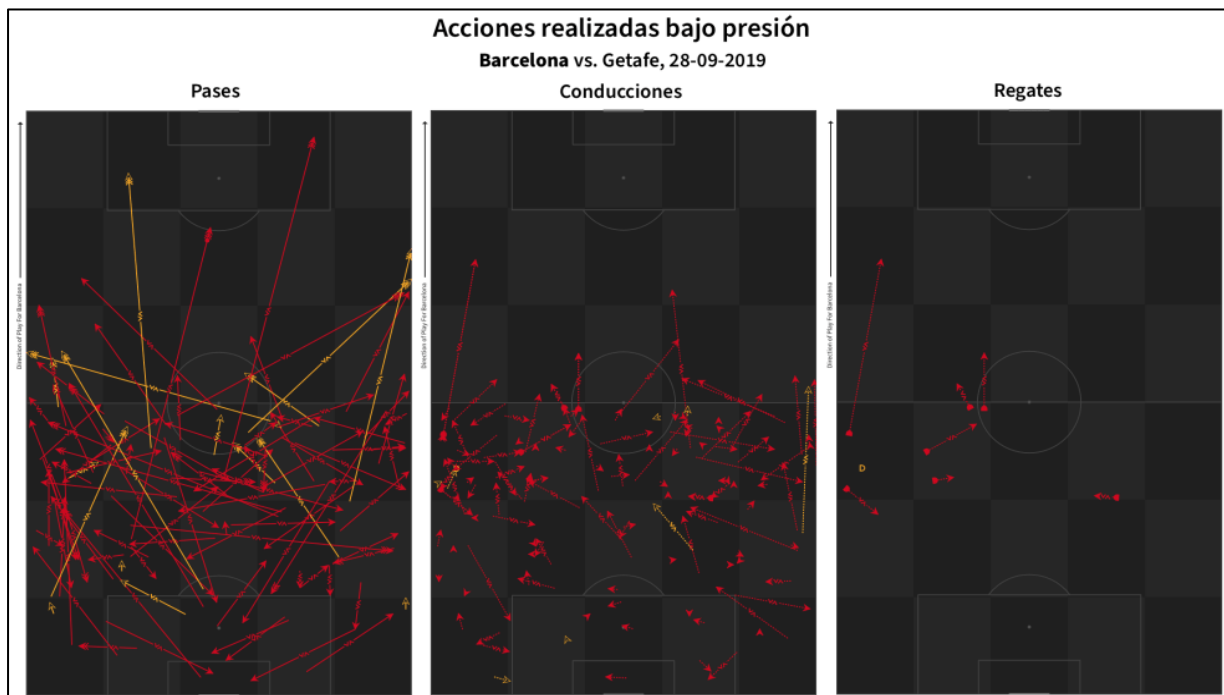
**Figure 32: Maps of actions performed under pressure (StatsBomb)**



Source: StatsBomb, February 11th 2021, <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-equipos-de-la-liga/>

Acciones realizadas bajo presión	Actions performed under pressure
Pases	Passes
Conducciones	Dribbles
Regates	Faints

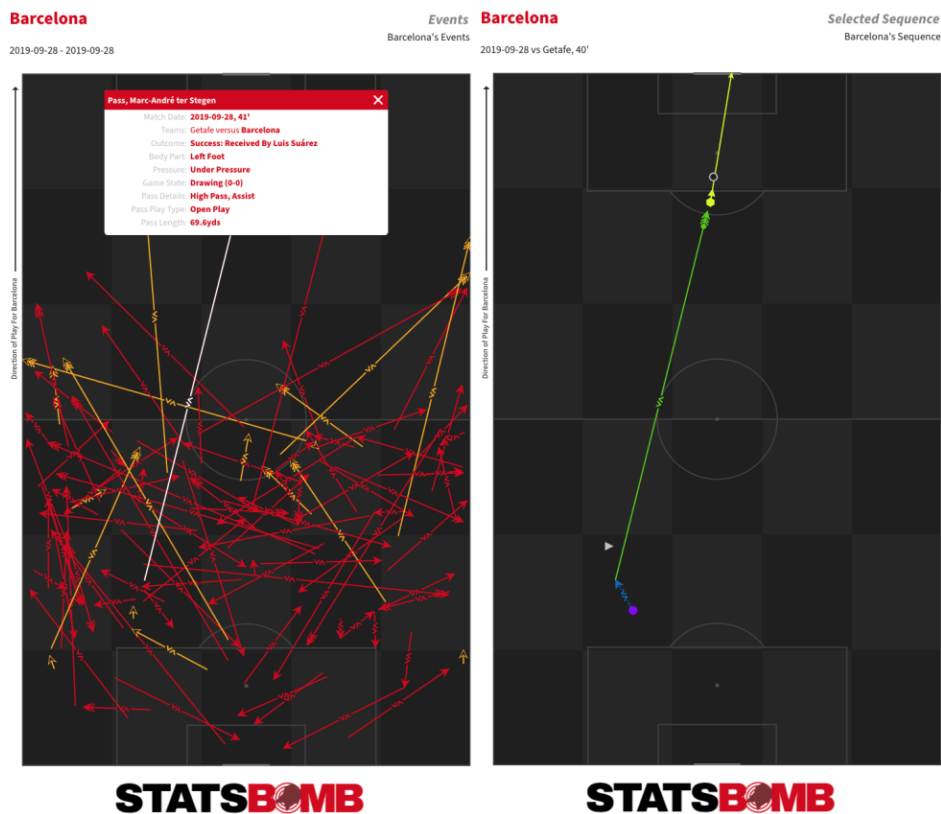
Figure 33: Maps of actions performed under pressure (StatsBomb)



