Comunidad Hispano

16 NOVIEMBRE, 2019

Análisis de datos en el fútbol

library(soccergraphR)
library(FootballBadges)

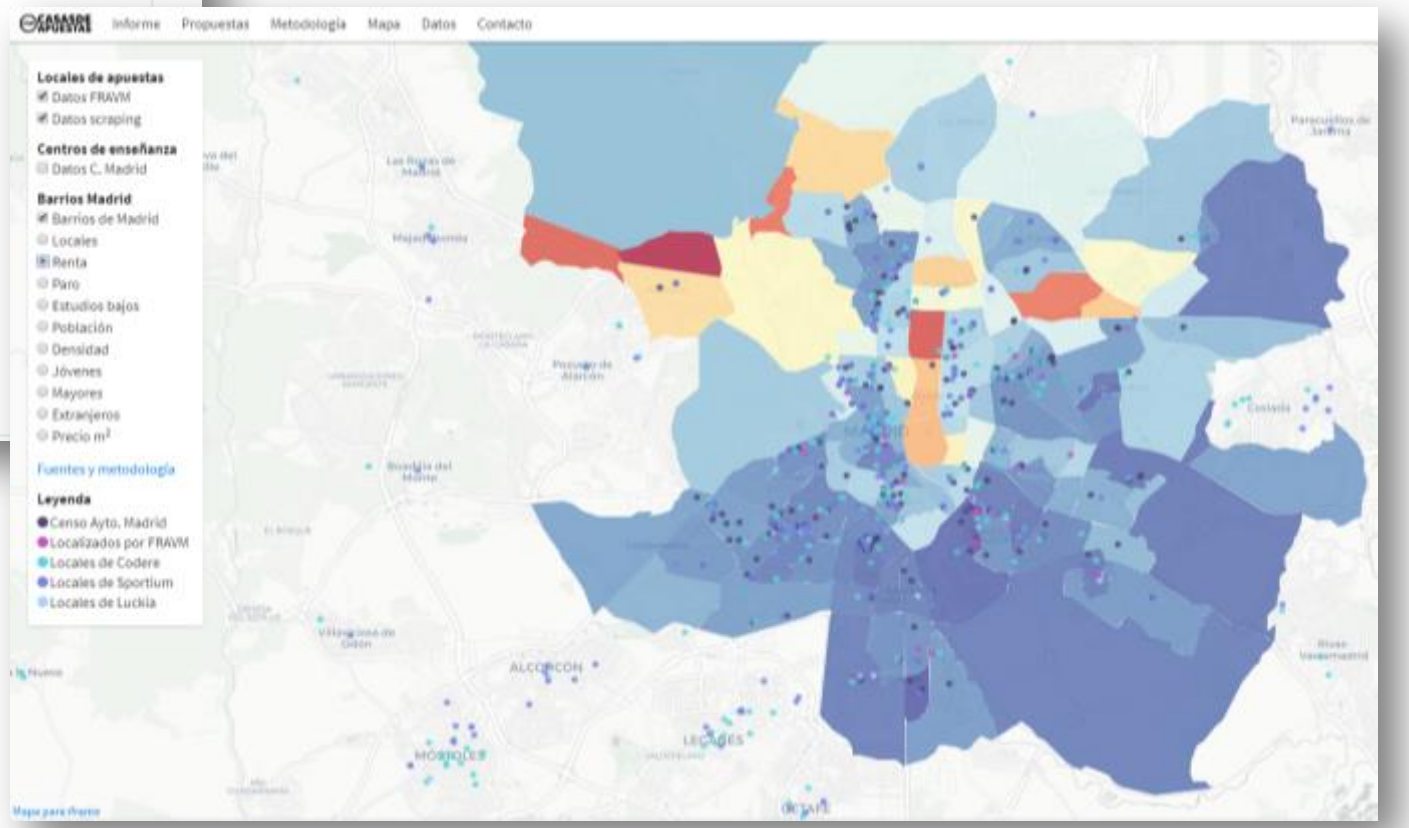
Jesús Lagos Milla

@Vdot_Spain

#XIJRes #BigDataFutbol

https://github.com/Jelagmil/

¿De qué no vamos a hablar? De apuestas.

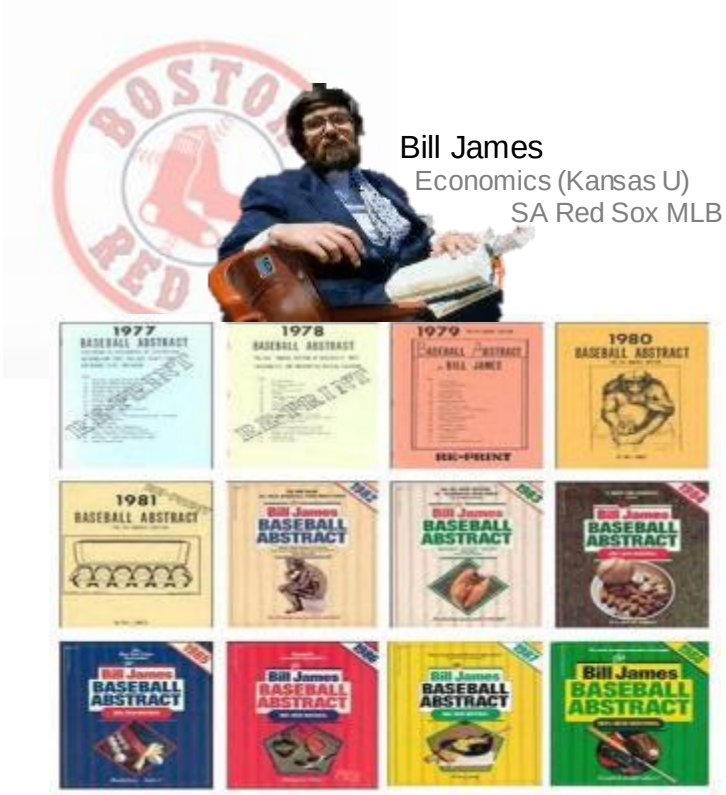


Índice

1. Introducción

Introducción

70’s-80’s ---------- 00’s ---------- 10’s ---------->



Bill James
Economics (Kansas U)
SA Red Sox MLB

Runs Created -> Expected Runs

Billy Beane
Economics* (UC San Diego)
VP Op OA’s MLB

Paul DePodesta
Economics (Harvard U)
CSO Cleveland Browns NFL

Record 20 victorias seguidas (actualemente 22)*
Coste plantilla 28 de 30 (40M) -> A. Rodriguez (Rangers) (22M)

Susana Ferreras
Telecomunication (U. Valladoilid)
Data Scientist Arsenal EPL

Vosse de Bood
Physiotherapy (Amsterdam)
Head of Sport Science AFC Ajax

```r
#' Esta funcion crea el grafo de pases para un equipo durante un partido a partir de una
#' segmentacion kmeans de posiciones. Es decir, si un pase lo hace en la misma zona 2 jugadores
#' computan como 1, porque lo que se dibuja es el cluster en esa zona de pase independiente quien
#' lo haya hecho
#'
#' @param df Un df
#' @param home Un número de 1 y 0 que indica si es home o away
#' @param cl El número de clusters que queremos calcular en la red
#' @return El grafo de \code{df}  del equipo local si \code{home} con un total de \code{cl} clusters
#' @examples
#' OptaMAPmatrixpasscluster(df,1,20)
#' OptaMAPmatrixpasscluster(df,1,10)
#' @export
#'
#'
#'
OptaMAPmatrixpasscluster <- function(df,home,cl){

polar<-dplyr::select(df,type_id,team_id,outcome,x,y,player_id,"2","107","123","213","home_team_id","away_team_id","140","141")
#Cambiamos los nombres que los numeros no gustan
names(polar)<-c("type_id","team_id","outcome","x","y","player_id","a","b","c","d","e","f","g","h")



if(home==1){
  polar <- dplyr::filter(polar,team_id==e)}
else{
  polar <- dplyr::filter(polar,team_id!=e)
}

#Nos quedamos los eventos de pase
polar <- dplyr::filter(polar,type_id==1 & is.na(a) & is.na(b) & is.na(c))

polar$yd<-polar$y*0.7
polar$hd<-polar$h*0.7

df1<-dplyr::select(polar,x,yd)
df2<-dplyr::select(polar,g,hd)
names(df2)<-c('x','yd')
matriz <- rbind(df1,df2)

set.seed(76964057) #Set the seed for reproducibility
k <-kmean
k$centers
table(k$c
matriz_cl

polar_clu
names(pol

matriz_ce
matriz_ce
```

2. Los datos

El problema a resolver

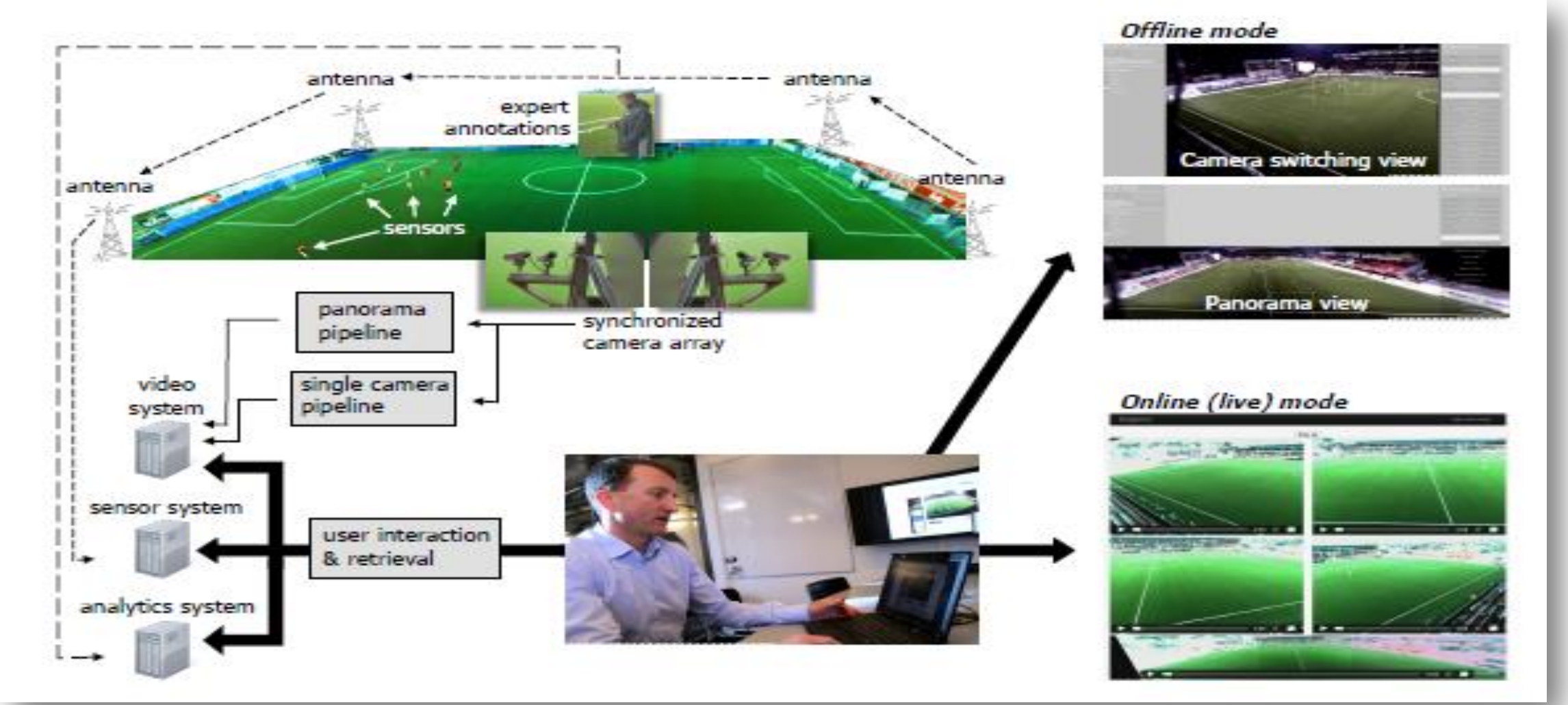
captura manual

- Se puede explotar con la librería para R #soccergraphR
- Se puede explotar con la librería para R #StatbombR

```
devtools::install_github('jelagmil/soccergraphR', build_opts = c("--no-resave-data", "--no-manual"), = TRUE)
```

El problema a resolver

captura manual **video**



El problema a resolver

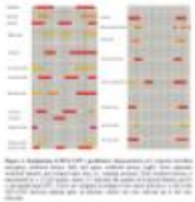
captura manual **video** **weareables**

3. Aplicaciones

Casos de uso en el fútbol

- Rendimiento físico y predicción de lesiones
- Scouting
- Métricas de juego
- Táctica y estrategia
- Acciones a Balón Parado
- Tracking

BARÇA INNOVATION HUB BARÇA SPORTS ANALYTICS SUMMIT

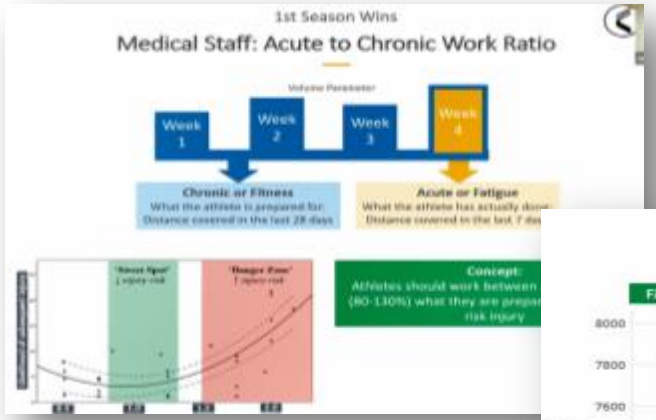


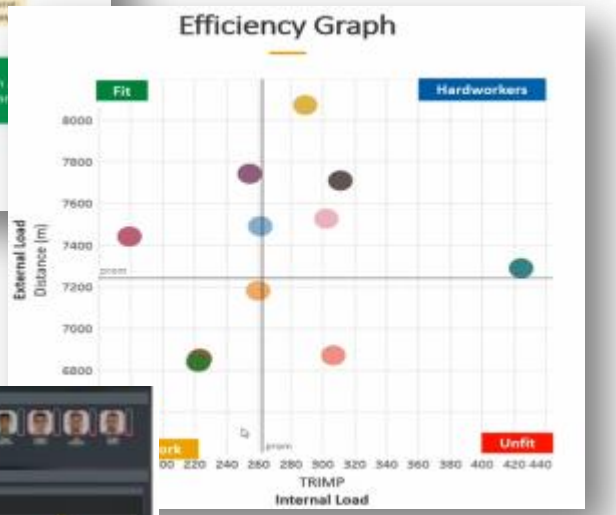
Explainable Injury Forecasting in Soccer via Multivariate Time Series and Convolutional Neural Networks

Luca Pappalardo[1], Luca Guerrini[2], Alessio Rossi[2], Paolo Cintia[2]

[1] Institute of Information Science and Technologies (ISTI), National Research Council of Italy (CNR), Pisa, Italy

