Spanish basketball analytics

Guillermo Vinué
Mathematician. PhD in Statistics
Data Analyst

July 7, 2023

Summary

Clarification

- Statistics: Branch of mathematics devoted to the analysis of data of any type.

- Stats: Numerical data representing teams' and players' performance.

Statistics allows us to analyze stats...

Motivation

- Love of sports, in particular, of basketball.

- Interest in joining my hobby with my profession.

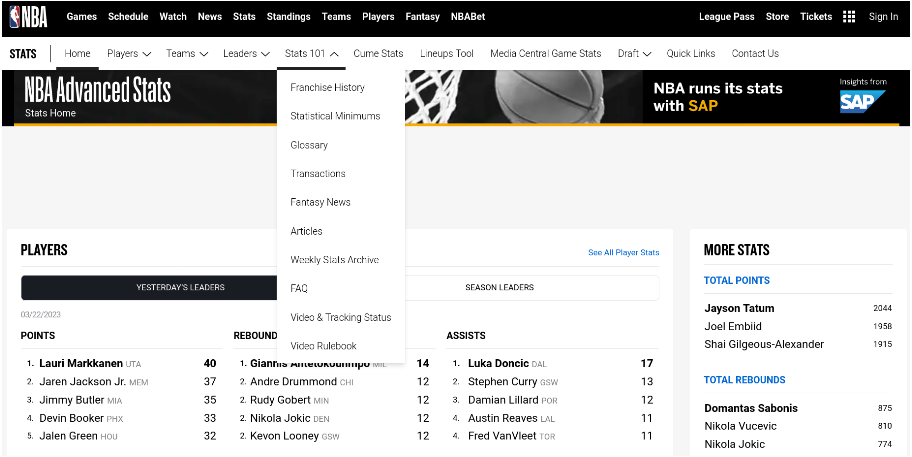
- Basketball generates a huge amount of data (*stats*), both individual and as a team.

- In the NBA (USA male professional basketball league), data collection and analysis has been usual practice for a long time.


32 - SHAQUILLE O'NEAL
FIELD GOALS
6/6
POINTS
14
BROKEN BACKBOARDS
1

- The NBA has an excellent website where a lot of information can be examined.

www.nba.com/stats



- In Spain, the treatment of data at the official level is still very rudimentary.
  - The male professional league is the ACB (Asociación de Clubes de Baloncesto, in Spanish) (for sponsorship reasons it is known as Liga Endesa) and is the one providing the audience with stats.
  - Basic stats, that, although informative, can be expanded.
  - Tabular representation, without option of interacting. Absence of visualizations.

ESTADÍSTICAS

VALENCIA BASKET

TEMPORADA 2022-2023 | LIGA ENDESA | TODAS LAS FASES | MEDIA

Liga Endesa - TEMPORADA 2022-2023 - TODAS LAS FASES

	PAR	MIN		PT	TIROS DE 3			TIROS DE 2			TIROS LIBRES			REBOTES			ASI	BALONES		TAPONES			FALTAS			VAL
	Jug	Jug	5i		Con	Int	%	Con	Int	%	Con	Int	%	Def	Ofe	Tot	Efe	Rec	Per	Fav	Con	Mat	Com	Rec	+/-	
Jasiel Rivero	18	21:21	0,2	14,4	0,6	1,1	55,0%	4,7	6,7	70,0%	3,2	5,3	60,4%	2,5	2,0	4,5	0,7	0,7	1,0	0,8	0,1	0,3	2,1	4,4	4,0	17,8
Bojan Dubljevic	22	21:34	0,9	9,9	1,4	3,7	37,0%	1,9	3,5	53,2%	2,0	2,5	83,3%	4,1	2,0	6,0	1,5	0,8	1,6	0,3	0,5	0,0	1,7	2,9	1,0	13,2
Chris Jones	25	22:04	0,8	12,2	1,2	3,0	40,5%	3,4	6,3	54,1%	1,8	2,1	84,6%	1,6	0,2	1,8	4,8	1,0	2,2	0,0	0,3	0,0	1,5	2,2	3,0	13,0
Xabi López-Arostegui	25	20:16	0,3	8,6	1,0	2,9	36,1%	2,0	3,7	52,7%	1,6	1,8	88,6%	2,8	0,7	3,5	1,4	1,5	1,2	0,1	0,2	0,0	1,9	2,5	0,0	10,5
James Webb III	18	19:42	0,7	9,1	1,7	4,4	37,5%	1,7	2,6	63,8%	0,8	1,0	77,8%	4,4	1,1	5,5	0,4	0,8	1,1	0,4	0,2	0,4	2,3	1,3	1,0	10,1
Jared Harper	13	15:11	0,1	8,6	1,4	3,5	40,0%	0,8	2,4	32,3%	2,9	3,3	88,4%	0,5	0,1	0,5	2,4	0,5	1,6	0,1	0,2	0,0	2,0	2,9	-2,0	7,2
Sam Van Rossom	6	16:39	0,3	7,8	1,5	3,7	40,9%	0,8	1,7	50,0%	1,7	1,8	90,9%	1,0	0,0	1,0	1,8	0,3	1,0	0,0	0,3	0,0	1,2	1,5	-4,0	6,8
Klemen Prepelic	21	16:16	0,2	6,2	0,9	3,9	22,0%	0,8	1,7	48,6%	2,0	2,6	76,4%	1,0	0,2	1,2	2,9	0,3	1,6	0,0	0,2	0,0	1,3	2,8	0,0	5,8
Jaime Pradilla	23	15:02	0,3	5,1	0,4	1,0	43,5%	1,6	3,3	48,1%	0,6	0,8	73,7%	1,2	1,1	2,3	1,0	0,7	0,7	0,1	0,4	0,2	1,1	1,0	0,0	5,6
Kyle Alexander	21	14:29	0,2	5,2	0,0	0,4	11,1%	2,1	3,1	66,7%	0,9	1,4	63,3%	1,7	1,0	2,7	0,4	0,3	0,9	0,6	0,2	0,9	2,1	1,3	-2,0	5,3
Shannon Evans	10	16:50	0,0	5,9	0,7	1,9	36,8%	1,3	2,7	48,1%	1,2	1,8	66,7%	1,4	0,0	1,4	2,4	0,8	1,9	0,1	0,0	0,1	1,3	1,5	-6,0	5,7
Jonah Radebaugh	26	19:02	0,7	6,6	1,0	3,0	32,5%	1,3	2,7	50,7%	1,0	1,1	89,7%	1,3	0,3	1,6	1,2	0,7	0,8	0,1	0,2	0,0	2,0	1,0	0,0	4,8
Víctor Claver	23	18:26	0,3	4,0	0,8	2,0	39,1%	0,7	1,6	40,5%	0,3	0,3	87,5%	2,2	0,5	2,7	0,8	0,5	0,8	0,3	0,1	0,1	1,6	0,9	-1,0	4,6
Martin Hermannsson	5	17:30	0,2	2,6	0,4	2,0	20,0%	0,4	2,0	20,0%	0,6	0,6	100,0%	1,4	0,0	1,4	3,6	0,6	1,0	0,0	0,0	0,0	0,8	1,0	0,0	4,2
Josep Puerto	25	16:10	0,7	5,2	0,8	2,4	34,4%	1,1	1,6	65,9%	0,5	0,7	76,5%	1,4	0,5	2,0	0,4	0,4	0,6	0,0	0,2	0,1	1,8	0,8	0,0	4,0
Guillem Ferrando	8	12:54	0,3	3,9	1,1	2,5	45,0%	0,1	0,1	100,0%	0,3	0,4	66,7%	0,4	0,4	0,8	1,9	0,5	1,4	0,0	0,0	0,0	2,6	0,4	3,0	1,9
Millán Jiménez	0	0:00	0,0	0,0	0,0	0,0	0,0%	0,0	0,0	0,0%	0,0	0,0	0,0%	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Sergio De Larrea	0	0:00	0,0	0,0	0,0	0,0	0,0%	0,0	0,0	0,0%	0,0	0,0	0,0%	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Gonzalo Bressan	1	1:54	0,0	0,0	0,0	0,0	0,0%	0,0	1,0	0,0%	0,0	0,0	0,0%	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,0	0,0	-7,0	-2,0
Lucas Marí	3	1:00	0,0	0,0	0,0	0,7	0,0%	0,0	0,0	0,0%	0,0	0,0	0,0%	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,3	0,0	0,0	-1,0
Totales	26			83,5	10,2	28,6	35,7%	19,0	34,6	55,1%	14,8	19,2	77,2%	23,5	9,5	32,9	17,5	10,5	13,4	2,2	2,4	1,7	19,5	20,5		93,5