Source: StatsBomb, February 11th 2021, <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-equipos-de-la-liga/>

Acciones realizadas bajo presión	Actions performed under pressure
Pases	Passes
Conducciones	Dribbles
Regates	Feints

Figure 34: Maps of actions performed under pressure (StatsBomb)

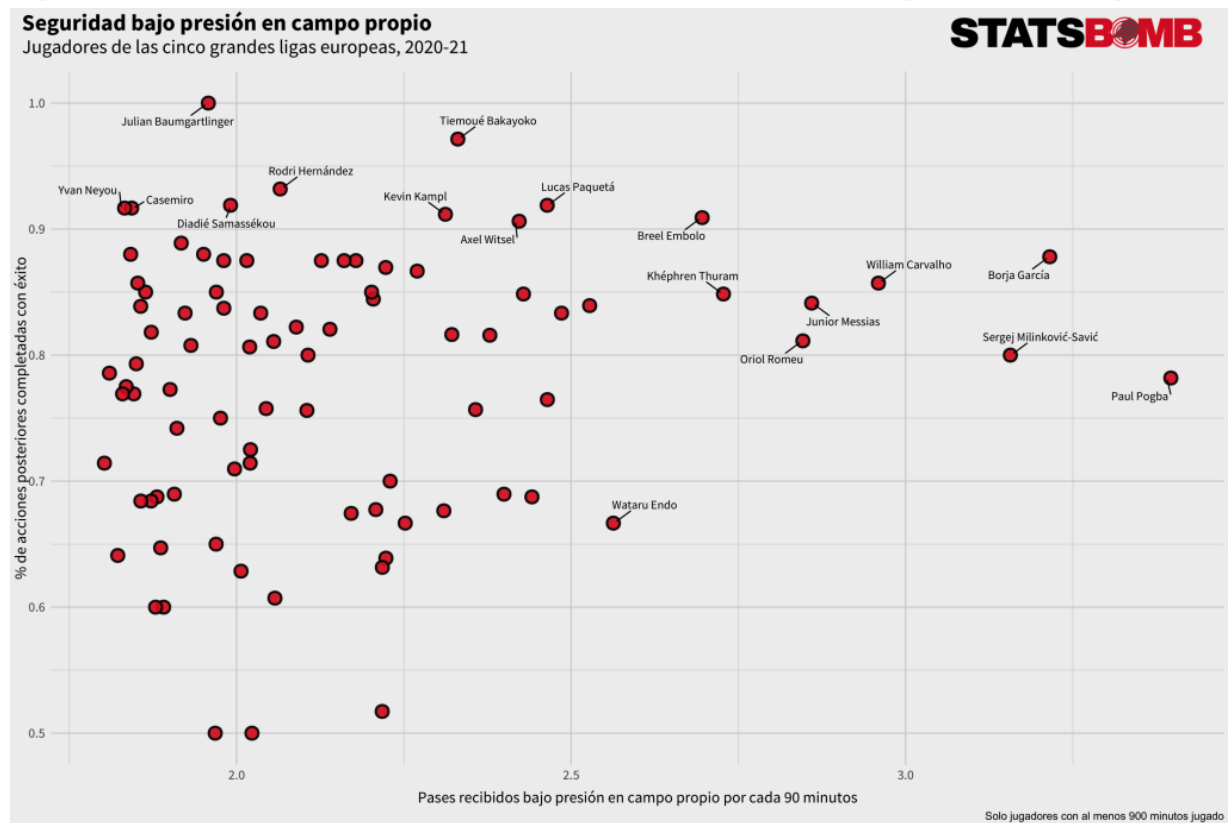


Source: StatsBomb, February 11th 2021, <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-equipos-de-la-liga/>

On the other hand, we can see on figure 34 that Barcelona did find solutions for counteracting the exerted high-press by Getafe, they being able to connect among teammates on the opponent's half court from their own half. They could even score a goal in one of these connections thanks to the goalkeeper Ter Stegen, who assisted with a pass into space from his own half to Luis Suarez to score.