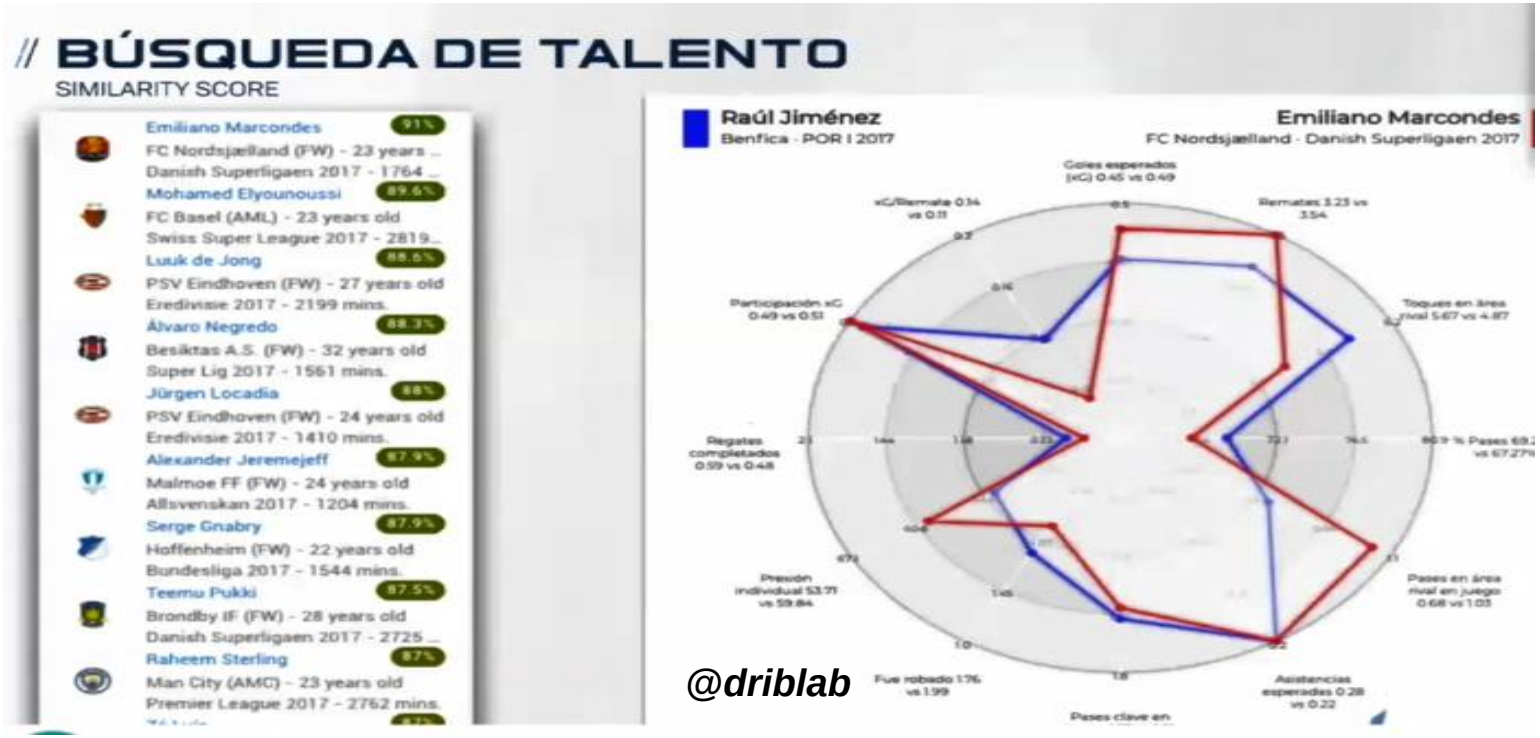
[2] Department of Computer Science, University of Pisa, Pisa, Italy





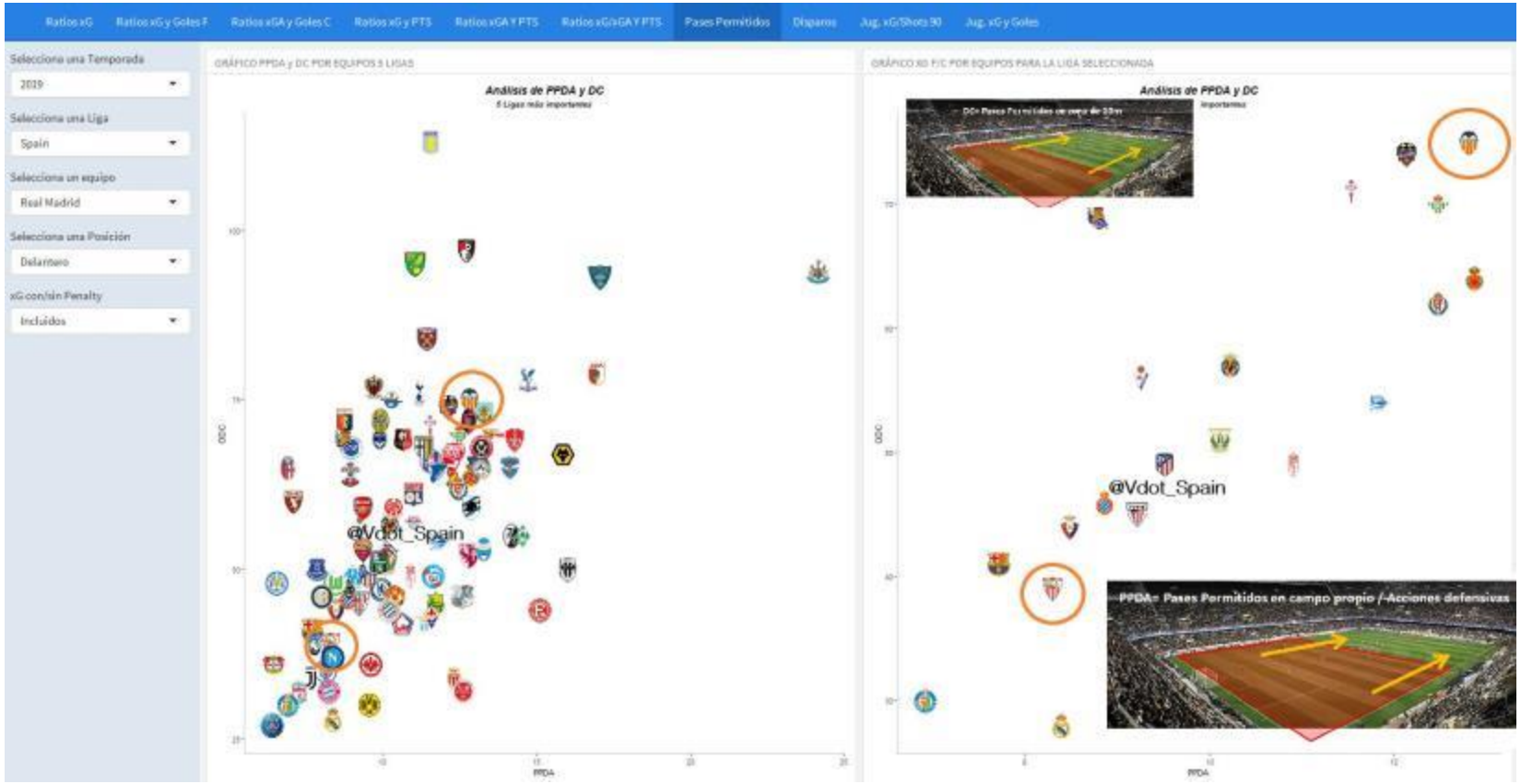
Casos de uso en el fútbol

- Rendimiento físico y predicción de lesiones
- Scouting
- Métricas de juego
- Táctica y estrategia
- Acciones a Balón Parado
- Tracking



Casos de uso en el fútbol

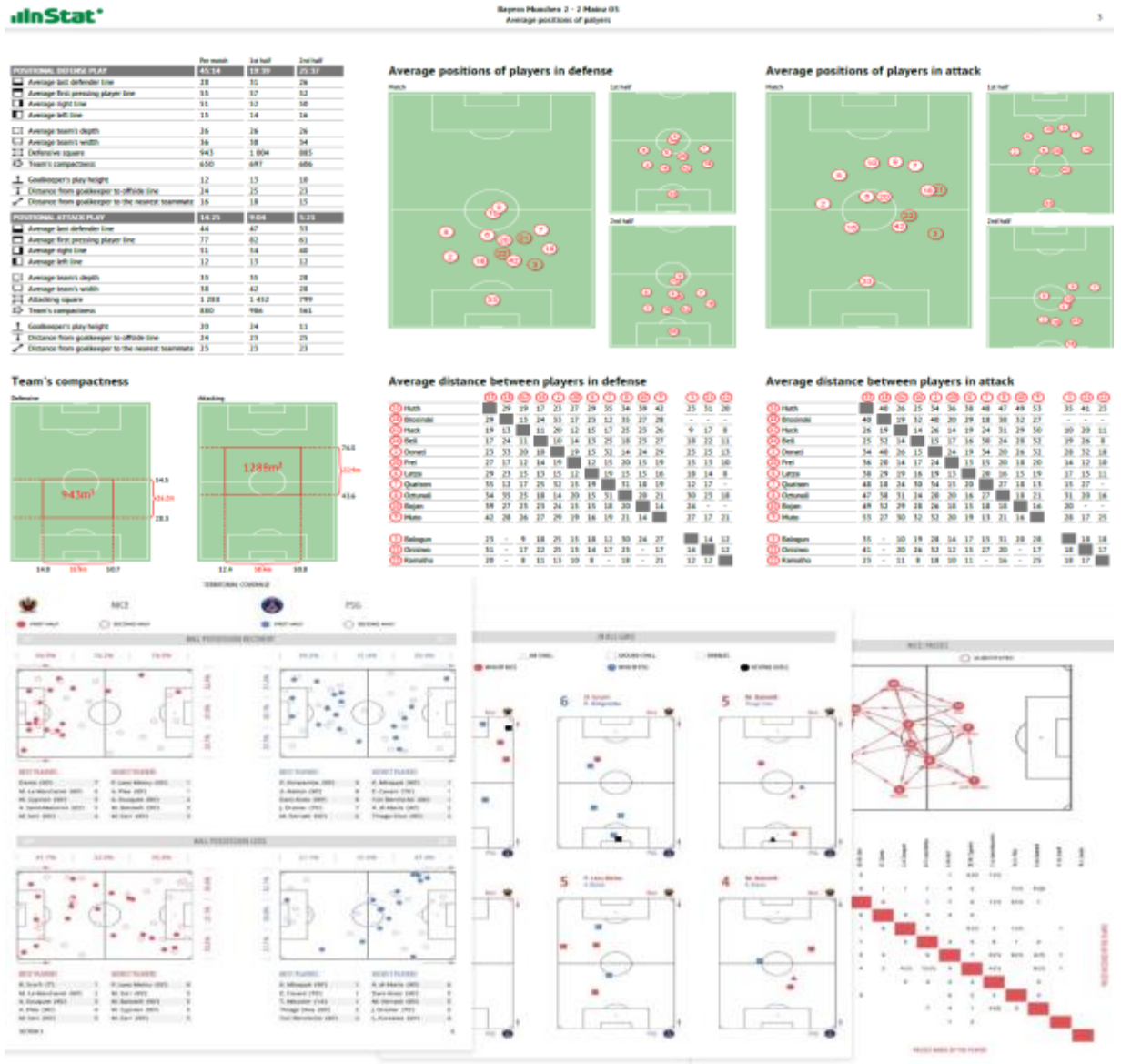
- Rendimiento físico y predicción de lesiones
- Scouting
- Métricas de juego
- Táctica y estrategia
- Acciones a Balón Parado
- Tracking



xG, npxG, xA, xT, xBuildup, xGChain, PPDA, DC, xPTS, etc.

Casos de uso en el fútbol

- Rendimiento físico y predicción de lesiones
- Scouting
- Métricas de juego
- Táctica y estrategia
- Acciones a Balón Parado
- Tracking



Casos de uso en el fútbol

- Rendimiento físico y predicción de lesiones
- Scouting
- Métricas de juego
- Táctica y estrategia
- Acciones a Balón Parado
- Tracking

Casos de uso en el fútbol

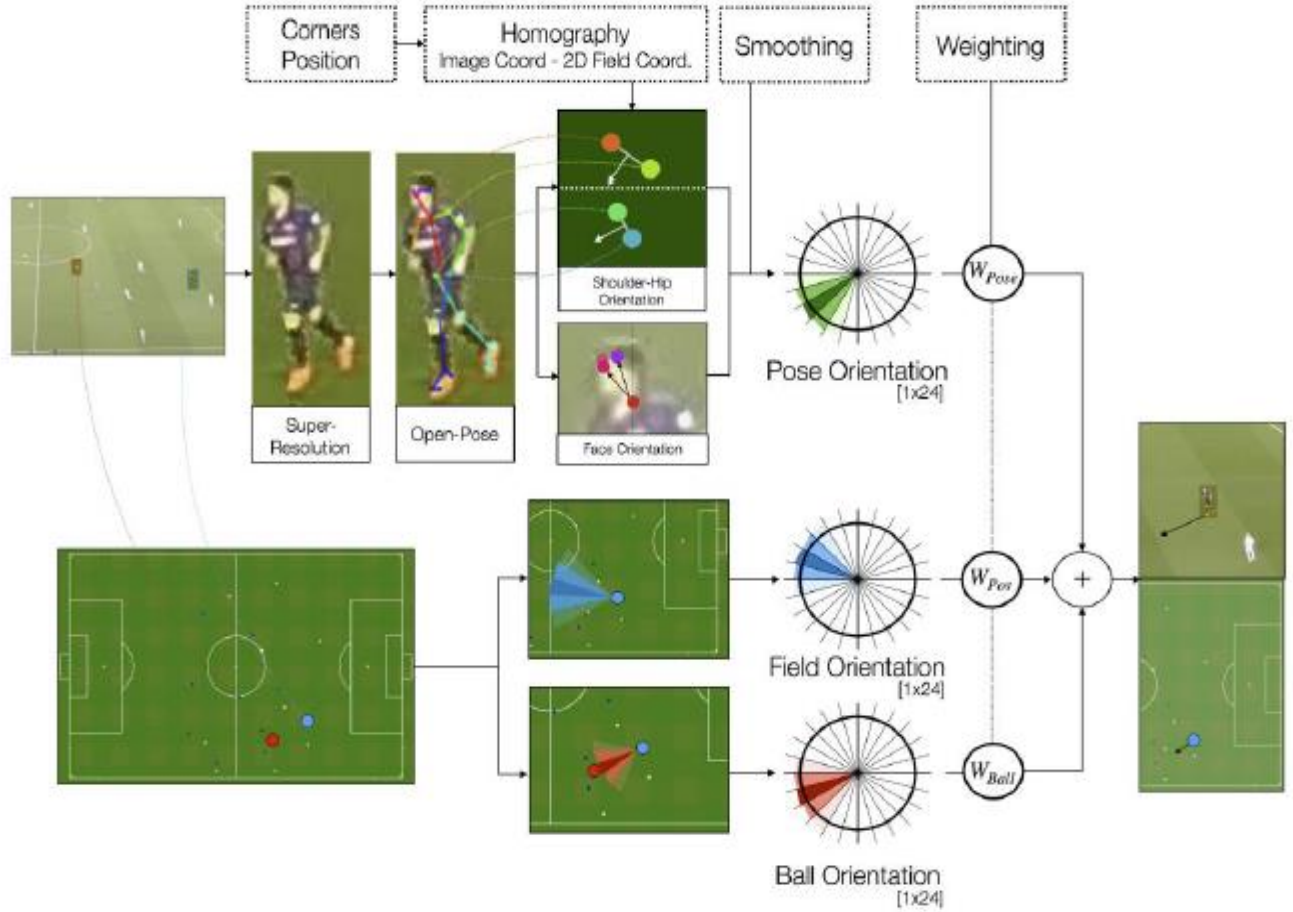
- Rendimiento físico y predicción de lesiones
- Scouting
- Métricas de juego
- Táctica y estrategia
- Acciones a Balón Parado
- Tracking