- Spain is a world basketball power, placed on the same level as the NBA (current European and world champions).

- But in the world of stats, the NBA is still holding a big advantage over us. If they analyze their data in an advanced way...

Why not do it too?

- Chance of innovation with the information generated from each game:
  - **New representation** of the **traditional stats.**
  - **New analyses** from the **play-by-play data.**
- Importance of visualizing data: Identify trends and patterns in data easily.
- Software tool: R, www.r-project.org

A short comment about R

python

We use SAS here...

We are looking for Tableau users...


WTF?!

- R is a programming language **specifically** developed to carry out statistical analyses of data.
- It is based on a system of packages, which are a collection of functions developed by and for the users community.
- They are mostly collected in the CRAN repository (The Comprehensive R Archive Network). Around 20.000 packages[1].
- The best statisticians of the world use it (Rob Tibshirani, Julia Silge, Rob Hyndman, etc).
- It's free.

[1] https://cran.r-project.org/web/packages/

Example of R packages

- tidyverse: Data processing and mining.
- tidymodels: Models and predictive systems with *machine learning*.
- shiny, flexdashboard: Dashboards for data visualization.
- roxygen2: Development of personal packages.
- RMarkdown: Writing of documents (interactive and with LaTeX).

I did not know that we could do so much with R...

Data collection

Type of data in the ACB

- Traditional stats: Summary of the players' activity in several game aspects (points, rebounds, etc) **once the game is finished**.

- Play-by-play data: Transcription of every event (player, action and time point (minute and second)) that happens **during the game**.

1. **Traditional stats.**

2. Play-by-play data.

Traditional stats

- They are provided as static tables (*box score data*).
- There is no downloading option (as an Excel file, for instance).
- Unfeasible copy and paste.

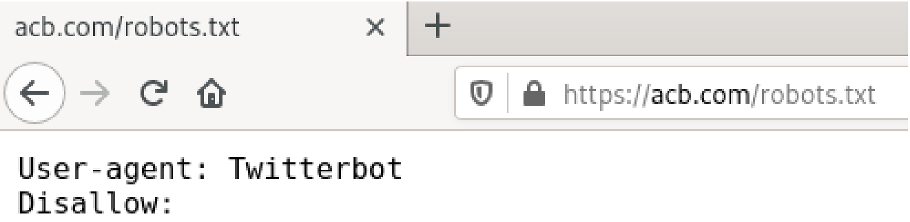
http://www.acb.com/partido/estadisticas/id/103170

Valencia Basket 75

										REB						TAP			FP			
D	**Nombre**	**Min**	**P**	**T2**	**T2 %**	**T3**	**T3 %**	**T1**	**T1 %**	**T**	**D+O**	**A**	**BR**	**BP**	**C**	**F**	**C**	**M**	**F**	**C**	**+/-**	**V**
1	V. Claver	21:55	14	0/2	0%	4/4	100%	2/2	100%	7	5+2	0	0	0	0	0	0	0	1	2	-1	20
*2	J. Puerto	15:19	7	2/4	50%	1/4	25%	0/0	0%	1	1+0	1	1	1	2	0	1	0	1	0	1	2
4	J. Pradilla	07:41	4	2/3	67%	0/1	0%	0/1	0%	0	0+0	0	0	1	0	0	0	0	1	1	-2	0
*5	J. Webb	14:50	0	0/1	0%	0/3	0%	0/0	0%	3	2+1	1	1	3	0	0	1	0	5	0	-4	-8
6	López-Arostegui	19:41	5	0/2	0%	1/5	20%	2/2	100%	4	4+0	1	1	1	0	0	1	0	2	2	-10	3
7	C. Jones	18:44	14	3/5	60%	2/5	40%	2/4	50%	2	2+0	2	0	2	0	0	0	0	1	2	-9	10
*12	J. Radebaugh	22:18	4	1/1	100%	0/3	0%	2/2	100%	4	2+2	0	1	0	0	0	0	0	2	2	-5	6
13	S. Evans	14:34	4	2/2	100%	0/1	0%	0/0	0%	1	1+0	2	1	0	0	0	0	0	1	0	-10	6
*14	B. Dubljevic	26:31	15	3/3	100%	2/6	33%	3/4	75%	3	1+2	2	3	1	0	0	0	0	3	5	-7	19
*24	M. Hermannsson	19:03	2	1/4	25%	0/2	0%	0/0	0%	2	2+0	6	0	1	0	0	0	0	2	2	-1	4
41	J. Rivero	19:24	6	2/4	50%	0/1	0%	2/5	40%	4	2+2	1	1	2	0	1	0	0	1	4	-7	8
55	L. Marí																					
	Equipo									2	0+2		3	1					0	0		4
	Total	200:00	75	16/31	52%	10/35	29%	13/20	65%	33	22+11	16	12	13	2	1	3	0	20	20	-55	74
E	Álex Mumbrú																					
5f		J. Webb																				

- Solution: Turn to the so-called *friendly web scraping*.
- *Web scraping* is a data analysis technique that allows us to extract information from websites by using software programs.
- But... what does *friendly* mean?