Another offensive metric derived from pressure events collected by StatsBomb is the one for the way in which pass receivers cope with the opponent's pressure when they receive the ball under pressure on their own half court. That way, we can measure their ability for a build-up play and for giving continuity to the game when they receive the ball under pressure on their own half court. We can observe it on this dispersion graph (figure 35):

Figure 35: Pass reception under pressure of players in the five big European leagues



Source: StatsBomb, February 18th 2021, <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-jugadores-de-las-cinco-grandes-ligas-europeas/>

Seguridad bajo presión en campo propio Jugadores de las cinco grandes ligas europeas 2020-21	Conviction under pressure on own half court Players of the five big European leagues 2020-21
% de acciones posteriores completadas con éxito	% of successful after actions
Pases recibidos bajo presión en campo propio por cada 90 minutos	Passes received under pressure on own half court for every 90 minutes
Solo jugadores con al menos 900 minutos jugado	Only players that played at least 900 minutes

But, which type of conclusions can we draw from this dispersion graph? Below there are some references to the ability that some of these players have for giving continuity to the game when they receive under pressure:

Some names that are worthy of mention: Paul Pogba, from Manchester United, receives more passes under pressure on his own half that any other player, although he loses the ball a bit more frequently than other players like Borja García in Huesca or William Carvalho in Real Betis, who also receive a lot of passes like that. Julian Baumgartlinger in Bayern Leverkusen receives a smaller volume of these passes (1.96 by 90), but he



has never not lost the ball in the current season. (StatsBomb, February 18th 2021, <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-jugadores-de-las-cinco-grandes-ligas-europeas/>).



## Unit 4.4. Keys to a good data interpretation

On our presentations, we make use of visualizations that providers offer us or of the visualizations that we have created. But the most important thing is the way we interpret data after seeing them on these visualizations.

To different people's eyes and with different points of view, one same visualization can have different interpretations, being one as acceptable as the other one. What will give them more or less acceptability is to think about the reason why we make that interpretation, or in other words, the causes or origins of that interpretation that usually is derived from an accumulation of game facts, actions or situations that are repeated during matches.

For example, on figure 36, on the left ranking, we can see highlighted in yellow that Sevilla is last in relation to the number of total exerted pressure in La Liga in season 2020/21, but instead, we can see it on 2nd position as the team with the best successful pressure percentage (with recovery).

If we only see the left ranking (the one for exerted pressures) we can interpret that Sevilla is one of the worst teams in pressure, but we have to contextualize data. We can do it as follows:

- a) By exploring data to the maximum for a better understanding: we should not be swayed by hurries or first impressions, but we should put an effort to understand the meaning of the values that are being registered (in this case, the value taken is the number of exerted pressures, without taking into account if the ball is recovered or not). Sevilla can perform a smaller number of pressures simply because it has bigger possession than most teams.
- b) By contrasting data with other values related to the data in question: this is going to give us an adequate context for valuing and giving that data a more acceptable interpretation. For example, with the right ranking visualization, we are adding context to the left ranking. It tells us that Sevilla is the team that presses the less, but the second team which recovers balls the most, after exerting that pressure. Therefore, we are talking about one of the most efficient teams when exerting pressure.