BARÇA INNOVATION HUB BARÇA SPORTS ANALYTICS SUMMIT

Head, Shoulders, Hip and Ball... Hip and Ball! Using Pose Data to Leverage Football Player Orientation

Adrià Arbués-Sangüesa, Gloria Haro, Coloma Ballester, Adrián Martín

Universitat Pompeu Fabra | adria.arbues@upf.edu



Casos de uso en el fútbol

- Rendimiento físico y predicción de lesiones
- Scouting
- Métricas de juego
- Táctica y estrategia
- Acciones a Balón Parado
- Tracking

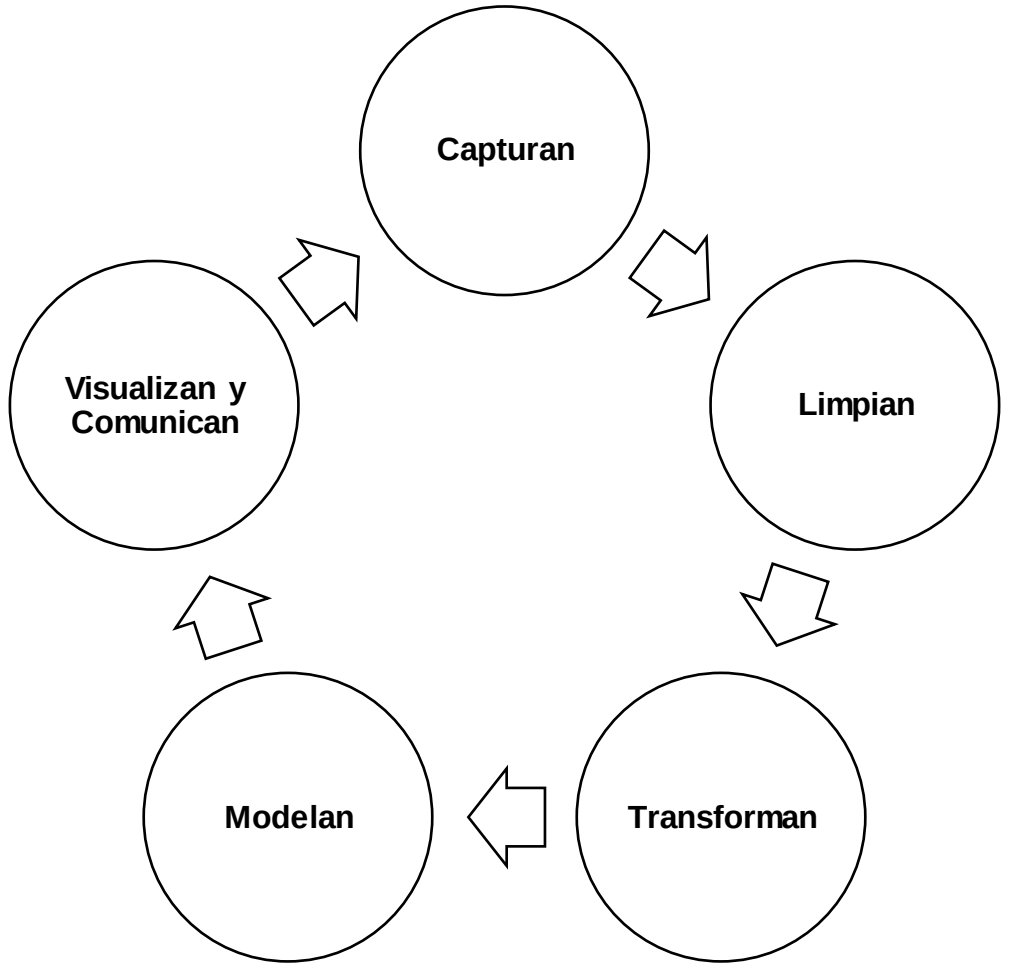
Casos de uso en el fútbol

- Rendimiento físico y predicción de lesiones
- Scouting
- Métricas de juego
- Táctica y estrategia
- Acciones a Balón Parado
- Tracking

4. Proveedores

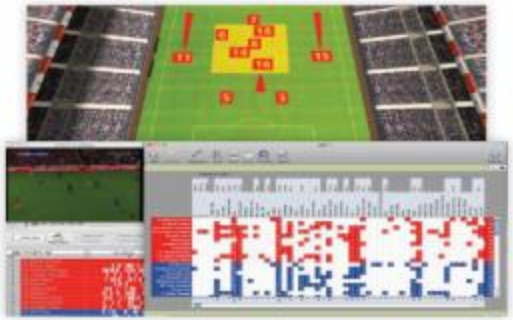
Principales proveedores

- opta
- instat
- wyscout
- mediacoach
- stats
- statsbomb
- understat



Principales proveedores

- opta
- instat
- wyscout
- mediacoach
- stats
- statsbomb
- understat



XML Data Feeds

Opta have made market leading XML feeds compatible with other market leading products such as *Scout7-Xeatre* and *SportsCode* to ensure clubs get the maximum value out of these systems. Clubs can now have a consistent data set across all of their analysis products.

Video Scout

Opta have developed a unique online interface that allows clubs to download match footage online and search the actions within it using *Opta*'s time coded events. Users can create unlimited playlists from clips across multiple matches and download them as a single movie.

Data Scout

Developed in conjunction with *StratBridge LLC*, *Data Scout* is the next generation of football player analysis software. This entirely online solution allows users to explore *Opta*'s massive player database, run comparisons and view trends over time.

Live Monitor

Live Monitor provides an online tool through which to monitor your teams performance live as a game progresses. View the impact of tactical changes via a pass matrix, player stats and graphics of average shape or player possessions. Compatible with any tablet platform *Live Monitor* offers the flexibility to access stats anywhere during your own matches or whilst scouting players and teams.

Reports & Analysis

Our skilled team of data analysts can provide clubs with reports and analysis packs based on pre-agreed criteria. This can be weekly opponent scouting reports or own-team performance over weeks, months or seasons. This can be provided in a wide variety of formats for ease of circulation within the club itself.

Data Query Tool

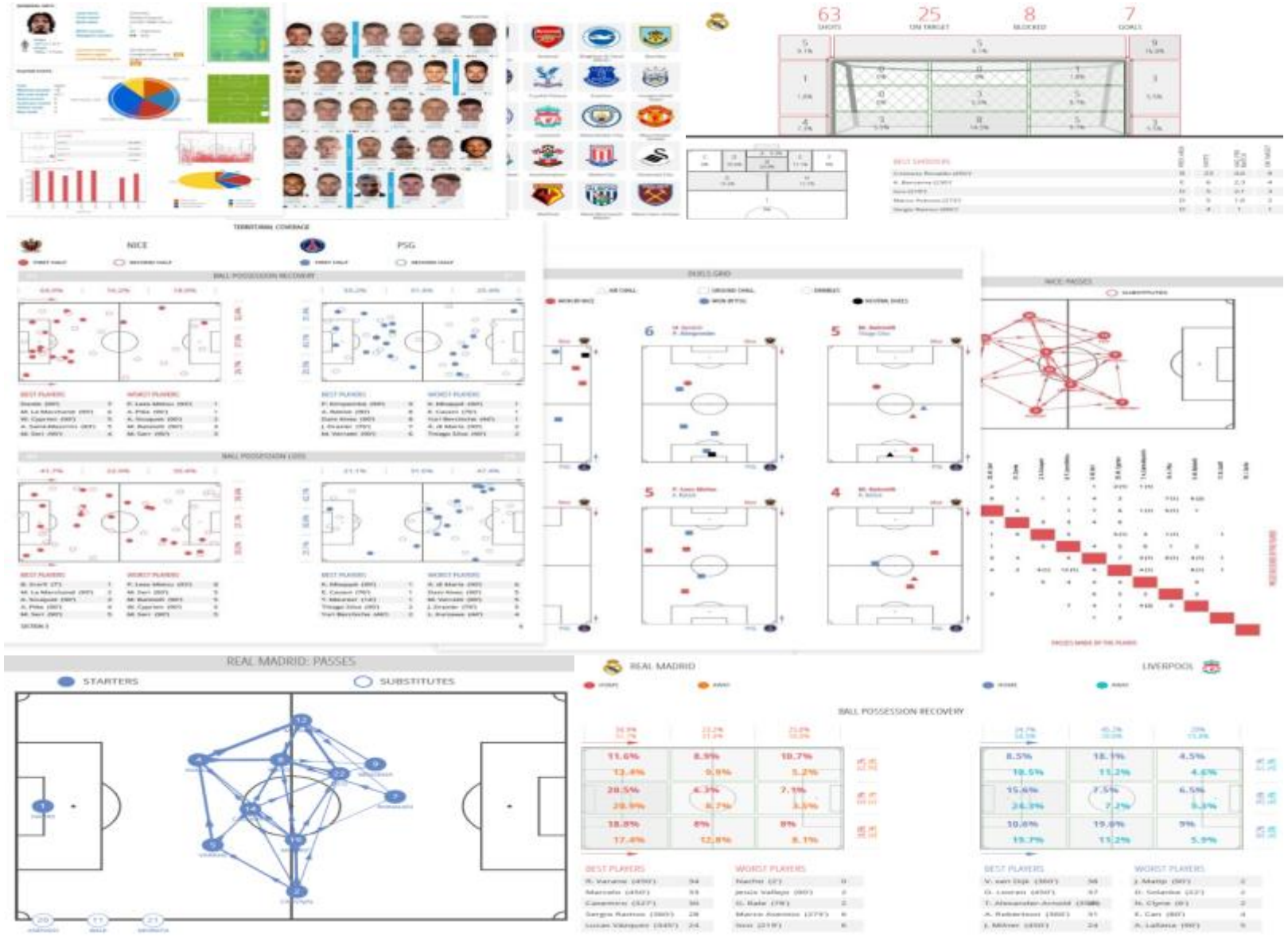
Accessed entirely online, the *Opta Data Query Tool* allows users to select any competition, followed by specific teams, players and positions to create a multivariate table. This can be customised across over 120 different variables across multiple or single seasons.