- Before getting data, we must check if there is permission for that.
- In the **robots.txt** file, the web owner specifies who can extract data and who can not.



- Once we know that we have permission to get the data, we must find out if data are available in *html* format.
- Procedure: Right mouse button → View of web source code:

```
    <td class="dorsal">1</td>
    <td class="nombre jugador ellipsis"><a href="/jugador/ver/20201901-V.-Claver">V. Claver</a></td>
    <td class="">21:55</td>
    <td class="grisclaro">14</td>
    <td class="">0/2</td>
    <td class="">0%</td>
    <td class="grisclaro">4/4</td>
    <td class="grisclaro">100%</td>
    <td class="">2/2</td>
    <td class="">100%</td>
    <td class="grisclaro">7</td>
    <td class="grisclaro">5+2</td>
    <td class="">0</td>
    <td class="grisclaro" >0</td>
    <td class="">0</td>
    <td class="grisclaro">0</td>
    <td class="">0</td>
    <td class="">0</td>
    <td class="grisclaro">0</td>
    <td class="">1</td>
    <td class="grisclaro">2</td>
    <td class="">-1</td>
    <td class="grisclaro">20</td>
</tr>
 <tr>
    <td class="dorsal">*2</td>
    <td class="nombre jugador ellipsis"><a href="/jugador/ver/20210392-J.-Puerto">J. Puerto</a></td>
    <td class="">15:19</td>
    <td class="grisclaro">7</td>
    <td class="">2/4</td>
    <td class="">50%</td>
    <td class="grisclaro">1/4</td>
    <td class="grisclaro">25%</td>
    <td class="">0/0</td>
    <td class="">0%</td>
```

- Once we know that the *html* format is available, the **rvest**[2] R package is providing us with the suitable functions.

```
library(rvest)
url_html <- read_html("http://www.acb.com/partido/estadisticas/id/103170")
url_html %>%
  html_table(fill = TRUE) %>%
  `[[`(3)
```

[2] https://CRAN.R-project.org/package=rvest

data_acb

Filter

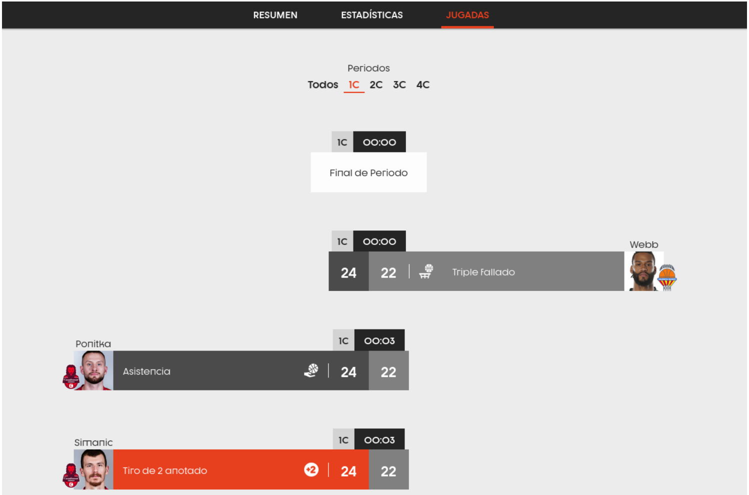
											REB	REB					TAP	TAP		FP	FP		
1	D	Nombre	Min	P	T2	T2 %	T3	T3 %	T1	T1 %	T	D+O	A	BR	BP	C	F	C	M	F	C	+/-	V
2	1	V. Claver	21:55	14	0/2	0%	4/4	100%	2/2	100%	7	5+2	0	0	0	0	0	0	0	1	2	-1	20
3	*2	J. Puerto	15:19	7	2/4	50%	1/4	25%	0/0	0%	1	1+0	1	1	1	2	0	1	0	1	0	1	2
4	4	J. Pradilla	07:41	4	2/3	67%	0/1	0%	0/1	0%	0	0+0	0	0	1	0	0	0	0	1	1	-2	0
5	*5	J. Webb	14:50	0	0/1	0%	0/3	0%	0/0	0%	3	2+1	1	1	3	0	0	1	0	5	0	-4	-8
6	6	López-Arostegui	19:41	5	0/2	0%	1/5	20%	2/2	100%	4	4+0	1	1	1	0	0	1	0	2	2	-10	3
7	7	C. Jones	18:44	14	3/5	60%	2/5	40%	2/4	50%	2	2+0	2	0	2	0	0	0	0	1	2	-9	10
8	*12	J. Radebaugh	22:18	4	1/1	100%	0/3	0%	2/2	100%	4	2+2	0	1	0	0	0	0	0	2	2	-5	6
9	13	S. Evans	14:34	4	2/2	100%	0/1	0%	0/0	0%	1	1+0	2	1	0	0	0	0	0	1	0	-10	6
10	*14	B. Dubljevic	26:31	15	3/3	100%	2/6	33%	3/4	75%	3	1+2	2	3	1	0	0	0	0	3	5	-7	19
11	*24	M. Hermannsson	19:03	2	1/4	25%	0/2	0%	0/0	0%	2	2+0	6	0	1	0	0	0	0	2	2	-1	4
12	41	J. Rivero	19:24	6	2/4	50%	0/1	0%	2/5	40%	4	2+2	1	1	2	0	1	0	0	1	4	-7	8
13	55	L. Marí																					
14		Equipo									2	0+2		3	1					0	0		4
15		Total	200:00	75	16/31	52%	10/35	29%	13/20	65%	33	22+11	16	12	13	2	1	3	0	20	20	-55	74

1. Traditional stats.
2. **Play-by-play data.**

Play-by-play data

- They are provided as an event sequence.
- There is no downloading option (as an Excel file, for instance).
- Unfeasible copy and paste.

http://jv.acb.com/es/103170/jugadas



- In this case, the *html* format is not available. Instead of using **rvest**, the suitable R package is **RSelenium**[3].
- The procedure is tougher, but it works.

```
library(RSelenium)
# Abre un servidor:
rD <- rsDriver(browser = "firefox", chromever = NULL)
# Copia el enlace del partido http://jv.acb.com/es/103170/jugadas
# Trabaja cuarto a cuarto. Empieza desde 1C, ve con el ratón hasta al final
# de la página y entonces, ejecuta este código:
remDr <- rD$client  # Driver remoto.

# Datos jugada a jugada:
game_elem <- remDr$getPageSource()[[1]]

# Cierra el cliente y el servidor:
remDr$close()
rD$server$stop()
```