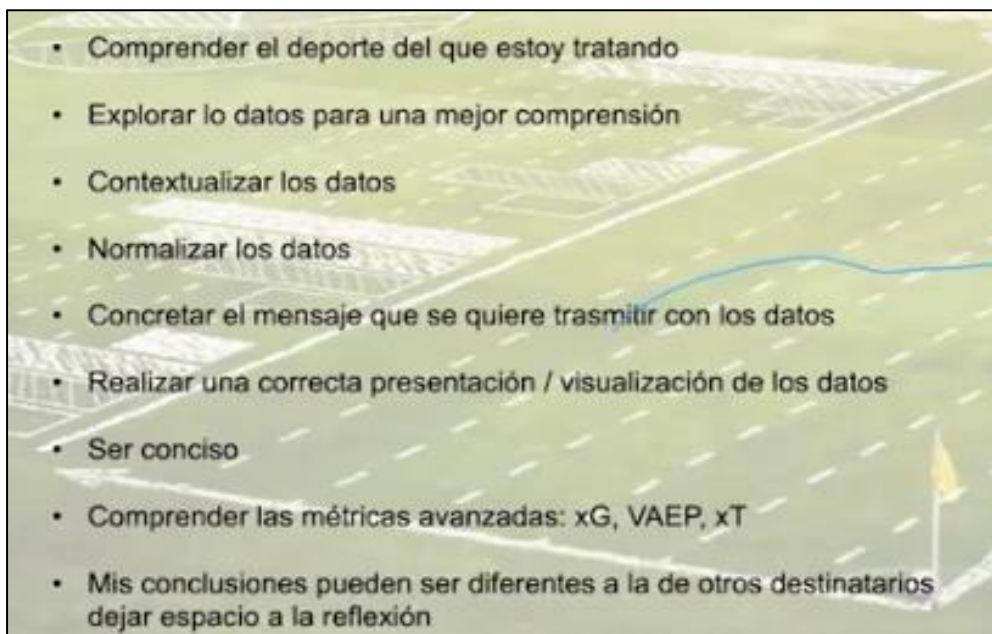
Figure 36: Rankings of the total pressing actions and by thirds in the field on the Fbref web site

Squad Stats										Successful Pressure Percentage														
Number of times applying pressure to opposing player who is receiving, carrying or releasing the ball										Percentage of time the squad gained possession within five seconds of applying pressure Minimum 6.44 pressures per squad game to qualify as a leader														
Squad	Press	Succ	%	Def 3rd	Mid 3rd	Att 3rd	Blocks	Sh	ShSv	Pass	Squad	Press	Succ	%	Def 3rd	Mid 3rd	Att 3rd	Blocks	Sh	ShSv	Pass	Int	Tkl	
Getafe	6085	1689	27.8	1602	3055	1428	557	68	4	489	Eibar	5015	1602	31.9	1272	2251	1492	553	75	2	478	390		
Celta Vigo	5997	1701	28.4	2063	2850	1084	610	97	1	513	Sevilla	4716	1477	31.3	1392	2199	1125	503	95	3	408	334		
Valladolid	5839	1624	27.8	2051	2600	1188	601	117	2	484	Barcelona	5075	1546	30.5	1404	2330	1341	519	82	0	437	354		
Huesca	5451	1578	28.9	1951	2462	1038	551	103	1	448	Atlético Madrid	5124	1540	30.1	1665	2328	1131	579	93	0	486	395		
Alavés	5447	1510	27.7	1752	2541	1154	568	105	4	463	Real Sociedad	5070	1487	29.3	1423	2340	1307	497	104	2	393	337		
Levante	5363	1499	28.0	1697	2350	1316	538	108	3	430	Real Madrid	4880	1425	29.2	1562	2175	1143	535	100	4	435	406		
Osasuna	5329	1550	29.1	1558	2487	1284	557	109	2	448	Osasuna	5329	1550	29.1	1558	2487	1284	557	109	2	448	394		
Cádiz	5264	1433	27.2	2094	2304	866	612	118	4	494	Huesca	5451	1578	28.9	1951	2462	1038	551	103	1	448	348		
Betis	5263	1514	28.8	1619	2495	1149	531	84	2	447	Betis	5263	1514	28.8	1619	2495	1149	531	84	2	447	432		
Athletic Club	5250	1451	27.6	1494	2380	1376	553	86	2	467	Villarreal	4994	1425	28.5	1587	2226	1181	578	117	3	461	354		
Eibar	5168	1455	28.2	2068	2228	872	553	111	2	442	Celta Vigo	5997	1701	28.4	2063	2850	1084	610	97	1	513	375		
Atlético Madrid	5124	1540	30.1	1665	2328	1131	579	93	0	486	Eibar	5168	1455	28.2	2068	2228	872	553	111	2	442	388		
Barcelona	5075	1546	30.5	1404	2330	1341	519	82	0	437	Levante	5363	1499	28.0	1697	2350	1316	538	108	3	430	445		
Real Sociedad	5070	1487	29.3	1423	2340	1307	497	104	2	393	Getafe	6085	1689	27.8	1602	3055	1428	557	68	4	489	420		
Granada	5027	1358	27.0	1616	2345	1066	611	142	5	469	Valladolid	5839	1624	27.8	2051	2600	1188	601	117	2	484	390		
Eibar	5015	1602	31.9	1272	2251	1492	553	75	2	478	Alavés	5447	1510	27.7	1752	2541	1154	568	105	4	463	432		
Villarreal	4994	1425	28.5	1587	2226	1181	578	117	3	461	Athletic Club	5250	1451	27.6	1494	2380	1376	553	86	2	467	348		
Real Madrid	4880	1425	29.2	1562	2175	1143	535	100	4	435	Cádiz	5264	1433	27.2	2094	2304	866	612	118	4	494	411		
Valencia	4855	1305	26.9	1794	2067	994	566	137	2	429	Granada	5027	1358	27.0	1616	2345	1066	611	142	5	469	406		
Sevilla	4716	1477	31.3	1392	2199	1125	503	95	3	408	Valencia	4855	1305	26.9	1794	2067	994	566	137	2	429	321		