Principales proveedores

- opta
- instat
- wyscout
- mediacoach
- stats
- statsbomb
- understat

Principales proveedores

- opta
- instat
- wyscout
- mediacoach
- stats
- statsbomb
- understat

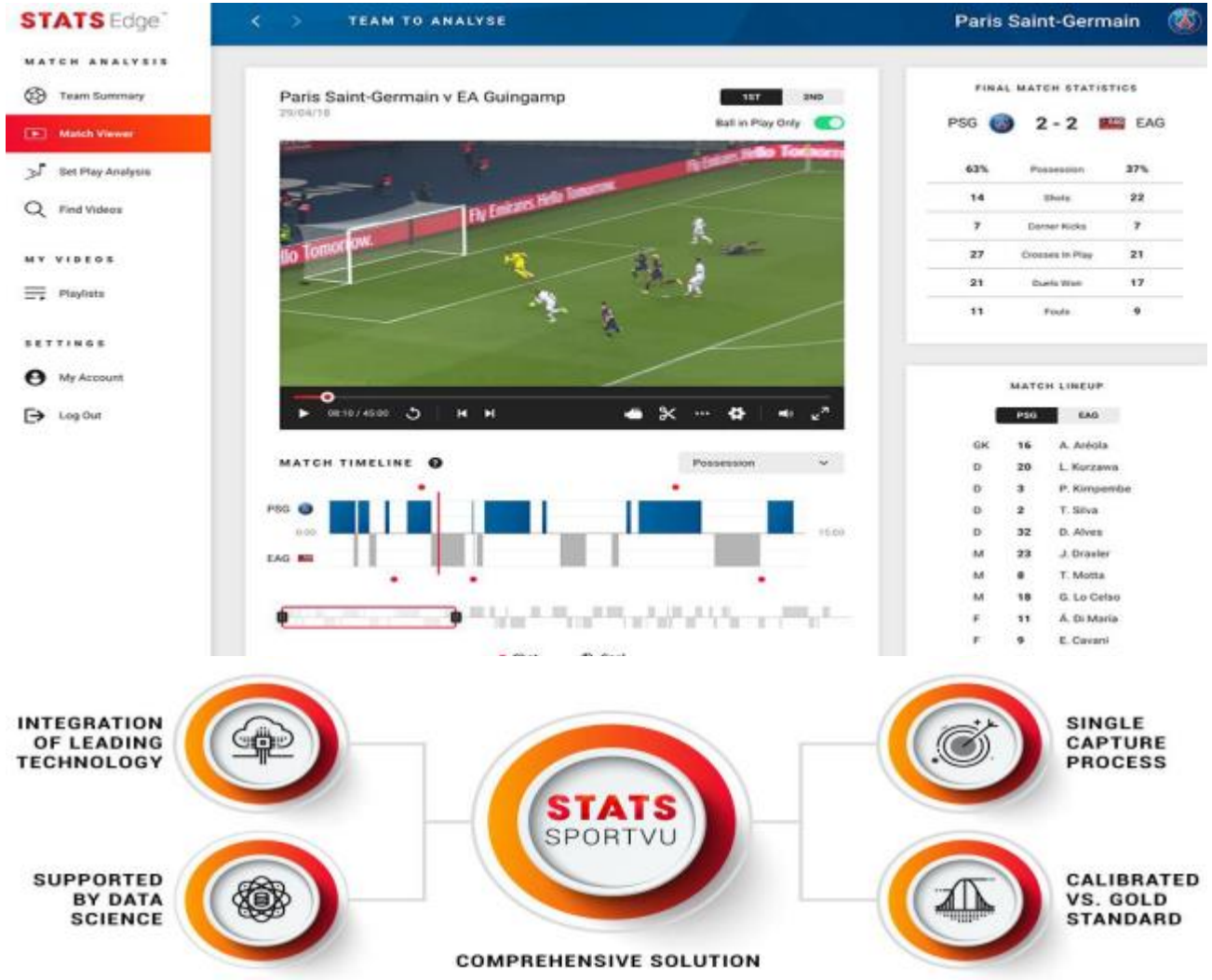


Principales proveedores

- opta
- instat
- wyscout
- mediacoach
- stats
- statsbomb
- understat

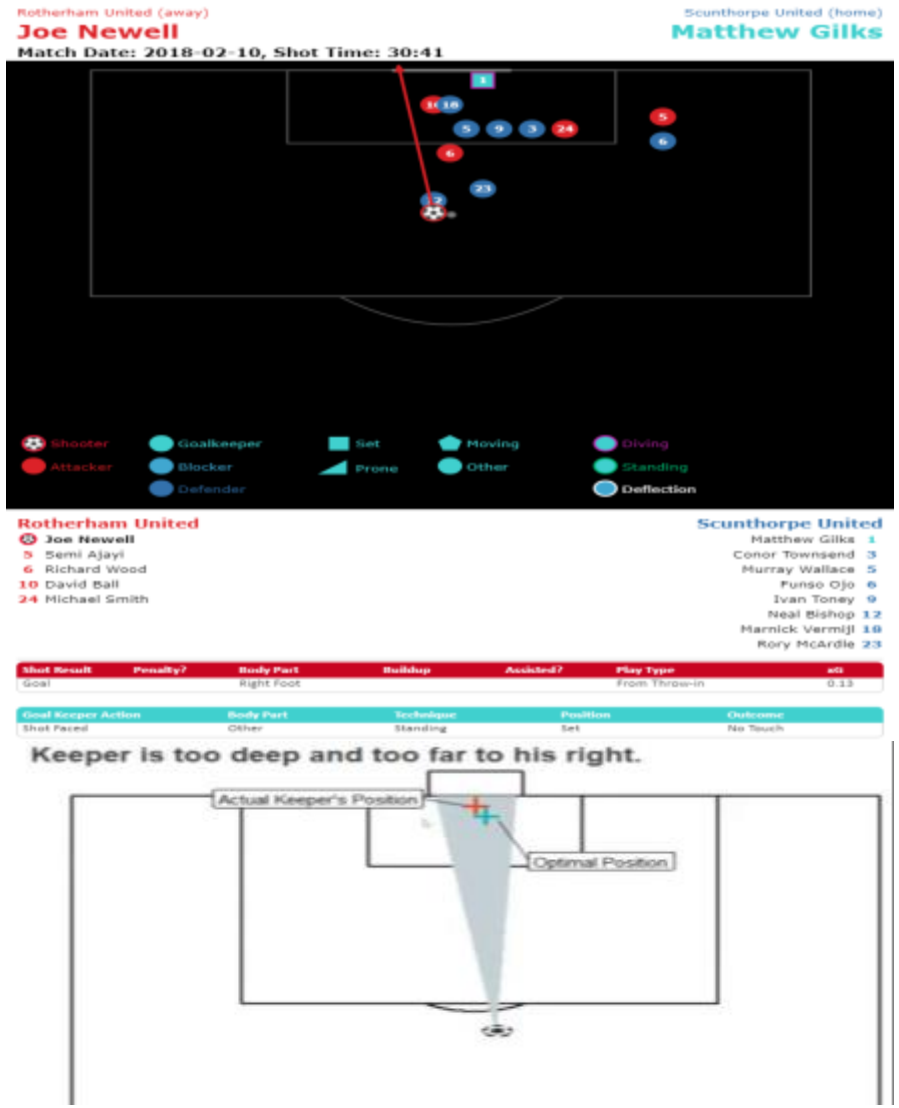
Principales proveedores

- opta
- instat
- wyscout
- mediacoach
- stats
- statsbomb
- understat



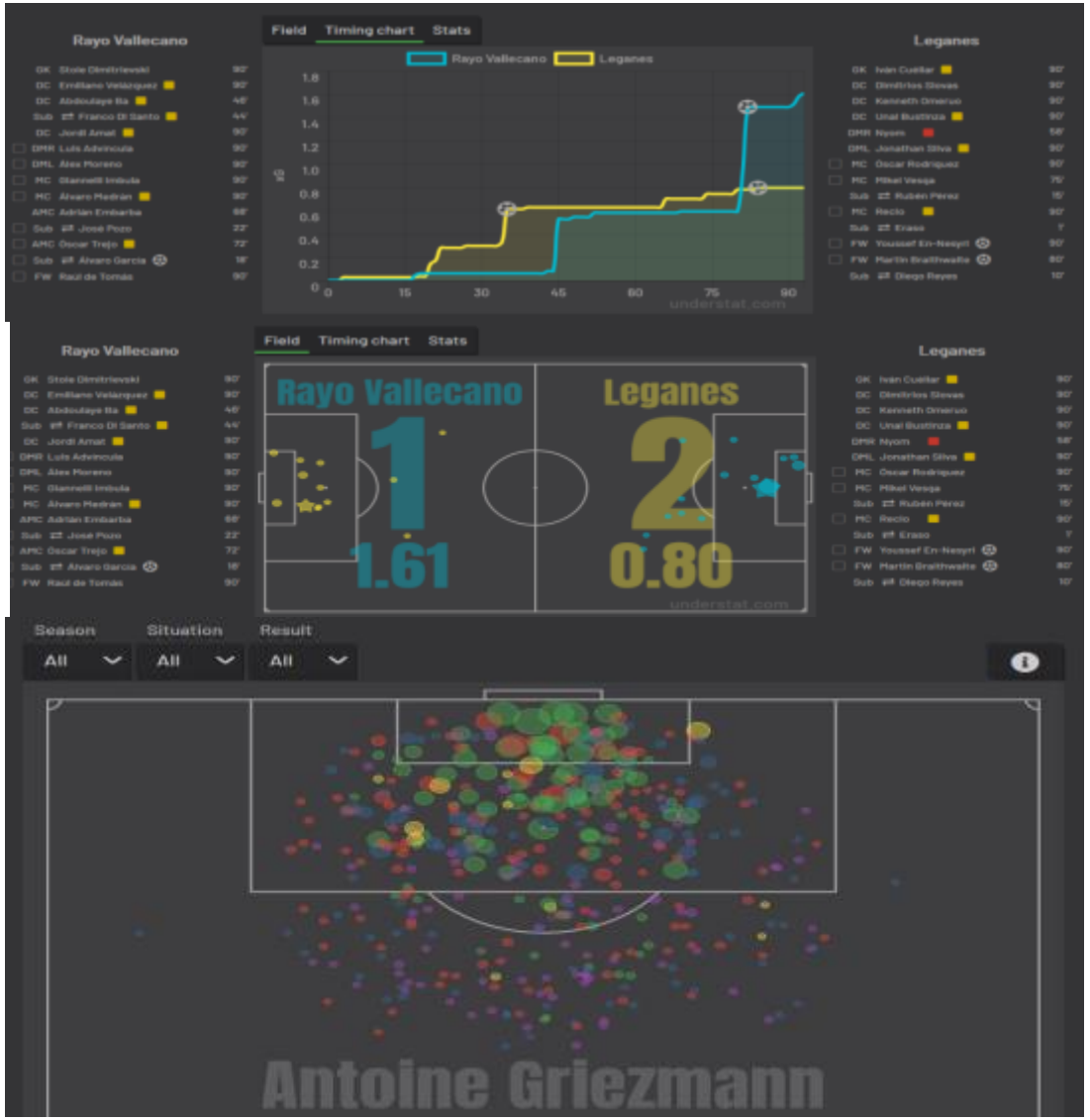
Principales proveedores

- opta
- instat
- wyscout
- mediacoach
- stats
- statsbomb
- understat



Principales proveedores

- opta
- instat
- wyscout
- mediacoach
- stats
- statsbomb
- understat

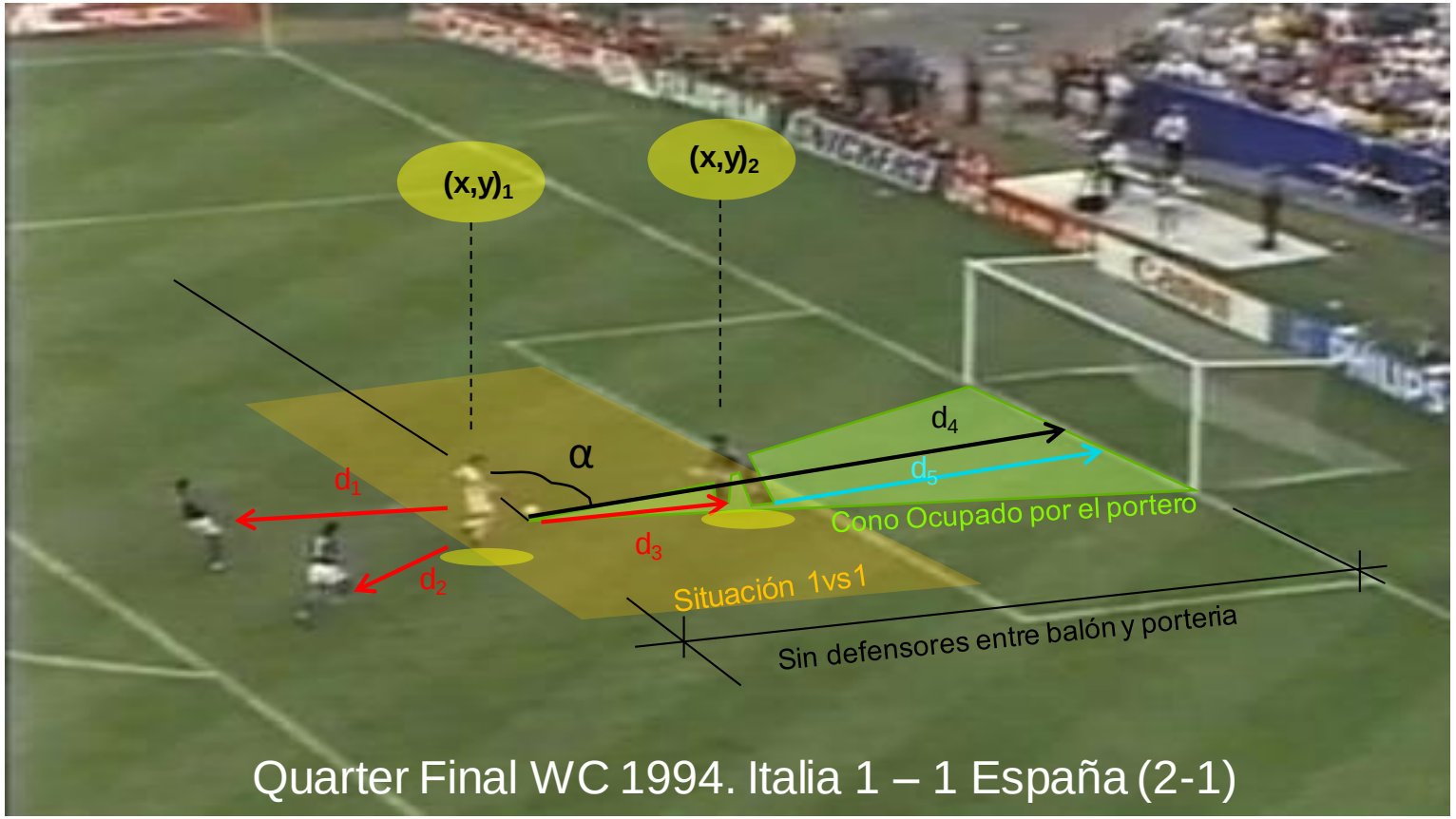


```r
#' Esta funcion crea el grafo de pases para un equipo durante un partido a partir de una
#' segmentacion kmeans de posiciones. Es decir, si un pase lo hace en la misma zona 2 jugadores
#' computan como 1, porque lo que se dibuja es el cluster en esa zona de pase independiente quien
#' lo haya hecho
#'
#' @param df Un df
#' @param home Un número de 1 y 0 que indica si es home o away
#' @param cl El número de clusters que queremos calcular en la red
#' @return El grafo de \code{df}  del equipo local si \code{home} con un total de \code{cl} clusters
#' @examples
#' OptaMAPmatrixpasscluster(df,1,20)
#' OptaMAPmatrixpasscluster(df,1,10)
#' @export
#'
#'
#'
OptaMAPmatrixpasscluster <- function(df,home,cl){

polar<-dplyr::select(df,type_id,team_id,outcome,x,y,player_id,"2","107","123","213","home_team_id","away_team_id","140","141")
#Cambiamos los nombres que los numeros no gustan
names(polar)<-c("type_id","team_id","outcome","x","y","player_id","a","b","c","d","e","f","g","h")



if(home==1){
  polar <- dplyr::filter(polar,team_id==e)}
else{
  polar <- dplyr::filter(polar,team_id!=e)
}

#Nos quedamos los eventos de pase
polar <- dplyr::filter(polar,type_id==1 & is.na(a) & is.na(b) & is.na(c))

polar$yd<-polar$y*0.7
polar$hd<-polar$h*0.7

df1<-dplyr::select(polar,x,yd)
df2<-dplyr::select(polar,g,hd)
names(df2)<-c('x','yd')
matriz <- rbind(df1,df2)

set.seed(76964057) #Set the seed for reproducibility
k <-kmean                                                    columns 1 and 2
k$centers
table(k$c                                                    ter
matriz_cl

polar_clu                                                    ",all.x=TRUE)
names(pol

matriz_ce
matriz_ce                                                    s, "VALUE")
```

5. Ejemplo xG

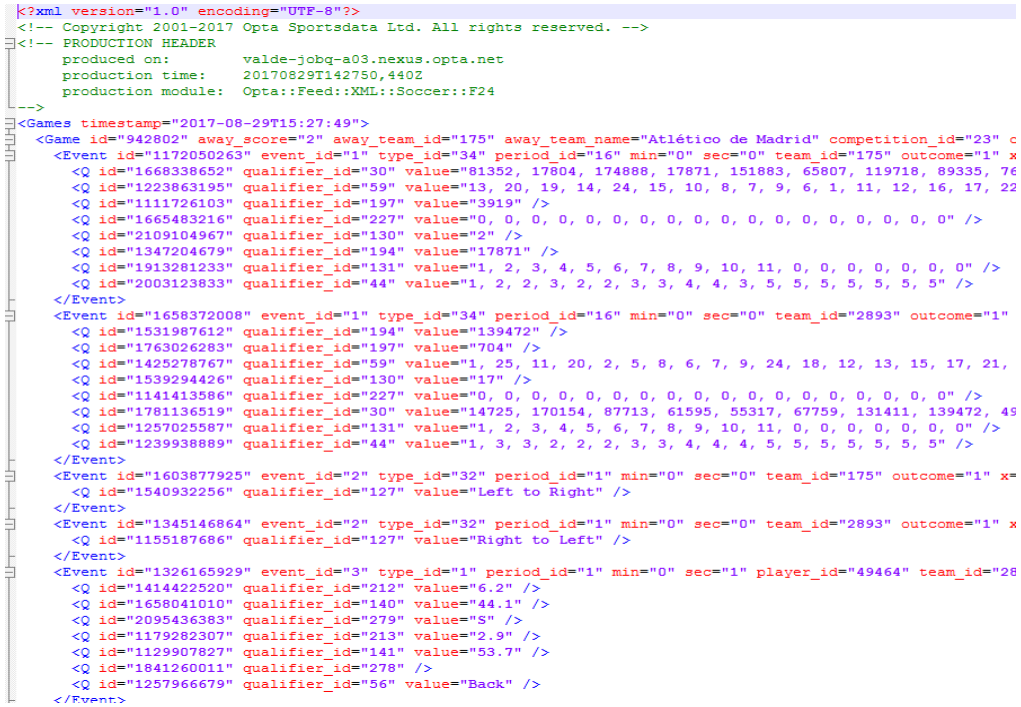
Expected Goals

Expected goals (xG) es una métrica que asigna a cada disparo la probabilidad de que sea gol a partir de variables como distancia, ángulo o situación. El ∑xG de un jugador o equipo nos ayudan a conocer la eficacia cara a puerta al comparar los goles reales y los esperados.