[3]https://CRAN.R-project.org/package=RSelenium

data_acb_pbp

Filter

	period	time_point	player	action	local_score	visitor_score	team
1	1C	00:00	Webb	Triple fallado	24	22	Valencia Basket
2	1C	00:03	Simanic	Tiro de 2 anotado	24	22	Casademont Zaragoza
3	1C	00:03	Ponitka	Asistencia	24	22	Casademont Zaragoza

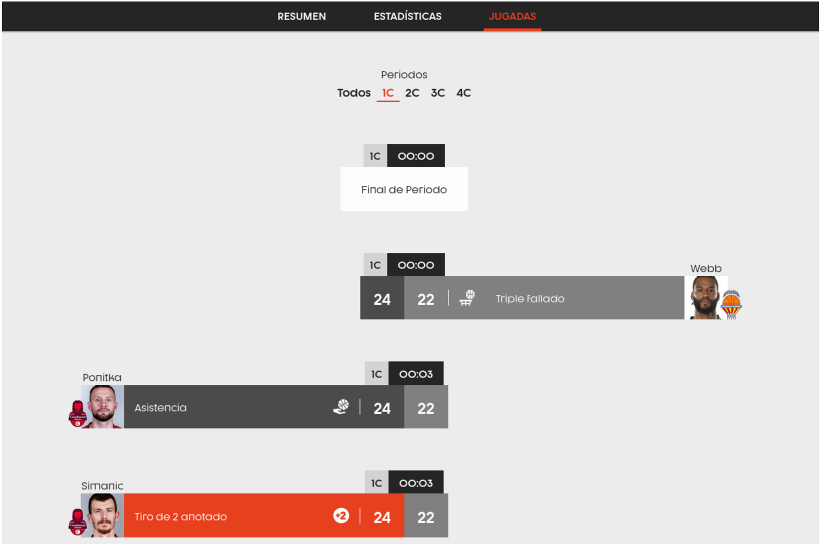
As a summary...

BEFORE

Valencia Basket 75

										REB						TAP			FP			
D	Nombre	Min	P	T2	T2 %	T3	T3 %	T1	T1 %	T	D+O	A	BR	BP	C	F	C	M	F	C	+/-	V
1	V. Claver	21:55	14	0/2	0%	4/4	100%	2/2	100%	7	5+2	0	0	0	0	0	0	0	1	2	-1	20
*2	J. Puerto	15:19	7	2/4	50%	1/4	25%	0/0	0%	1	1+0	1	1	1	2	0	1	0	1	0	1	2
4	J. Pradilla	07:41	4	2/3	67%	0/1	0%	0/1	0%	0	0+0	0	0	1	0	0	0	0	1	1	-2	0
*5	J. Webb	14:50	0	0/1	0%	0/3	0%	0/0	0%	3	2+1	1	1	3	0	0	1	0	5	0	-4	-8
6	López-Arostegui	19:41	5	0/2	0%	1/5	20%	2/2	100%	4	4+0	1	1	1	0	0	1	0	2	2	-10	3
7	C. Jones	18:44	14	3/5	60%	2/5	40%	2/4	50%	2	2+0	2	0	2	0	0	0	0	1	2	-9	10
*12	J. Radebaugh	22:18	4	1/1	100%	0/3	0%	2/2	100%	4	2+2	0	1	0	0	0	0	0	2	2	-5	6
13	S. Evans	14:34	4	2/2	100%	0/1	0%	0/0	0%	1	1+0	2	1	0	0	0	0	0	1	0	-10	6
*14	B. Dubljevic	26:31	15	3/3	100%	2/6	33%	3/4	75%	3	1+2	2	3	1	0	0	0	0	3	5	-7	19
*24	M. Hermannsson	19:03	2	1/4	25%	0/2	0%	0/0	0%	2	2+0	6	0	1	0	0	0	0	2	2	-1	4
41	J. Rivero	19:24	6	2/4	50%	0/1	0%	2/5	40%	4	2+2	1	1	2	0	1	0	0	1	4	-7	8
55	L. Marí																					
	Equipo									2	0+2		3	1					0	0		4
	Total	**200:00**	**75**	**16/31**	**52%**	**10/35**	**29%**	**13/20**	**65%**	**33**	**22+11**	**16**	**12**	**13**	**2**	**1**	**3**	**0**	**20**	**20**	**-55**	**74**
E	Álex Mumbrú																					
5f		J. Webb																				

BEFORE


RESUMEN
ESTADÍSTICAS
JUGADAS
Periodos
Todos 1C 2C 3C 4C
1C 00:00
Final de Periodo
1C 00:00
Webb
24 22
Triple Fallado
1C 00:03
Ponitka
Asistencia
24 22
1C 00:03
Simanic
Tiro de 2 anotado
24 22

BEFORE

R to the rescue!

AFTER

data_acb

											REB	REB					TAP	TAP		FP	FP		
1	D	Nombre	Min	P	T2	T2 %	T3	T3 %	T1	T1 %	T	D+O	A	BR	BP	C	F	C	M	F	C	+/-	V
2	1	V. Claver	21:55	14	0/2	0%	4/4	100%	2/2	100%	7	5+2	0	0	0	0	0	0	0	1	2	-1	20
3	*2	J. Puerto	15:19	7	2/4	50%	1/4	25%	0/0	0%	1	1+0	1	1	1	2	0	1	0	1	0	1	2
4	4	J. Pradilla	07:41	4	2/3	67%	0/1	0%	0/1	0%	0	0+0	0	0	1	0	0	0	0	1	1	-2	0
5	*5	J. Webb	14:50	0	0/1	0%	0/3	0%	0/0	0%	3	2+1	1	1	3	0	0	1	0	5	0	-4	-8
6	6	López-Arostegui	19:41	5	0/2	0%	1/5	20%	2/2	100%	4	4+0	1	1	1	0	0	1	0	2	2	-10	3
7	7	C. Jones	18:44	14	3/5	60%	2/5	40%	2/4	50%	2	2+0	2	0	2	0	0	0	0	1	2	-9	10
8	*12	J. Radebaugh	22:18	4	1/1	100%	0/3	0%	2/2	100%	4	2+2	0	1	0	0	0	0	0	2	2	-5	6
9	13	S. Evans	14:34	4	2/2	100%	0/1	0%	0/0	0%	1	1+0	2	1	0	0	0	0	0	1	0	-10	6
10	*14	B. Dubljevic	26:31	15	3/3	100%	2/6	33%	3/4	75%	3	1+2	2	3	1	0	0	0	0	3	5	-7	19
11	*24	M. Hermannsson	19:03	2	1/4	25%	0/2	0%	0/0	0%	2	2+0	6	0	1	0	0	0	0	2	2	-1	4
12	41	J. Rivero	19:24	6	2/4	50%	0/1	0%	2/5	40%	4	2+2	1	1	2	0	1	0	0	1	4	-7	8
13	55	L. Marí																					
14		Equipo									2	0+2		3	1					0	0		4
15		Total	200:00	75	16/31	52%	10/35	29%	13/20	65%	33	22+11	16	12	13	2	1	3	0	20	20	-55	74