Source: own adaptation of the Fbref.com web site, n. d., <https://fbref.com/en/comps/12/La-Liga-Stats>

In figure 37, Fredi Martín provide us with a series of keys—with which we completely agree—for treating and giving data a correct interpretation.

Figure 37: Keys to a correct interpretation of data



Source: [Untitled image about the keys for a correct interpretation of data]. (n. d.). Retrieved from [https://lh3.googleusercontent.com/D30igKfmY4FIR3Zv\\_0IF5HEm-PM2akwwbRIOR3d9me5JMY8UKXVB9U\\_v4W0rmnWZLiyUxv4=s109](https://lh3.googleusercontent.com/D30igKfmY4FIR3Zv_0IF5HEm-PM2akwwbRIOR3d9me5JMY8UKXVB9U_v4W0rmnWZLiyUxv4=s109)



- Comprender el deporte del que estoy tratando	- Understand the sports I am talking about
- Explorar lo datos para una mejor comprensión	- Explore data for a better understanding
- Contextualizar los datos	- Contextualize data
- Normalizar los datos	- Normalize data
- Concretar el mensaje que se quiere transmitir con los datos	- Transmit a concrete the message someone wants to give with data
- Realizar una correcta presentación / visualización de los datos	- Make a correct data presentation / visualization
-Ser conciso	-Be concise
- Comprender las métricas avanzadas: xG, VAEP, xT	- Understand advanced metrics: xG, VAEP, xT
- Mis conclusiones pueden ser diferentes a la de otros destinatarios dejar espacio a la reflexión	- My conclusions can be different to the others. Leave room for thinking about things

From all these keys, we are going to make comments about some of them. We can conclude that, for offering a good interpretation of data on our presentations, we should do as follows:

- a) Understand football as a sport in its whole dimension: the more we know, the more we can give data more context and, therefore, a better interpretation.
- b) Explore data to the maximum: for having a wider vision of the spectrum to study (pressures and others). This will give us the possibility of having a better understanding of data and, therefore, we will be able to understand more.
- c) Contextualize data: there is a need to search for and find the reasons and the causes of the interpretation we are giving to data. For example, data show that Marcos Llorente has more goals and assistances in Atlético de Madrid than when he played in Real Madrid. For giving context to this or that data, we have to make the adequate questions and hypotheses: ¿Why when playing in Real Madrid Llorente did not walk the opponent's half court and did not mark or assist as much as he does in Atlético de Madrid following Cholo Simeone's orders? One reason could be because the current coach asks him different things (that he holds more balance, that he does not engage so much in attacks because the team already has other players that could end the play, etc.) or maybe because the player shows a different mentality and he has evolved offensively for having more presence in the area. There might be many answers to the previous question. The purpose is to compare them and take them into account for having an interpretation that is as close as possible to reality.
- d) Normalize data: when we observe visualizations on defensive data about some teams or players, such as interceptions, tackles, duels, etc., we should take into account that, in order to make a good interpretation of this data, we need to



normalize it. That is to say, adjust these defensive metrics in relation to possession. The teams' game style affects data that are registered on absolute values.

Players can perform defensive actions only when their team is not in possession of the ball. Therefore, in relation to the time the team is on possession, it will have more or less opportunities of adding defensive actions. This causes the teams with high volumes of possession have smaller numbers of defensive actions and, therefore, the absolute numbers for tackles, interceptions (or pressures) have more significance than the desirable. (Peña Rodríguez, P. 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

The adjustment in relation to possession is a practical solution for standardize and normalize values, so that they can be used for comparisons among different players or teams, correcting aspects that are external to their game ability or style respectively.