Quarter Final WC 1994. Italia 1 – 1 España (2-1)

Los feeds



- Se puede explotar con la librería para R #soccergraphR
- Se puede explotar con la librería para R #StatbombR

```
devtools::install_github('jelagmil/soccergraphR', build_opts = c("--no-resave-data", "--no-manual"), = TRUE)
```

Ejemplo cálculo xG

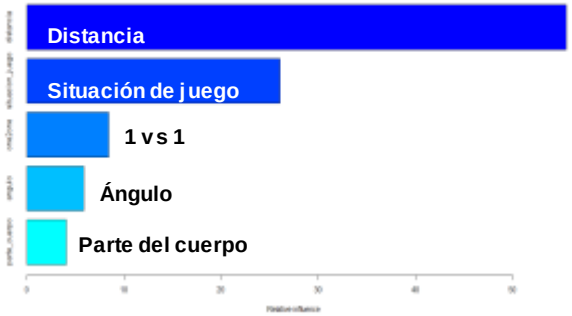
Datos de partida:
- F24_17/18
- F24_18/19
- **1.339.254 eventos**

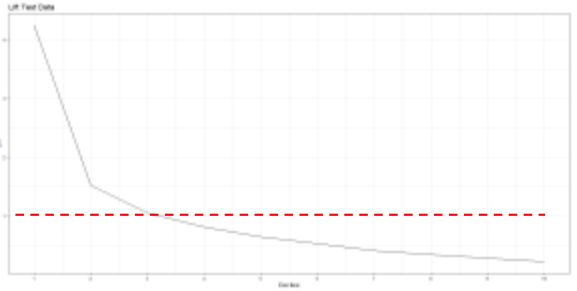
Disparos:
- Type_id=13,14,15,16
- **18.976 disparos**

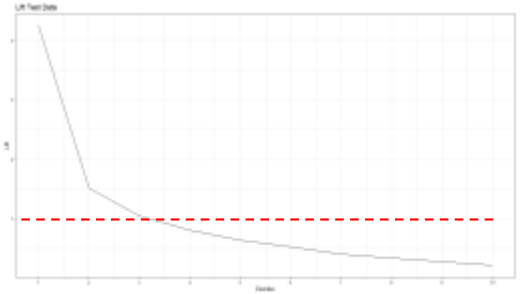
- **Goles 2056**
- **No goles 16901**
  - Distancia (num)
  - Ángulo (num)
  - 1vs1(factor 2)
  - Situación de juego (factor 8)
  - Parte cuerpo (factor 4)

library(gbm)
- Train/test (50%)
- cv.folds=5
- In.depth 1
- Minobsinnode 10
- Shrinkage 0.01

AUC Train=0.807
AUC Test=0.804

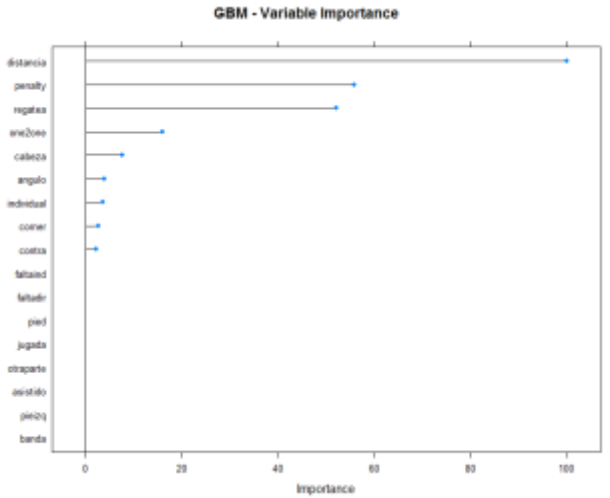


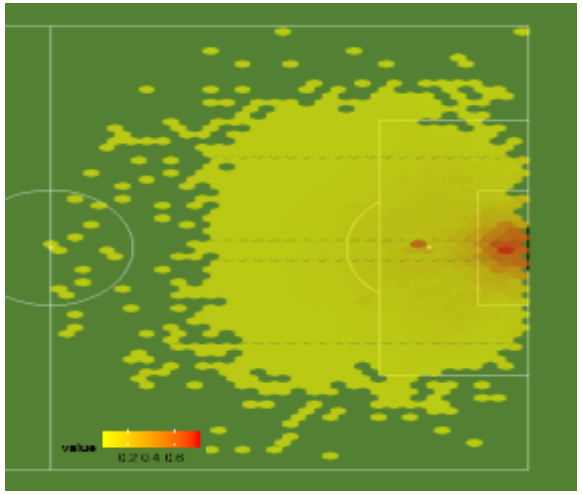




Library(caret)
- Train/test (50%)
- cv.folds=5
- In.depth 1
- Minobsinnode 5
- Shrinkage 0.1

AUC Train=0.807
AUC Test=0.819



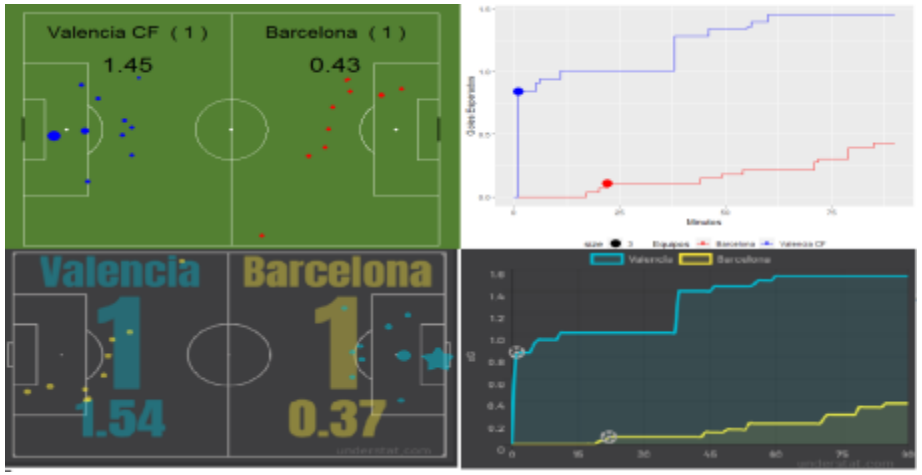


Expected Goals con #soccergraphR

En el paquete #soccergraph se incluye la función ***OptaMAPshootxG()*** que incluye un model xG propio basado en datos de opta como el modelo anterior y que permite pintar los mapas de xG que podemos ver en redes sociales a partir de un data entry en consola:

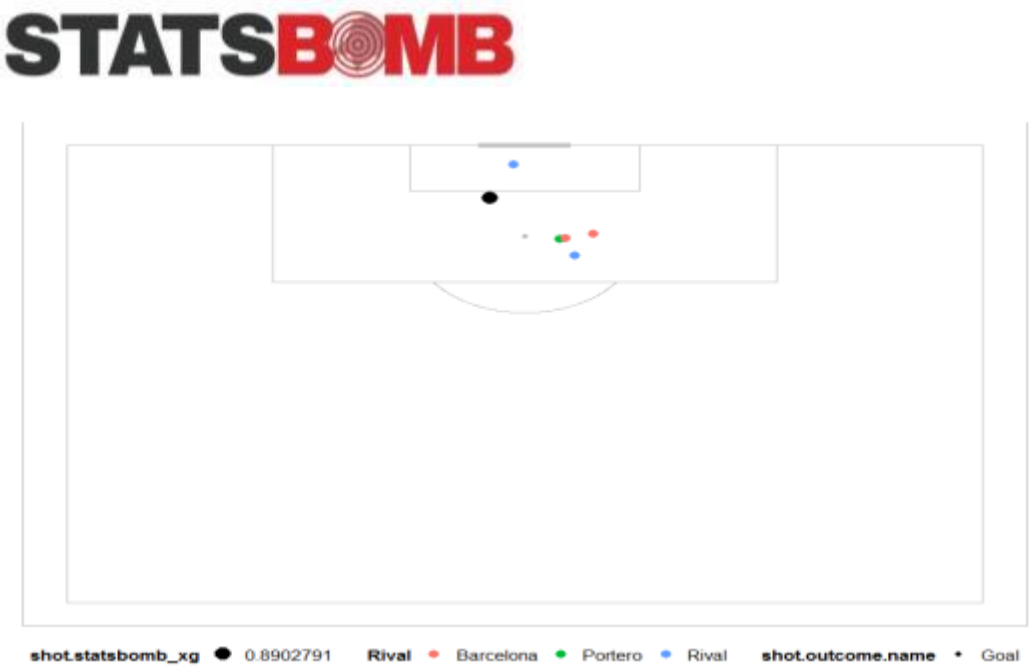
```
> OptaMAPshootxG(2)
Escribe el nombre del equipo 1: Valencia CF
¿Cuántos tiros vas a incluir del equipo 1?: 2
Escribe el nombre del equipo 2: Real Madrid
¿Cuántos tiros vas a incluir del equipo 2?: 1
Localiza en el plano la coordenada X y escríbela aquí (equipo 1): 98
Localiza en el plano la coordenada Y y escríbela aquí (equipo 1): 32
Selecciona una parte del cuerpo:

1: cabeza
2: pied
3: piei
4: otraparte

Selection: 2
Selecciona una parte del cuerpo:

1: juegoestatico
2: contra
3: faltaind
4: corner
5: faltadir
6: penalty
7: banda

Selection: 2
¿Es un tiro de penalty?:

1: Sí
2: No

Selection: 2
¿Es un uno contra uno contra el portero?:

1: Sí
2: No

Selection: 1
Localiza en el plano la coordenada X y escríbela aquí (equipo 1): 104
Localiza en el plano la coordenada Y y escríbela aquí (equipo 1): 48
Selecciona una parte del cuerpo:
```

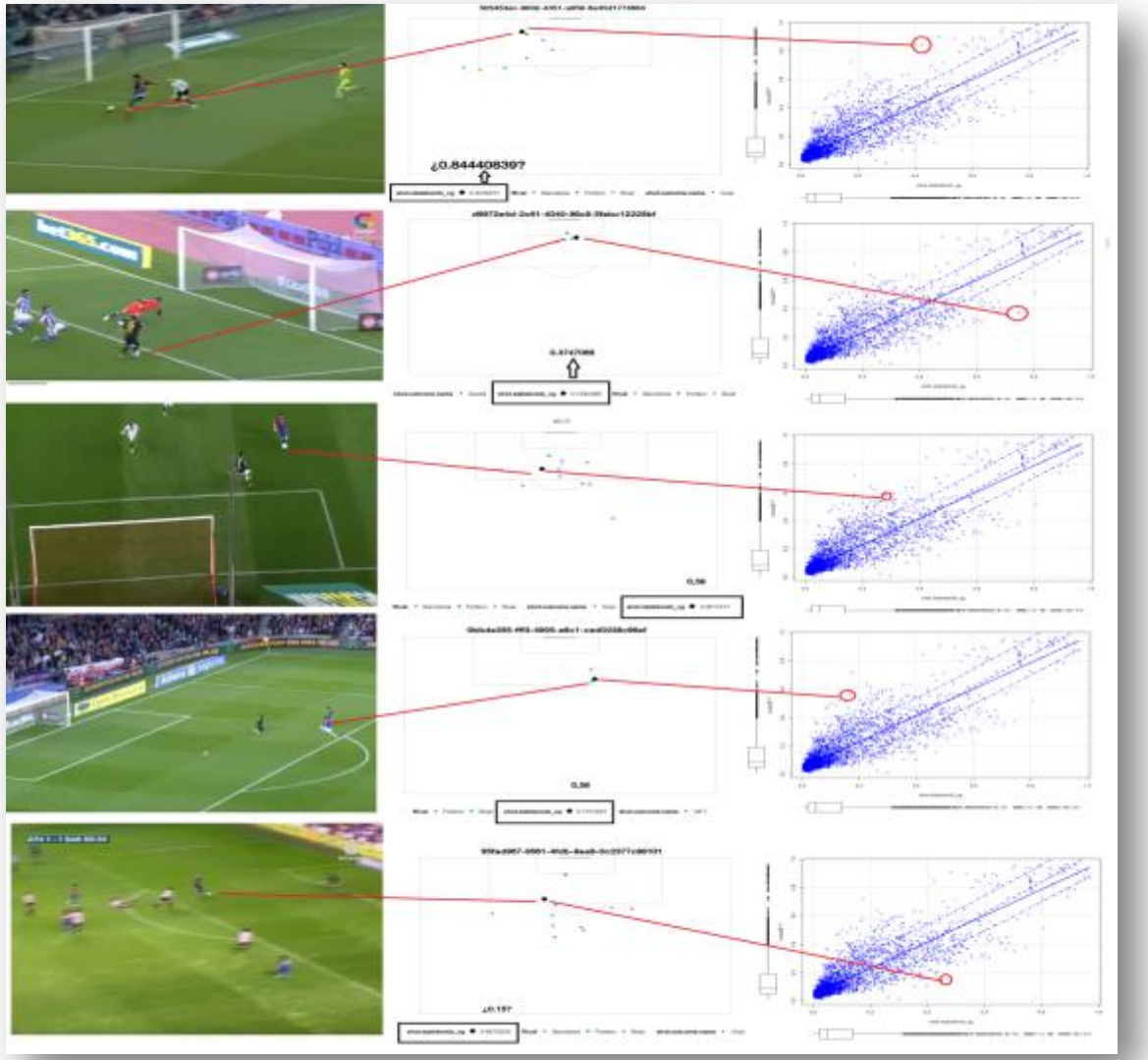
Con la función ***OptaMAPsummarisexG(df)*** pasamos un f24 de Opta y nos construye el resumen del partido



Expected Goals con #soccergraphR

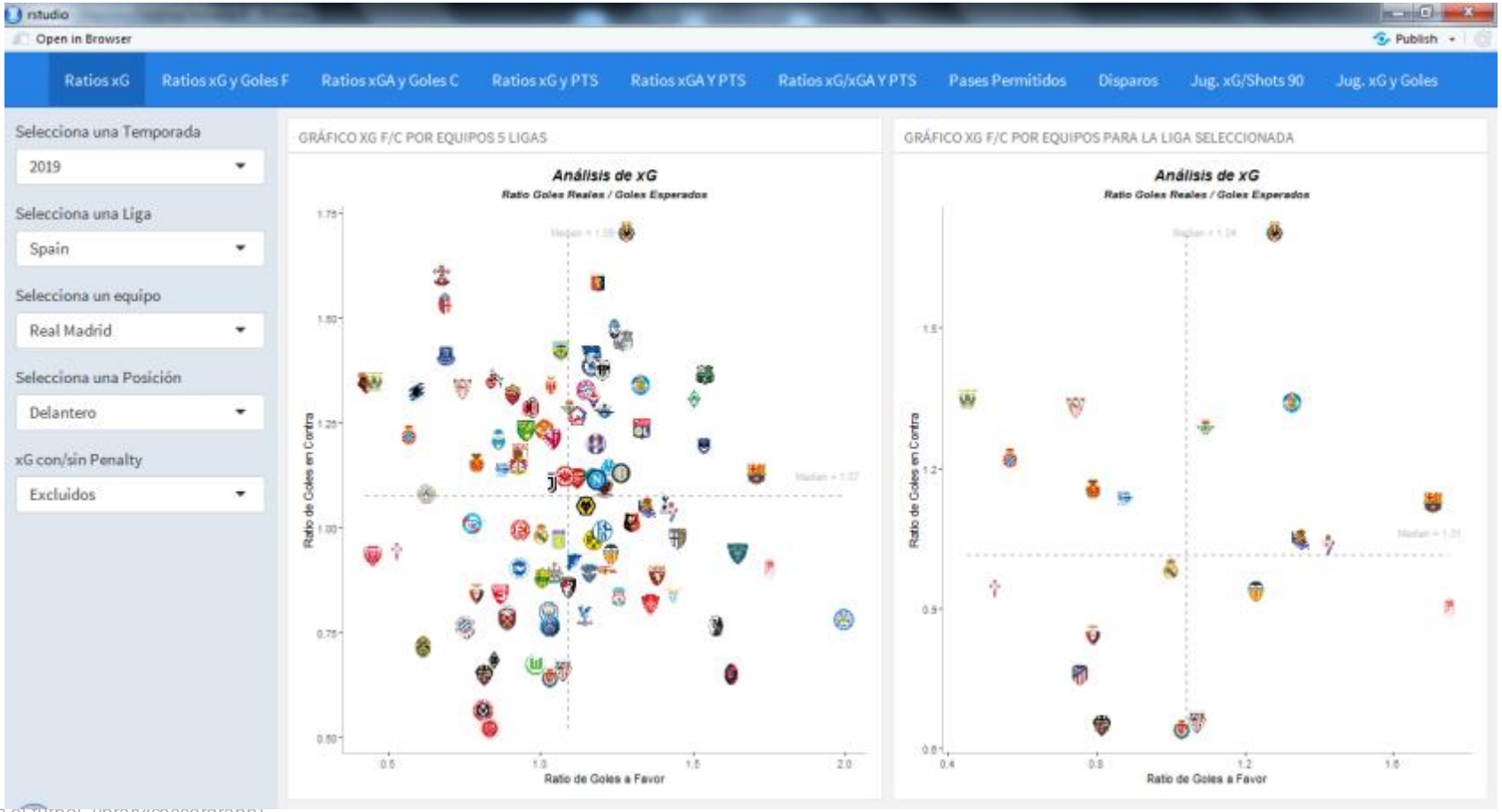
Statsbomb en su feed incluye el “frezze frame” que nos permite calcular las métricas anteriores.