AFTER

data_acb_pbp

Filter

	period	time_point	player	action	local_score	visitor_score	team
1	1C	00:00	Webb	Triple fallado	24	22	Valencia Basket
2	1C	00:03	Simanic	Tiro de 2 anotado	24	22	Casademont Zaragoza
3	1C	00:03	Ponitka	Asistencia	24	22	Casademont Zaragoza

AFTER

Descriptive statistics

Data processing, visualization and analysis

1. **Traditional stats.**

2. Play-by-play data.

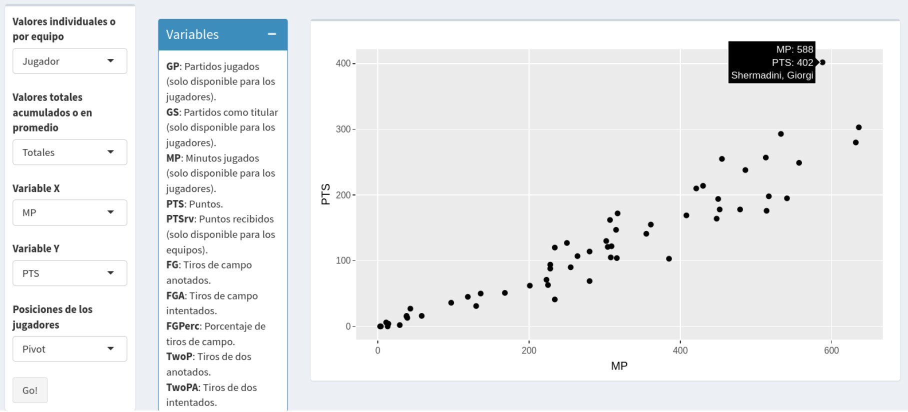
New representation of the traditional stats

https://www.uv.es/vivigui/AppEuroACB.html

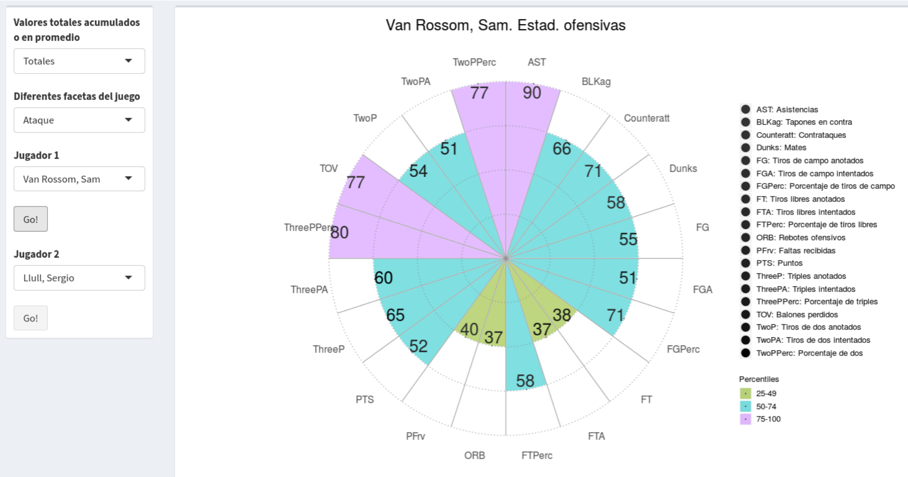
- Development of the first web application for the interactive visualization of European basketball box score data. Created in 2020. Available in Spanish and English.

- Players' and teams' data can be analyzed very easily and intuitively.
- Includes the traditional stats from the three most important European competitions:
  - ACB league (from the 1985-1986 season to the 2019-2020 including *Copa del Rey* and *Supercopa*).
  - Euroleague (2000-2001 to 2019-2020).
  - EuroCup (2002-2003 to 2019-2020).
- A total of 40 available stats and several visualizations.

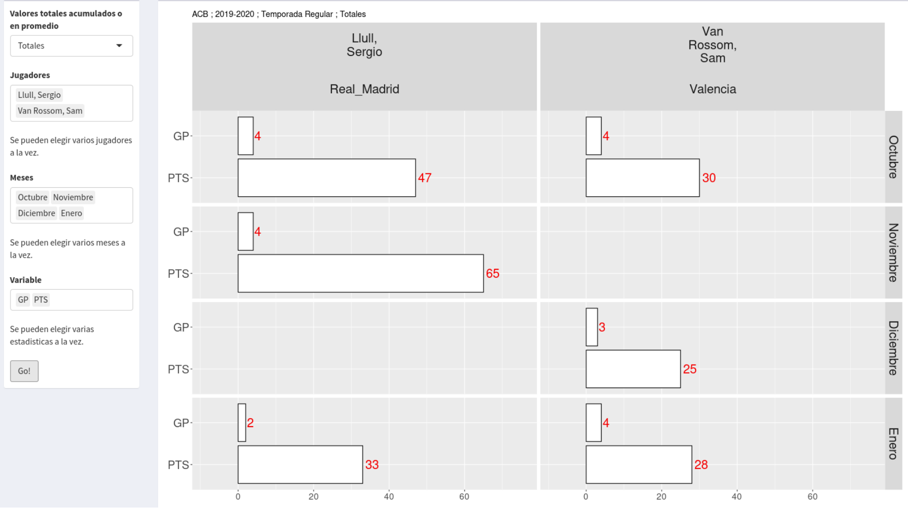
- Correlation between stats. Comparison between players and teams.



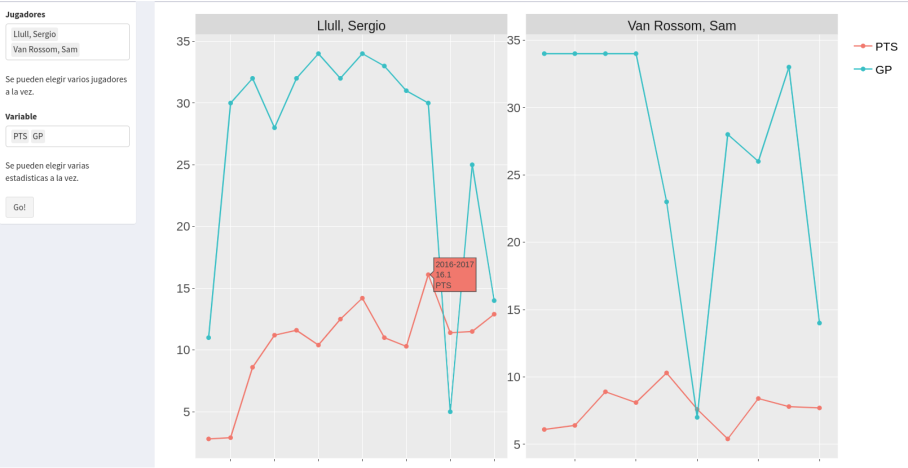
- Percentiles of stats (where the player is ranged regarding all the other players). Comparison between players.