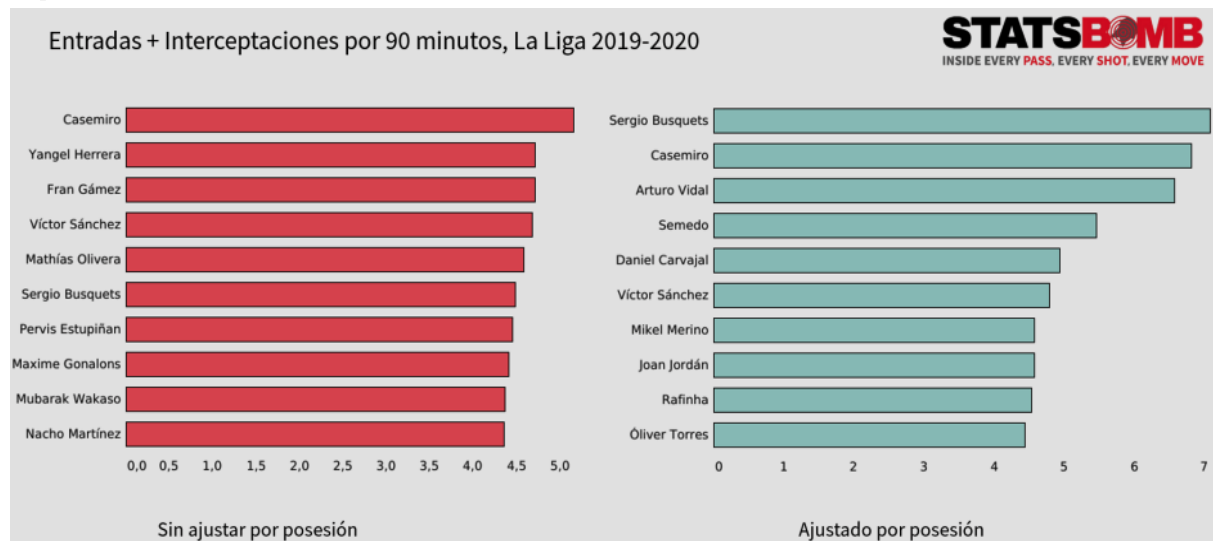
Nevertheless, it is important to highlight that, once possession is adjusted, there is no longer a veridical correspondence between the metric value and the real actions we can watch on videos. That is why, absolute numbers are still useful when evaluating players' performance (for example, for defining limits). (Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>).

On figure 38, we can see a ranking for tackles and interceptions (x90 minutes) of La Liga players. On the left, we can find absolute data, while on the right, we find the data adjusted in relation to their teams' possession.

As we can observe, players like Casemiro not even appear in the absolute values' ranking, simply by the fact that they have less opportunities of adding defensive actions, since they hold the ball a lot more than players in other teams who have less possession and are forced to defend a lot more, which in turn means more defensive events.



Figure 38: Data normalization by adjustment by possession



Source: Peña Rodríguez, 2020, <https://StatsBomb.com/es/2020/03/metricas-defensivas/>

Entradas + Interceptaciones por 90 minutos, La Liga 2019-2020	Tackles+Interceptions by 90 minutes, La Liga 2019-2020
Sin ajustar por posesión	Without adjusting by possession
Ajustado por posesión	Adjusted to possession

- e) To make a good data presentation/visualization, as well as to write a concrete and concise message: we already know that simplicity in the visualizations we present, as well as brevity and clarity in our opinions, will make it easier for the receiver to understand.
- f) To leave space for thinking, knowing that the rest can have different opinions: the subjectivity that football takes in its DNA is also reflected on the data it generates (if one same game action can be judged differently by each interlocutor who see that action, how that is not going to happen with the data the game generates itself?). For that reason, we should be respectful with the rest's opinions, as long as those opinions have a theoretical basis that was worked and has results like ours, after we have carried out all the previously mentioned processes and having applied all these keys for a good interpretation.

## Conclusions

It is clear that football has evolved throughout its history and it will always do. But in times of application of big data and artificial intelligence (AI), football has started a modernization process that forces coaches, assistants, sports directors and all the ones who want to be part of it to be trained for being up to date and understand it. This time it



becomes more important for performance improvement processes in football players and teams' game.

Having said this, it is important to have notions about football, not only in its technical-tactical dimension, but also in the training process and in the analysis of the data it generates. Data is being increasingly used by clubs and coaching staffs, who give room to professionals for them to make a useful and beneficial analysis for improving the daily decision-making that affects every area previously mentioned.

Therefore, it is very important to learn about this emerging world of data, in order not to be left out, for taking advantage of it and improving our professional features as the team performance.