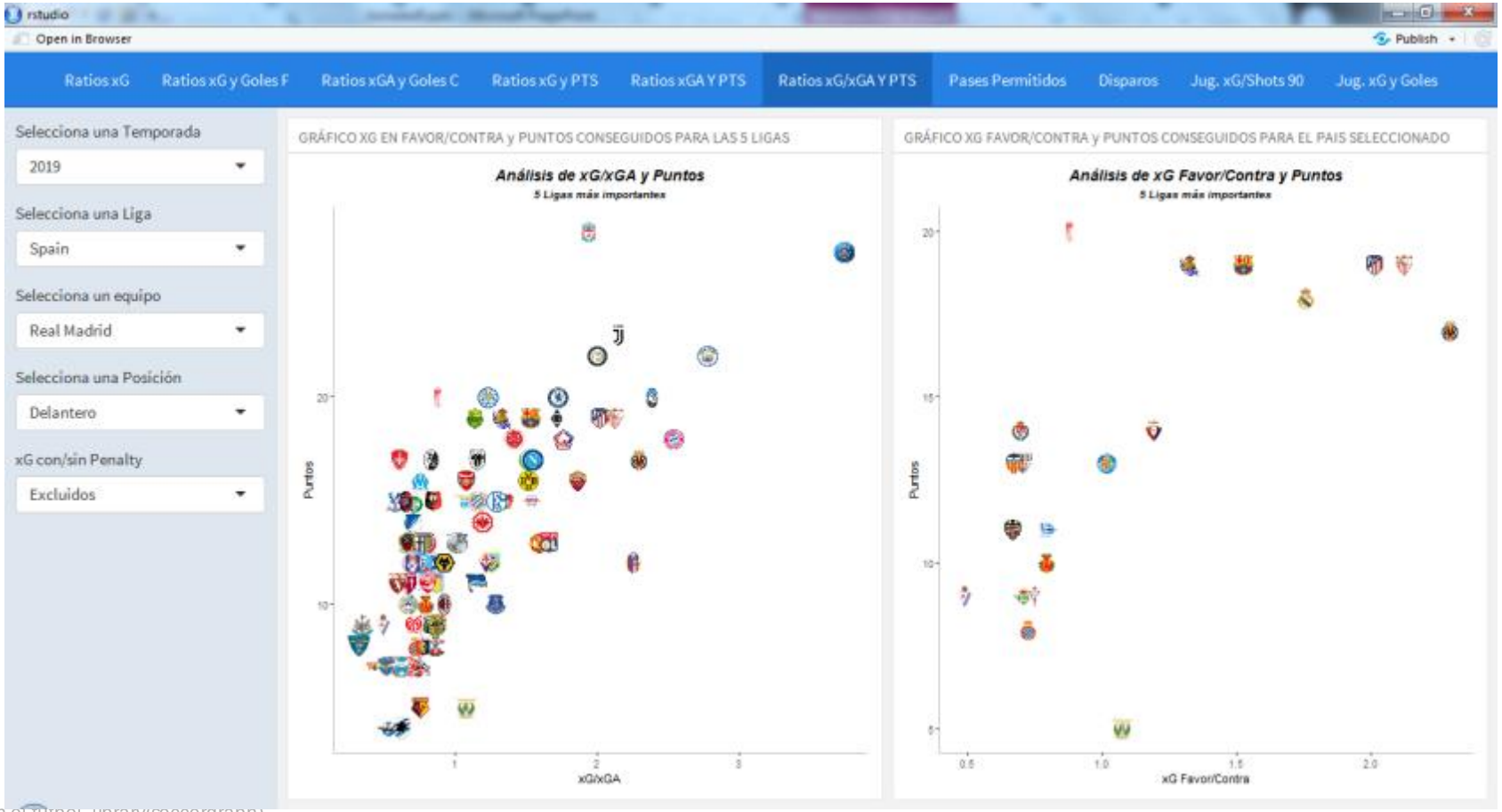
Aplicaciones del xG

- Rendimiento de un equipo en ataque y defensa
- Rendimiento de un jugador
- ¿quién se mereció ganar? ¿fue un resultado justo?



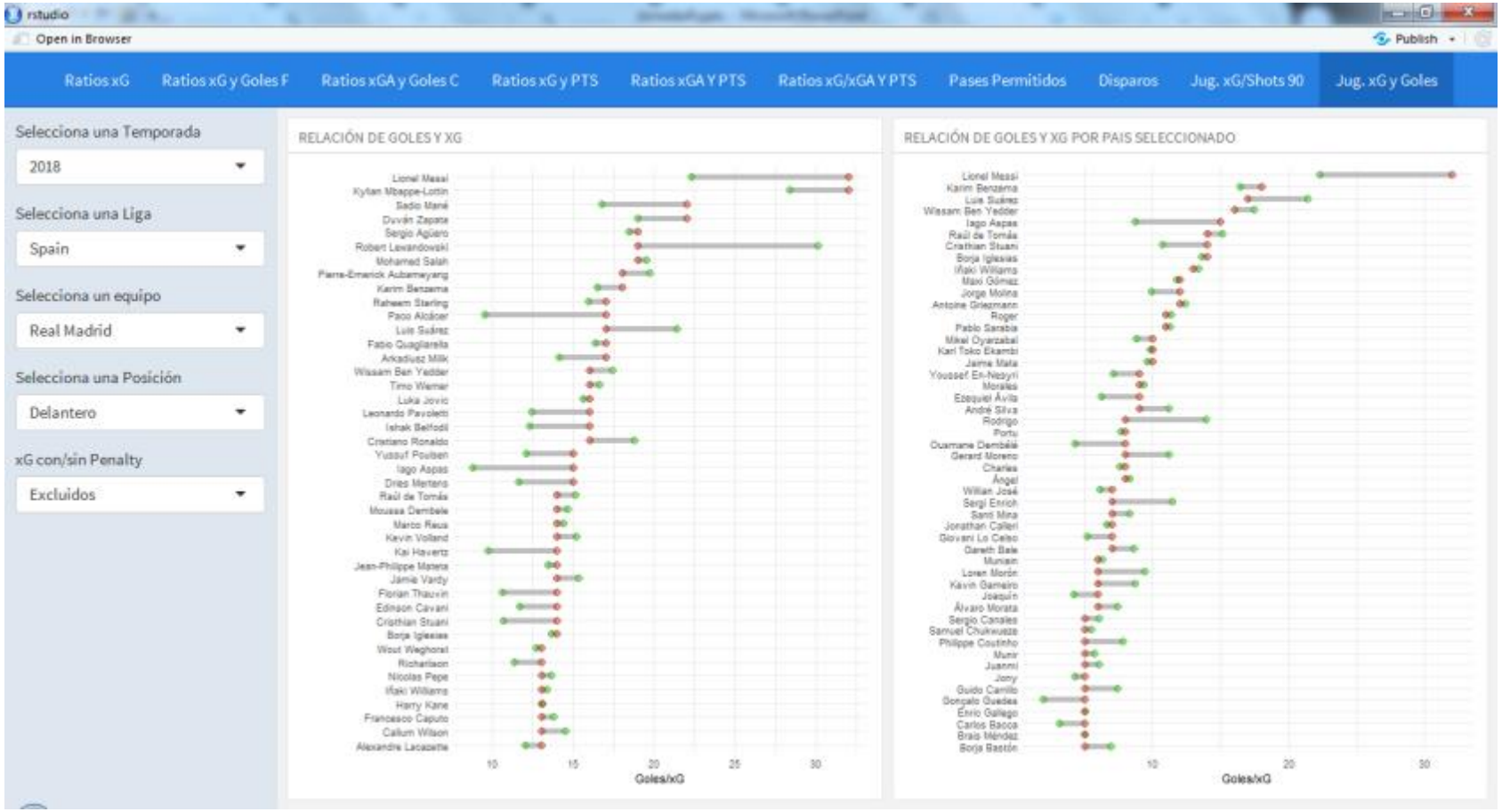
Aplicaciones del xG

- Rendimiento de un equipo en ataque y defensa
- Rendimiento de un jugador
- ¿quién se mereció ganar? ¿fue un resultado justo?



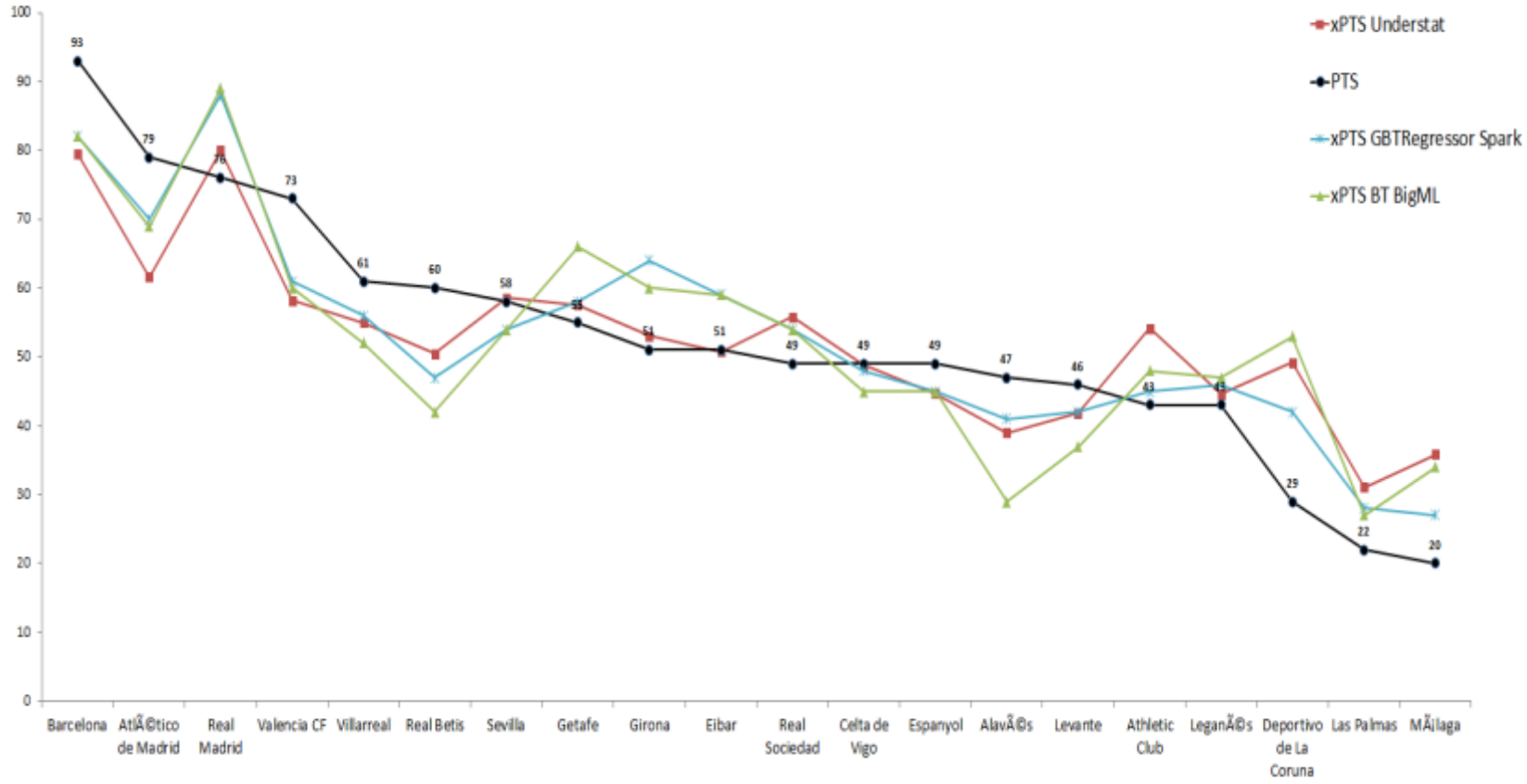
Aplicaciones del xG

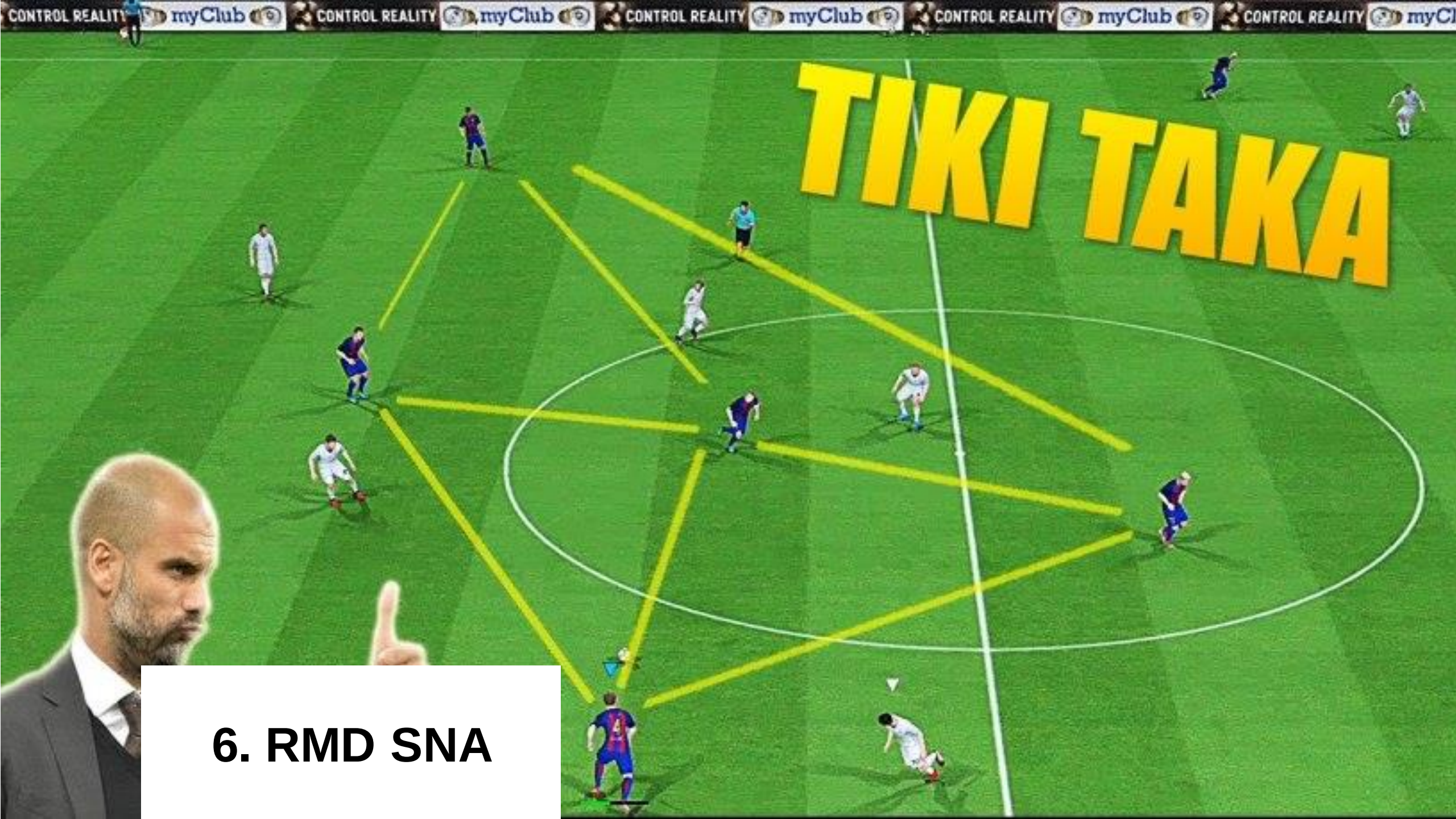
- Rendimiento de un equipo en ataque y defensa
- Rendimiento de un jugador
- ¿quién se mereció ganar? ¿fue un resultado justo?