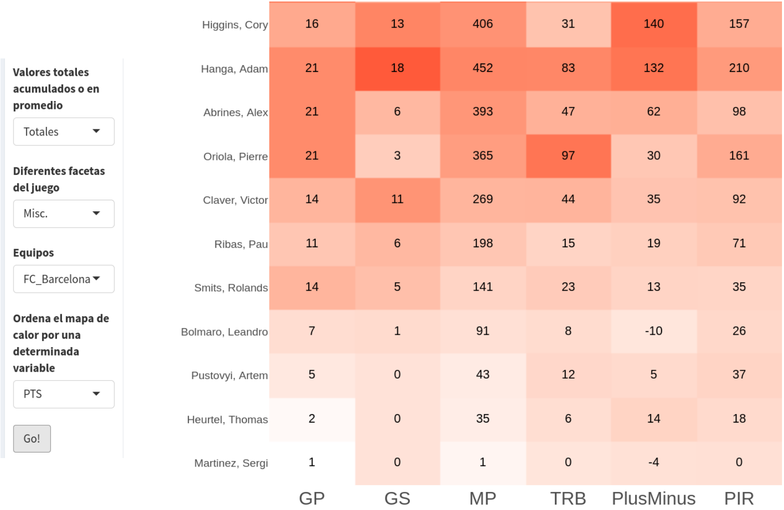
- Follow-up of the players' monthly evolution all season long. Comparison between players.



- Follow-up of the players' yearly evolution over the seasons. Comparison between players.

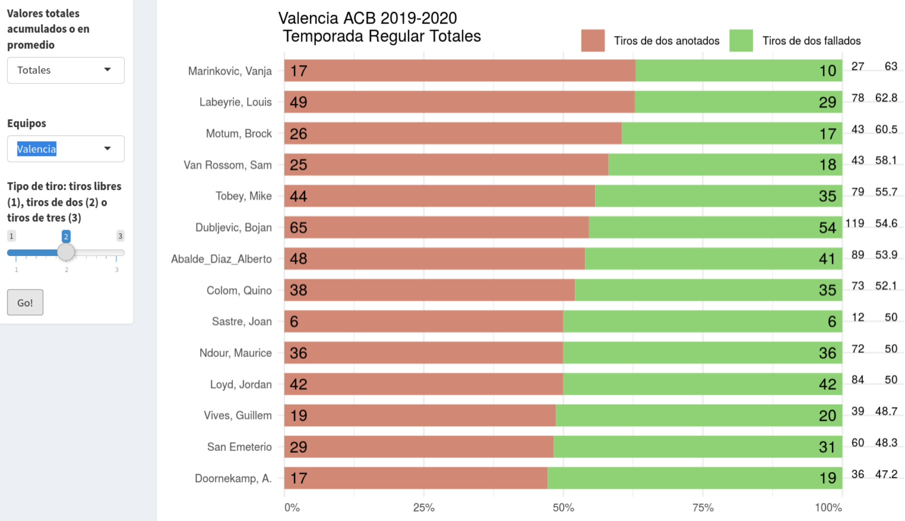


- Heatmap to compare players' stats from the same team.

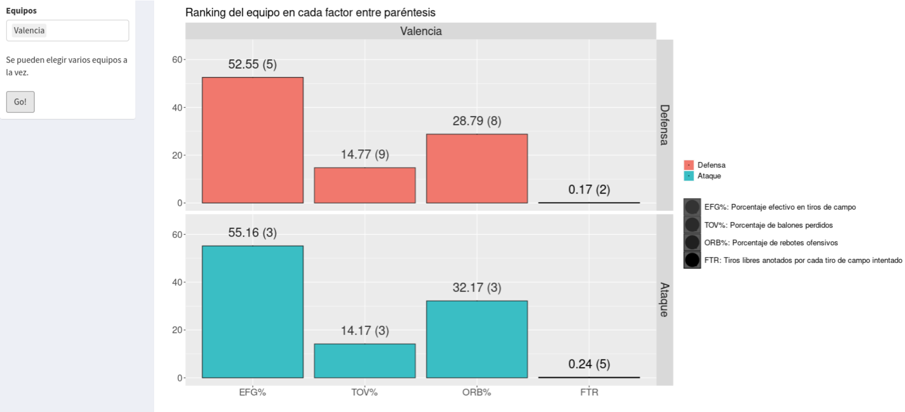


	GP	GS	MP	TRB	PlusMinus	PIR
Higgins, Cory	16	13	406	31	140	157
Hanga, Adam	21	18	452	83	132	210
Abrines, Alex	21	6	393	47	62	98
Oriola, Pierre	21	3	365	97	30	161
Claver, Victor	14	11	269	44	35	92
Ribas, Pau	11	6	198	15	19	71
Smits, Rolands	14	5	141	23	13	35
Bolmaro, Leandro	7	1	91	8	-10	26
Pustovyi, Artem	5	0	43	12	5	37
Heurtel, Thomas	2	0	35	6	14	18
Martinez, Sergi	1	0	1	0	-4	0

- Shooting percentage of players from the same team.



- Four-factors. Comparison between teams.



Publication

Big Data
Volume 8, Number 1, 2020

DOI: 10.1089/big.2018.0124

A Web Application for Interactive Visualization of European Basketball Data

Guillermo Vinué*

Abstract

The statistical analysis of basketball games is a fast-growing field. Certainly, basketball data are scientifically relevant because an appropriate analysis provides a great deal of information about the performance of both players and teams. The number of games played each season generates a large amount of data worth analyzing. Basketball analytics is well established in U.S. leagues. In Europe, however, it has not been duly developed. This study focuses on the top three European team competitions: the EuroLeague, the EuroCup, and the Spanish ACB (Association of Basketball Clubs, acronym in Spanish) league. Their official websites provide access to game data for anyone who is interested, but they are only represented in a static tabular form. As a consequence, it is difficult to gain any valuable insights from them. This article presents a highly useful interactive tool, created with the free statistical software R, which makes it possible to visualize and explore basketball data from a large number of seasons. We will demonstrate its core functionality. An accompanying R package is presented in the Supplementary Data.