Knowledge about the processing and visualization of the defensive data presented in this module 4 will be useful for improving our abilities as analysts and/or coaches to decode the game. We have learned to do it by video-analysis and now we have to incorporate data as an additional element, in which we get support for broadening our point of view and providing it with bidirectionality.

- a) Going from the game to data: getting support on data visualizations for reinforcing the opinions we have achieved by analyzing the play by the video-analysis and by observing the own as well as the opponent's behavior.
- b) Going from data to game: obtaining relevant information about the game by the data which we would hardly obtain with video-analysis (due to lack of time, attention or simply because of the high volume of matches we want to compare). For example, obtaining data about all passes into space performed by a team from their own half court in all matches played in La Liga and others.

As a farewell to this course, I just want to tell you that it has been a challenge for me to express in these files all the knowledge I have about the high-press topic and to share it with all of you.

I consider myself lucky for having learned from coaches and colleagues who are so good and qualified. For me, it is a great pleasure to share this knowledge with you, knowledge that is the result of my learning and previous experiences.

I hope you have liked doing this course and it is useful for you at the professional level.

I hope to see you all soon working in a professional coaching staff, if you are not doing it already.



## References

**De Juana, R.** (2019). Tableau: así es la empresa por la que Salesforce ha pagado más de 13.000 millones [Publication in Muy Computer Pro blog]. Retrieved from <https://www.muycomputerpro.com/2019/06/11/tableau-asi-es-la-empresa-por-la-que-salesforce-ha-pagado-mas-de-13-000-millones>

**Deutsche Fussball Liga (DFL).** (2020). DFL and Amazon Web Services to provide new real-time match analysis [Publication on Deutsche Fussball Liga web site]. Retrieved from <https://www.dfl.de/en/news/dfl-and-amazon-web-services-to-provide-new-real-time-match-analysis/>

**Fbref.** (n. d.). League Table [Publication on Fbref web site]. Retrieved from <https://fbref.com/en/comps/12/La-Liga-Stats>

**Fombella Pombal, D.** (2020a). Module 4. Sports data analysis with R and Python [Class notes]. Master in Sports Big Data, Universidad Católica San Antonio de Murcia.

**Fombella Pombal, D.** (2020b). Module 7. Sports data providers-Opta, InStat, Wyscout, Mediacoach, STATS, StatsBomb [Class notes]. Master in Sports Big Data, Universidad Católica San Antonio de Murcia.

**[Untitled image about Pentaho tools].** (n. d.). Retrieved from [https://lh3.googleusercontent.com/4y-ZxvheJkzXhDva8cLQOAccgCwF1wFHRD9PP\\_SWnRQarn\\_OIyVU5Ktwc1cgrJY71bYvxQ=s130](https://lh3.googleusercontent.com/4y-ZxvheJkzXhDva8cLQOAccgCwF1wFHRD9PP_SWnRQarn_OIyVU5Ktwc1cgrJY71bYvxQ=s130)

**[Untitled image about a team's pressure ability]** (n. d.). Retrieved from [https://lh3.googleusercontent.com/dilrueUqUvtmDf4VyI\\_rTuVSwgSi\\_81ZkBzbaifEcEz6YIAGxBIQR04xmOKSxsmbllqUPw=s85](https://lh3.googleusercontent.com/dilrueUqUvtmDf4VyI_rTuVSwgSi_81ZkBzbaifEcEz6YIAGxBIQR04xmOKSxsmbllqUPw=s85)

**[Untitled image about the 4 variables that measure a player's pressure ability]** (n. d.). Retrieved from <https://lh3.googleusercontent.com/D56U0KRjGbQ4I1Y-E73kKoKHqs-wrtDf0tL6Uh0QcERi7P8NwOIZpjYge1bGLg6QEML-t7M=s85>

**[Untitled image about the counterpressure]** (n. d.). Retrieved from [https://lh3.googleusercontent.com/C2rHhOr2Wm9Vqzk4xFYKkC6604tLENRDvJ7dDv1wYjnkHuYR\\_cmgkCmqNiv-TDYcxr7-HkM=s114](https://lh3.googleusercontent.com/C2rHhOr2Wm9Vqzk4xFYKkC6604tLENRDvJ7dDv1wYjnkHuYR_cmgkCmqNiv-TDYcxr7-HkM=s114)

**[Untitled image about keys for a correct data interpretation].** (n. d.). Retrieved from [https://lh3.googleusercontent.com/D30igKfmY4FIR3Zv\\_OIF5HEm-PM2akwwbRIOR3d9me5JMY8UKXVB9U\\_v4W0rmnWZLiyUxv4=s109](https://lh3.googleusercontent.com/D30igKfmY4FIR3Zv_OIF5HEm-PM2akwwbRIOR3d9me5JMY8UKXVB9U_v4W0rmnWZLiyUxv4=s109)