Aplicaciones del xG

- Rendimiento de un equipo en ataque y defensa
- Rendimiento de un jugador
- ¿quién se mereció ganar? ¿fue un resultado justo?

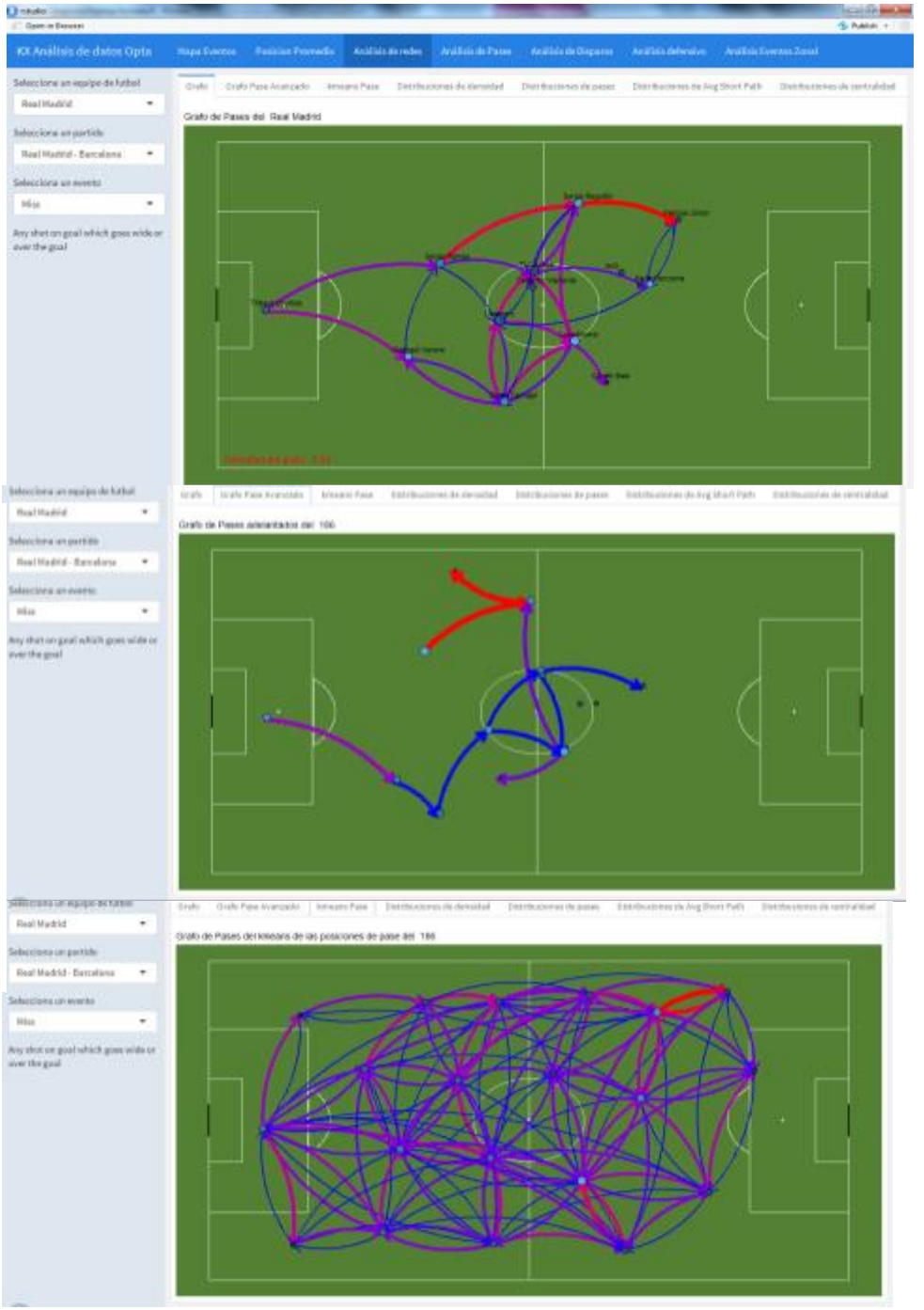




6. RMD SNA

Análisis de redes en el fútbol

- Métricas y grafos, pase avanzado y *kmeans*
- Densidades
- Centralidad
- Fútbol Masculino vs Femenino



Betweenness

The betweenness centrality of a node v is given by the expression:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v.

Density [edit]

The density D of a network is defined as a ratio of the number of edges E to the number of possible edges in a network with N nodes, given (in the case of simple graphs) by the binomial coefficient $\binom{N}{2}$, giving $D = \frac{E-(N-1)}{Emax-(N-1)} = \frac{2(E-N+1)}{N(N-3)+2}$ Another possible equation is $D = \frac{T-2N+2}{N(N-3)+2}$, whereas the ties T are unidirectional (Wasserman & Faust 1994) [2] This gives a better overview over the network density, because unidirectional relationships can be measured

Average shortest path length (or characteristic path length) [edit]

The average shortest path length is calculated by finding the shortest path between all pairs of nodes, and taking the average over all paths of the length thereof (the length being the number of intermediate edges contained in the path, i.e., the distance $d_{u,v}$ between the two vertices u, v within the graph). This shows us, on average, the number of steps it takes to get from one member of the network to another. The behavior of the expected average shortest path length (that is, the ensemble average of the average shortest path length) as a function of the number of vertices N of a random network model defines whether that model exhibits the small-world effect; if it scales as $O(\ln N)$, the model generates small-world nets. For faster-than-logarithmic growth, the model does not produce small worlds. The special case of $O(\ln \ln N)$ is known as ultra-small world effect.

Using network science to analyze football passing networks: dynamics, space, time and the multilayer nature of the game

J.M. Buldú, J. Busquets, J.H. Martínez, J.L. Herrera-Diestra, I. Echegoyen, J. Galeano, J. Luque

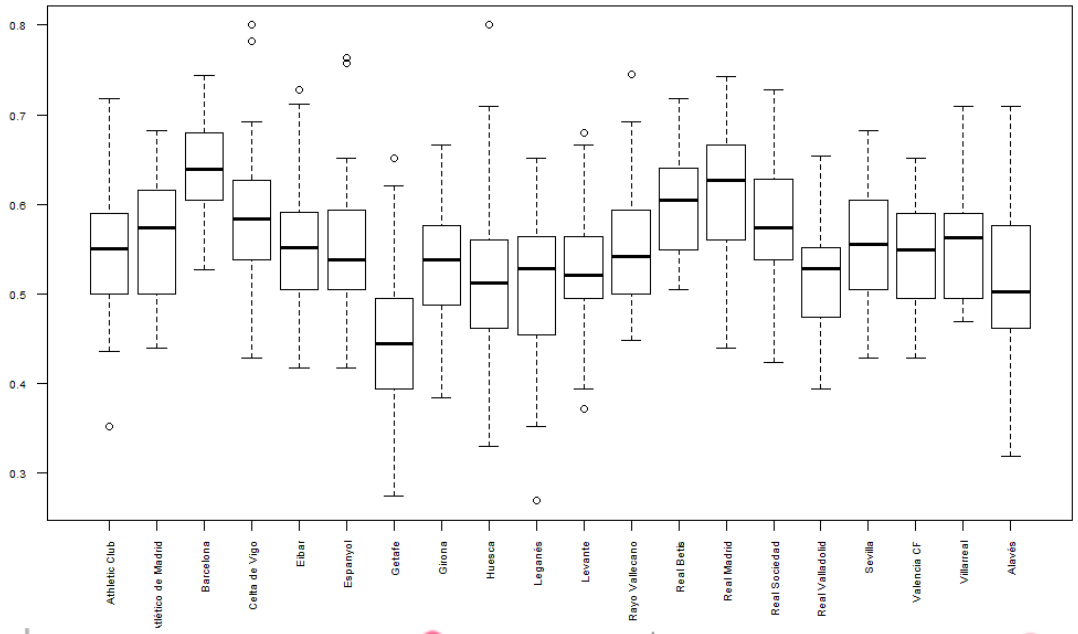
(Submitted on 2 Jul 2018 (v1), last revised 4 Jul 2018 (this version, v3))

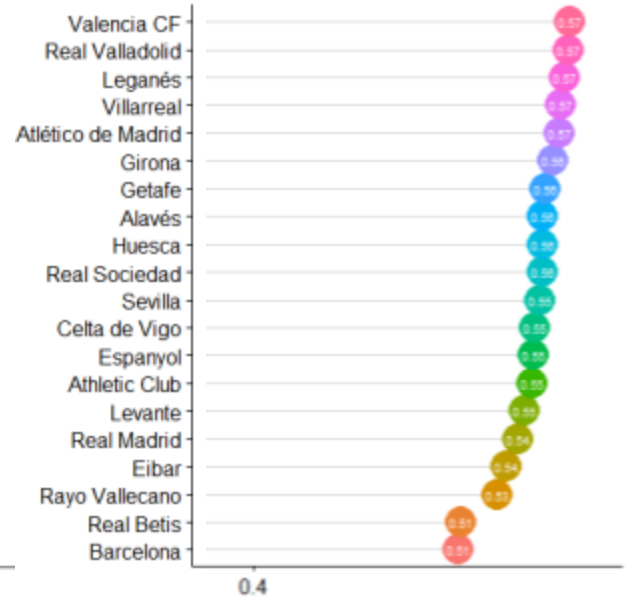
From the diversity of applications of Network Science, in this Opinion Paper we are concerned about its potential to analyze one of the most extended group sports: Football (soccer in U.S. terminology). As we will see, Network Science allows addressing different aspects of the team organization and performance not captured by classical analyses based on the performance of individual players. The reason behind relies on the complex nature of the game, which, paraphrasing the foundational paradigm of complexity sciences "can not be analyzed by looking at its components (i.e., players) individually but, on the contrary, considering the system as a whole" or, in the classical words of after-match interviews "it's not just me, it's the team".

Comments: 7 pages, 1 figure
Subjects: **Physics and Society (physics.soc-ph)**
Cite as: arXiv:1807.00534 [physics.soc-ph]
(or arXiv:1807.00534v3 [physics.soc-ph] for this version)

Análisis de redes en el fútbol

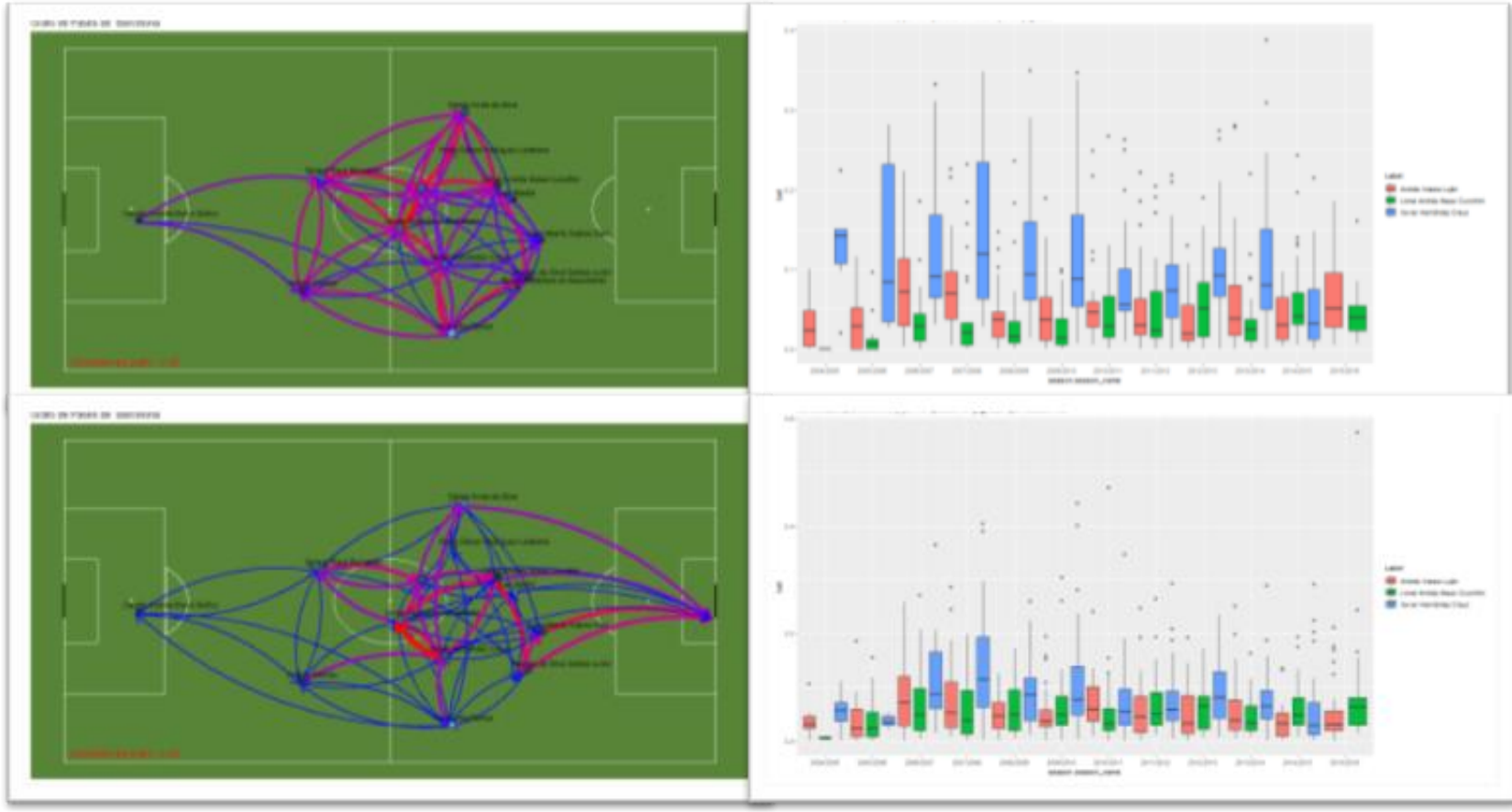
- Métricas y grafos, pase avanzado y *kmeans*
- Densidades
- Centralidad
- Fútbol Masculino vs Femenino