Keywords: European basketball; Big data mining; web application; interactive visualization; R

- Impact factor: 2.128 ; Position: 37/110 = 0.34 (Computer Science, Theory & Methods) ; Base: JCR (2020) ; Cites: 5 ; https://doi.org/10.1089/big.2018.0124

Publication

`https://CRAN.R-project.org/package=BAwiR`

BAwiR: Analysis of Basketball Data

Collection of tools to work with basketball data. Functions available are related to friendly web scraping and visualization. Data were obtained from <https://www.euroleague.net/>, <https://www.eurocupbasketball.com/> and <https://www.acb.com/>, following the instructions of their respectives robots.txt files, when available. Tools for visualization include a population pyramid, 2D plots, circular plots of players' percentiles, plots of players' monthly/yearly stats, team heatmaps, team shooting plots, team four factors plots, cross-tables with the results of regular season games and maps of nationalities. Please see Vinue (2020) <doi:10.1089/big.2018.0124>.

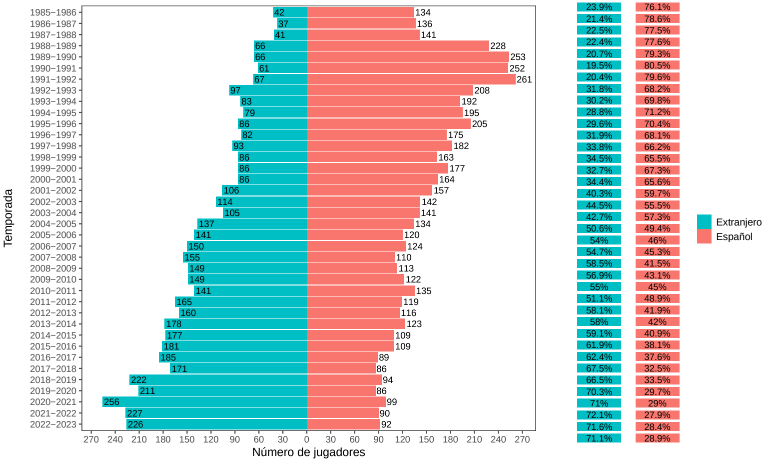
Version:	1.2.7
Depends:	R (≥ 3.4.0)
Imports:	Anthropometry, plyr, dplyr, ggplot2, ggthemes, grid, httr, lubridate, magrittr, purrr, reshape2, rvest, rworldmap, scales, stringi, stringr, tibble, tidyr, xml2
Suggests:	knitr, markdown, rmarkdown
Published:	2021-07-19
Author:	Guillermo Vinue
Maintainer:	Guillermo Vinue <Guillermo.Vinue at uv.es>
License:	GPL-2 | GPL-3 [expanded from: GPL (≥ 2)]
URL:	https://www.R-project.org, https://www.uv.es/vivigui/, https://www.uv.es/vivigui/AppEuroACB.html
NeedsCompilation:	no
Materials:	NEWS
In views:	SportsAnalytics
CRAN checks:	BAwiR results

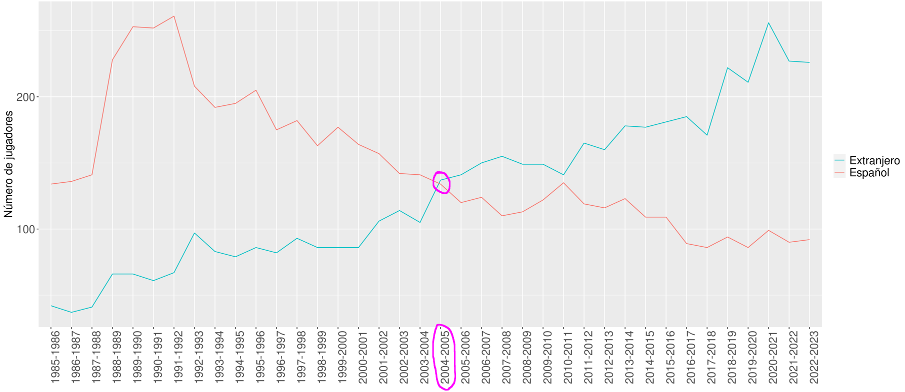
Documentation:

Reference manual:	BAwiR.pdf
Vignettes:	Visualization of European basketball data

Other studies: Demographic analysis

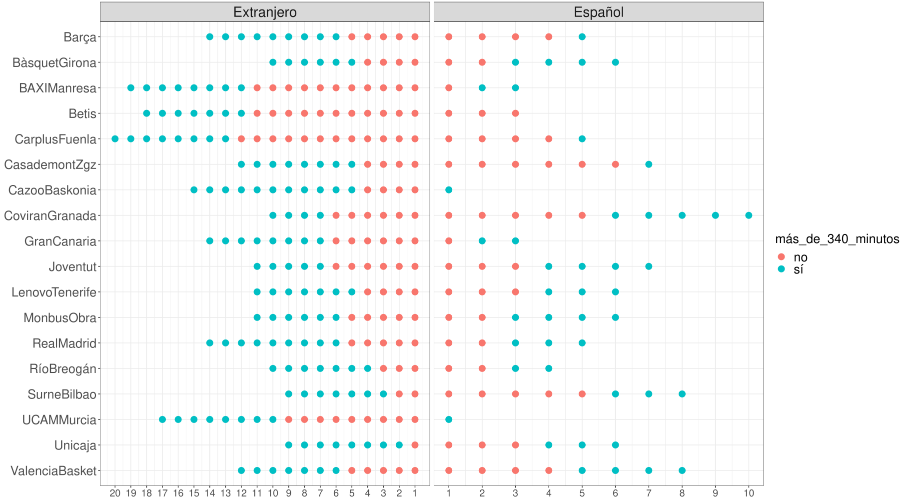
- There are fewer and fewer Spanish players in the ACB (less than 30%).





- In the 80s and 90s there were more Spaniards than foreigners.
- Trend change in the 2004-2005 season.
- From then on, the gap only increases.

- In addition, a large part of these Spaniards is little protagonist.



Publication

GIGANTES

Producto nacional a la baja

Esta temporada el número de jugadores nacidos en España que han jugado en la Liga Endesa ha disminuido. Una tendencia que analizamos desde un punto de vista estadístico una vez finalizado el curso 2021/2022

Guillermo Vinué (analista de datos)

- Gigantes del Basket, number 1522, 2022.

1. Traditional stats.

2. **Play-by-play data.**

New analyses from the play-by-play data

1. Lineups.
2. Breakdown by periods and *clutch time*.
3. Detailed personal fouls.
4. Offensive rebounds turned to scoring.
5. Possessions.
6. Offensive efficacy of time-outs.