**[Untitled image about PPDA on a StatsBomb ranking]**. (n. d.). Retrieved from [https://lh3.googleusercontent.com/4QRdqGWNi0\\_bJ\\_Zs392Z0mH96LDHnIenpyl145mhA9-n0sTvqJTvUzR4NS2CSroCXb0U=s85](https://lh3.googleusercontent.com/4QRdqGWNi0_bJ_Zs392Z0mH96LDHnIenpyl145mhA9-n0sTvqJTvUzR4NS2CSroCXb0U=s85)

**[Untitled image about ranking of high-press on the opponent's half court]** (n. d.). Retrieved from <https://lh3.googleusercontent.com/CFilwr8zUC-i9Gk-kyX5x-NEZRYcgj2QH4WoN63PAWrea3qbb1Avq4JGoepUMUvt1feuA=s85>

**López, L.** (2019). Posesión, pero con pegada: así es el ADN del campeón de La Liga Santander. *El mundo* [Online version]. Retrieved from <https://noesfutoleslaliga.elmundo.es/directos-al-futuro/posesion-pero-con-pegada-asi-es-el-adn-del-campeon-de-laliga-santander>

**Microsoft.** (2021). Power BI Desktop [data analysis software]. Redmond, US: Microsoft.

**Olmeda, M.** (2020). La tecnología que ayuda a los equipos a preparar los partidos. *Diario ABC* [Online version]. Retrieved from <https://www.abc.es/contentfactory/post/eslaliga/laliga-apuesta-por-la-tecnologia-para-hacerse-mas-competitiva/>

**Ovalle, C.** (2021). Qué dicen los datos: Bravo fue el tercer arquero que más goles evitó en La Liga la temporada pasada. *Diario BioBio* [Online version] Retrieved from <https://www.biobiochile.cl/noticias/futbol-internacional/chilenos-en-el-exterior/2021/05/27/que-dicen-los-datos-bravo-fue-el-tercer-arquero-que-mas-goles-evito-en-la-liga-la-temporada-pasada.shtml>

**Oviedo Fernández, F.** (2020). Module 1. Introduction to big data and theories: theory about games and theory about decision-making [Class notes]. Master in Sports Big Data, Universidad Católica San Antonio de Murcia.

**Peña Rodríguez, P.** (2020). Introducción a las métricas defensivas en fútbol. Retrieved from <https://StatsBomb.com/es/2020/03/metricas-defensivas/>

**Pérez, F.** (2021). Proyecto Jupyter [Programming language software]. Berkeley, US: Proyecto Jupyter.

**RStudio Inc.** (2021). RStudio [Programming language software]. Boston, US: RStudio Inc.

**Second Spectrum.** (n. d.). Win more games [Publication on Second Spectrum web site]. Retrieved from <https://www.secondspectrum.com/ourwork/teams-leagues.html>

**SoftwarePara.** (n. d.). Tableau [Publication on SoftwarePara blog]. Retrieved from <https://softwarepara.net/tableau/>



**StatsBomb.** (February 11th, 2021). ¿Cómo se desenvuelven ante presión rival los equipos de la liga? [Publication on StatsBomb web site]. Retrieved from <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-equipos-de-la-liga/>

**StatsBomb.** (February 18th, 2021). ¿Cómo se desenvuelven ante presión rival los jugadores de las cinco grandes ligas europeas? [Publication on StatsBomb web site]. Retrieved from <https://StatsBomb.com/es/2021/02/como-se-desenvuelven-ante-presion-rival-los-jugadores-de-las-cinco-grandes-ligas-europeas/>

**StatsBomb.** (n. d.). StatsBomb 360 [Publication on StatsBomb web site]. Retrieved from <https://StatsBomb.com/360-data/>

**Unir.** (2019). Lenguaje R, ¿qué es y por qué es tan usado en *Big data*? [Publication on Unir blog]. Retrieved from <https://www.unir.net/ingenieria/revista/lenguaje-r-big-data/>

**Web-Futbol.** (n. d.). *Big data* en el fútbol [Publication on Web-Futbol blog]. Retrieved from <https://www.web-futbol.com/big-data-en-el-futbol/>

**Wyscout.** (n. d.). Datos, videos y productos: una combinación perfecta [Publication on Wyscout web site]. Retrieved from <https://wyscout.com/es/plataforma-futbol/>