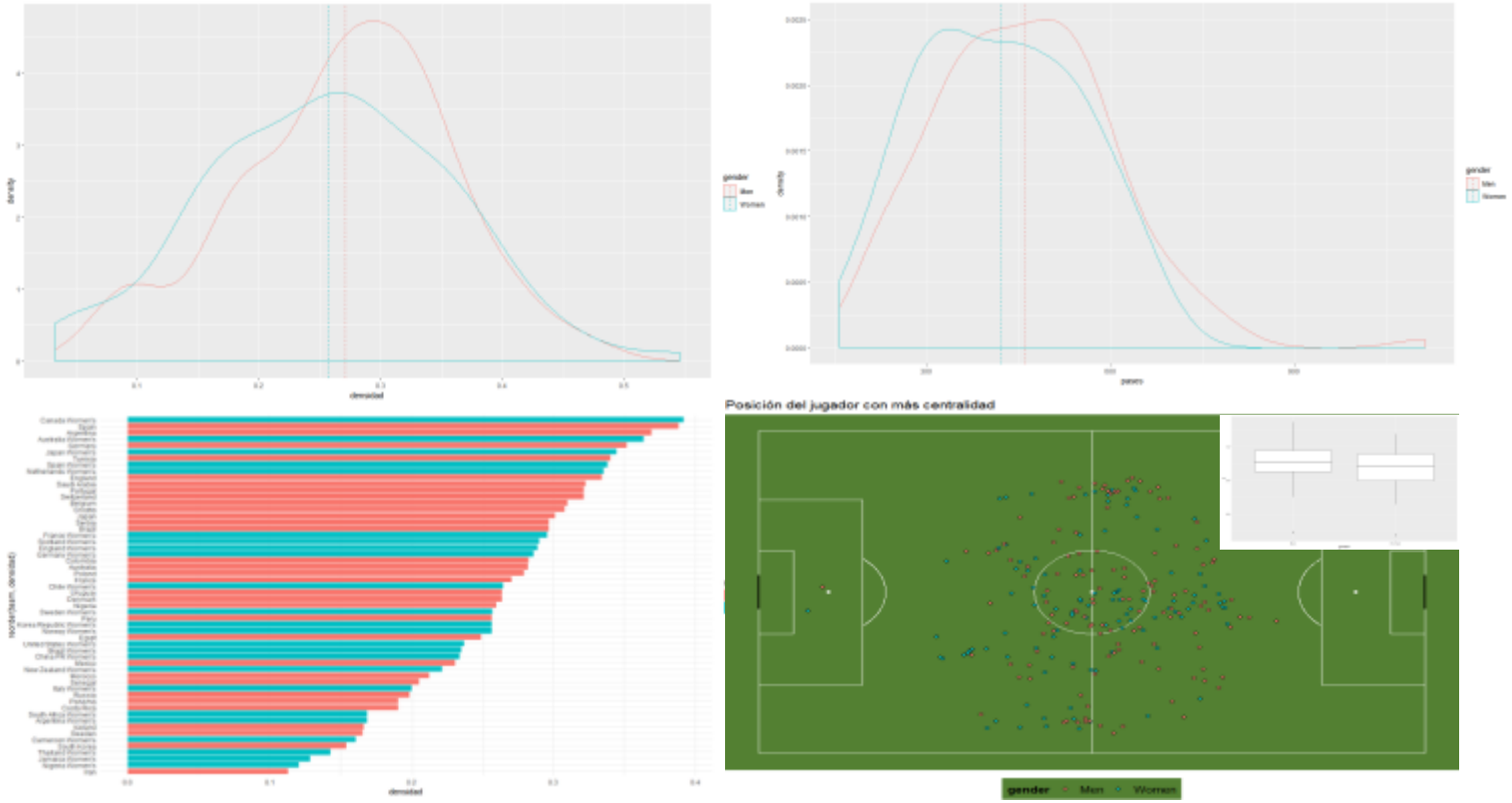
Análisis de redes en el fútbol

- Métricas y grafos, pase avanzado y *kmeans*
- Densidades
- Centralidad
- Fútbol Masculino vs Femenino

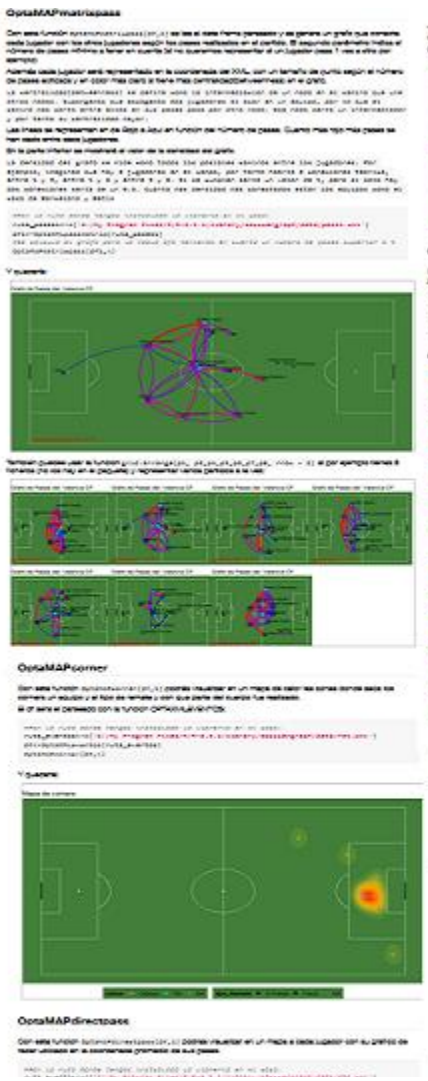


Análisis de redes en el fútbol

- Métricas y grafos, pase avanzado y *kmeans*
- Densidades
- Centralidad
- Fútbol Masculino vs Femenino

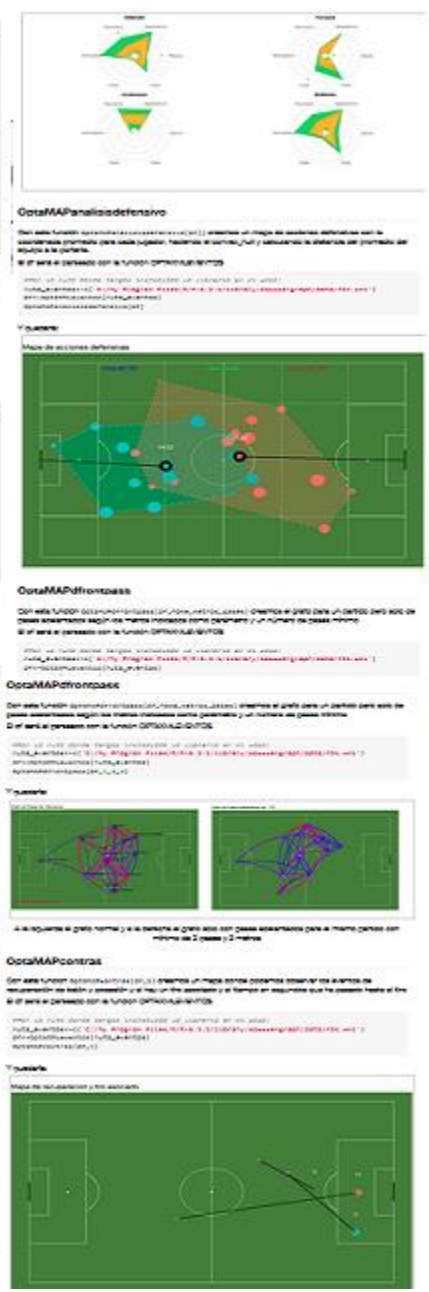


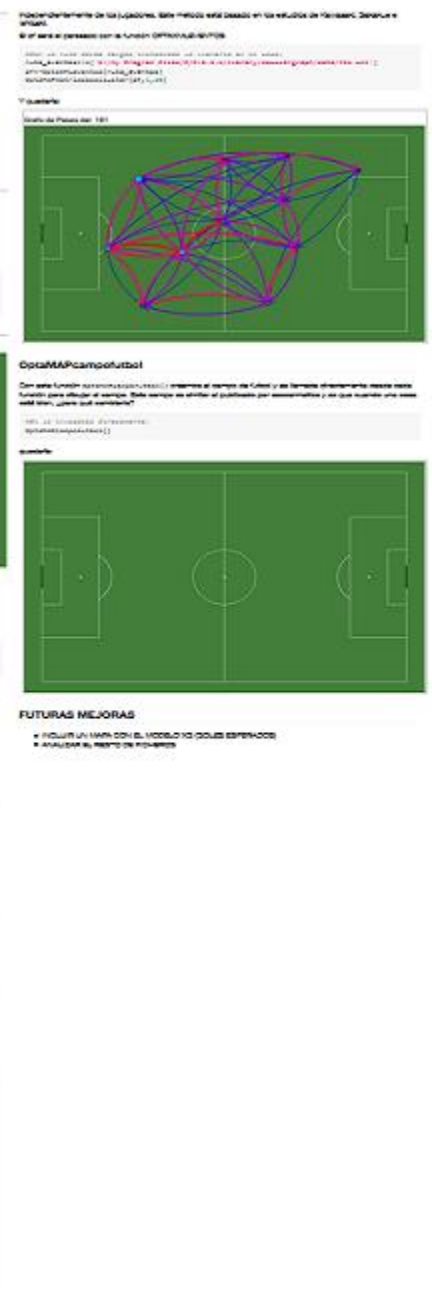

SoccerGraph
Instalación y Ayuda
Esta funcion crea el grafo de pases para un equipo durante un partido cuando los pases son avanzados
Description
Usage
Argumento
OptaXMLeventos
OPTAXMLPASSMATRIX


OptaMAPmatrixpass
OptaMAPcorner
OptaMAPdirectpass


OptaMAPheatMapFrom y OptaMAPheatMapTo
Zonas desde donde pasan los jugadores
Zonas donde reciben los jugadores
OptaMAPshoot
OPTAXMLPOSESION


Mapa de posesiones
OptaMAPpossessionzonal
OptaXMLestats
OptaMAPradarevent

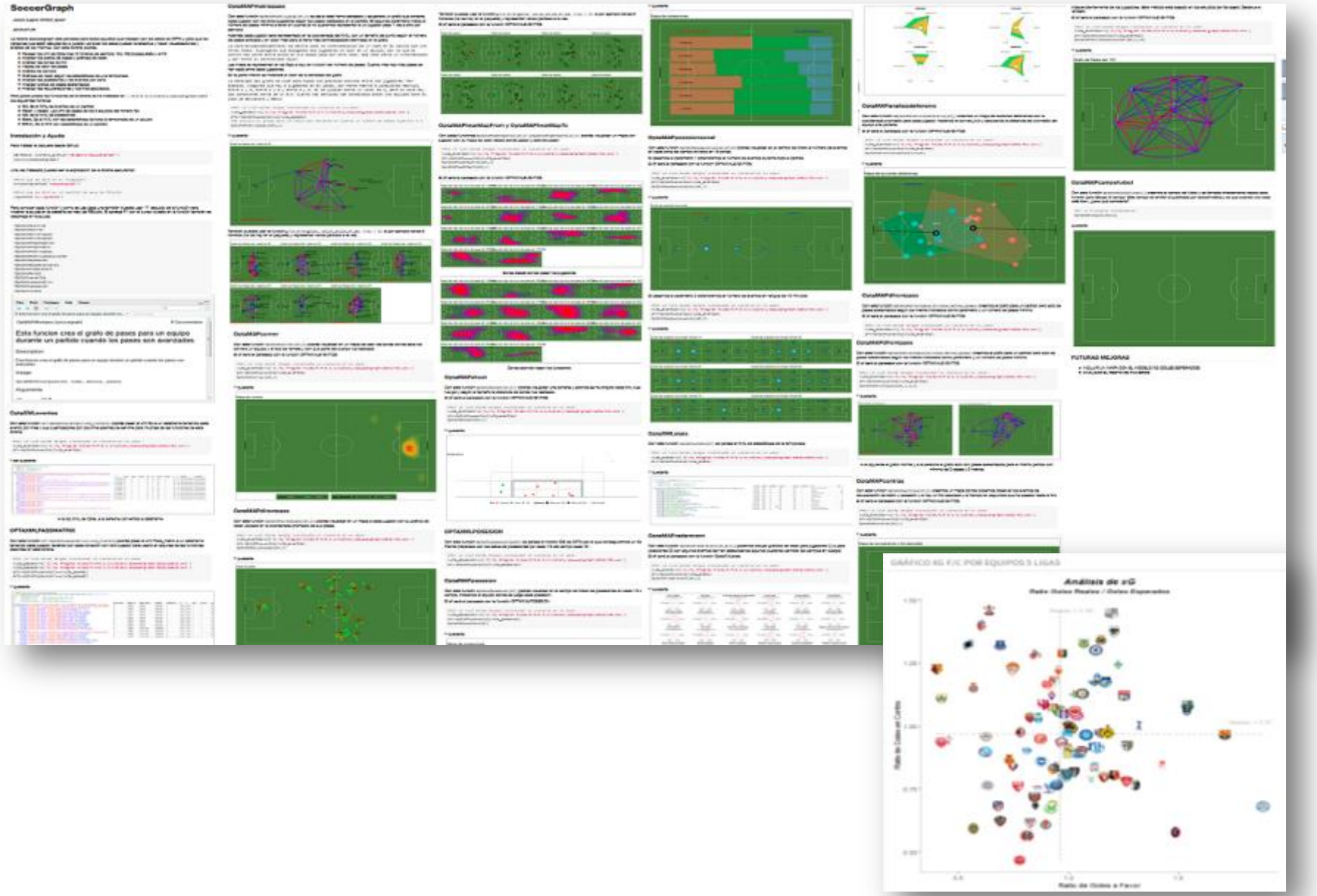

OptaMAPanalisisdefensivo
Mapa de acciones defensivas
OptaMAPdfrontpass
OptaMAPdfrontpass
OptaMAPcontras


OptaMAPcampofutbol
FUTURAS MEJORAS

7. Paquetes

Paquetes para iniciarse

- soccergraphR
- FootballBadges
- UnderstatPlots (próximamente)
- statsbombR
- FCrSTATS
- soccermatics
- ggsoccer



¡Muchas gracias!

16 NOVIEMBRE, 2019

Análisis de datos en el fútbol

library(soccergraphR)
library(FootballBadges)

Jesús Lagos Milla

@Vdot_Spain

#XIJRes #Futbol

https://github.com/Jelagmil/