Clarification

- The following results have been computed for the four periods of each game from the regular season.

- In other words, **excepting for the computation of clutch time**, overtimes have not been taken into account. Thus, each game has the same duration and results are exactly comparable.

- Out of the 306 regular season games, there were 17 with overtime: 15 with one and 2 with two (5.6%).

1. **Lineups.**
2. Breakdown by periods and *clutch time*.
3. Detailed personal fouls.
4. Offensive rebounds turned to scoring.
5. Possessions.
6. Offensive efficacy of time-outs.

Lineups

- Basketball is a collective game played 5 against 5.
- The plus / minus stat indicates the points difference between one team and the other.

cuarto	minutos	equipo	quinteto en pista	marcador	más_menos
1	10:00 – 5:13	Valencia Basket	Dubljevic, Ferrando, Puerto, Radebaugh, Webb	7-8	-1
1	10:00 – 5:13	Casademont Zaragoza	Cruz, Hlinason, Jovic, Sant-Roos, Yusta	8-7	1

- Nowadays, the plus / minus is only provided for players individually.

MÁS/MENOS			Ver lista
1	NIKOLA MIROTIC	BARÇA	9,9
2	Aaron Doornekamp	Lenovo Tenerife	8,9
3	Sertac Sanli	Barça	8,6
4	Dzanan Musa	Real Madrid	8,4
5	Tomas Satoransky	Barça	8,3

- However, it is much more informative to know how players interact.
- From the 5 player lineups, we can obtain the 4 player lineups, 3 player lineups and 2 player lineups that are part of them.
- Combinatorial theory: $C_{5,x} = \binom{5}{x} = \frac{5!}{x!(5-x)!}$, for $x = 4, 3, 2$
- Additional information for a coach: Find out what the best and worst combinations are during the season.

Results

- The best and worst three 5 PLAYER LINEUPS of the league in terms of **plus** / **minus** are the following ones:

Best 5				
team	lineup	occurrences	time	+ / -
Valencia	Dubljevic, Jones, Puerto, Radebaugh, Webb	15	58M 23S	52
Gran Canaria	Albicy, Balcerowski, Brussino, Shurna, Slaughter	52	2H 29M 44S	51
Murcia	Bellas, Diop, McFadden, Radovic, Rojas	17	54M 33S	45

- **Note:** From the 15 times that Dubljevic, Jones, Puerto, Radebaugh and Webb played together, only one was in the last period.

Worst 5				
team	lineup	occurrences	time	+ / -
Bilbao	Alonso, Kyser, Radicevic, Reyes, Sulejmanovic	11	26M 27S	-47
Granada	Bropleh, Caicedo, Costa, Moore, Ndoye	8	33M 2S	-45
Manresa	Harding, Pérez, Sagnia, Steinbergs, Tyson	6	9M 10S	-28

- **Note:** If the number of occurrences or the time played seems to be scarce to achieve a relevant conclusion, it is recommended to filter by a suitable threshold.

- The best and worst three 4 PLAYER LINEUPS of the league in terms of **plus** / **minus** are the following ones:

Best 4				
team	lineup	occurrences	time	+ / -
Gran Canaria	Albicy, Brussino, Shurna, Slaughter	111	4H 38M 52S	88
Barcelona	Laprovittola, Mirotic, Sanli, Satoransky	37	1H 20M 24S	79
Valencia	Dubljevic, Jones, Puerto, Radebaugh	49	2H 46M 53S	79

Worst 4				
team	lineup	occurrences	time	+ / -
Bilbao	Alonso, Kyser, Reyes, Sulejmanovic	34	1H 30M 46S	-69
Granada	Bropleh, Costa, Moore, Ndoye	16	49M 15S	-57
Bilbao	Hakanson, Kyser, Smith, Sulejmanovic	51	1H 21M 49S	-52

- The best and worst three 3 PLAYER LINEUPS of the league in terms of **plus** / **minus** are the following ones:

Best 3				
team	lineup	occurrences	time	+ / -
Barcelona	Laprovittola, Mirotic, Satoransky	112	4H 40M 32S	125
Barcelona	Laprovittola, Sanli, Satoransky	82	2H 53M 36S	119
Tenerife	Cook, Doornekamp, Salin	106	3H 4M 7S	114

Worst 3				
team	lineup	occurrences	time	+ / -
Bilbao	Alonso, Kyser, Sulejmanovic	49	1H 48M 15S	-83
Bilbao	Alonso, Kyser, Reyes	90	3H 25M 11S	-80
Bilbao	Kyser, Reyes, Sulejmanovic	88	2H 48M 20S	-80

- The best and worst three 2 PLAYER LINEUPS of the league in terms of **plus** / **minus** are the following ones:

Best 2				
team	lineup	occurrences	time	+ / -
Barcelona	Laprovittola, Satoransky	214	8H 8M 1S	186
Barcelona	Laprovittola, Sanli	156	4H 54M 16S	181
Tenerife	Doornekamp, Salin	249	6H 17M 54S	178

Worst 2				
team	lineup	occurrences	time	+ / -
Bilbao	Kyser, Sulejmanovic	189	5H 21M 45S	-139
Fuenlabrada	Horton, Kromah	103	4H 13M 44S	-135
Bilbao	Kyser, Reyes	233	6H 57M 1S	-129

- Extension with other stats. Same analysis as in the NBA.

https://www.nba.com/stats/lineups/advanced

SEASON: 2022-23
SEASON TYPE: Regular Season
TEAM: All Teams
LINEUPS: 5 Player Lineups

- 5 Player Lineups
- 4 Player Lineups
- 3 Player Lineups
- 2 Player Lineups

2000 Rows • Page

LINEUPS	TEAM	GP	MIN	OFFRTG	DEFRTG	NETRTG	AST%	AST/TO	AST RATIO	OREB%	DREB%	REB%	TO RATIO
H. Barnes - D. Sabonis - D. Fox - K. Huerter - K. Murray	SAC	63	900	118.3	116.1	2.2	62.0	2.30	19.4	25.8	76.4	50.6	0.1
C. Capela - D. Murray - J. Collins - T. Young - D. Hunter	ATL	44	738	115.9	110.6	5.3	56.9	2.25	18.4	28.8	74.2	51.1	0.1
K. Caldwell-Pope - A. Gordon - N. Jokic - J. Murray - M. Porter Jr.	DEN	41	706	124.3	111.2	13.1	69.8	2.66	23.2	28.3	75.1	52.5	0.1

1. Lineups.
2. **Breakdown by periods and clutch time.**
3. Detailed personal fouls.
4. Offensive rebounds turned to scoring.
5. Possessions.
6. Offensive efficacy of time-outs.

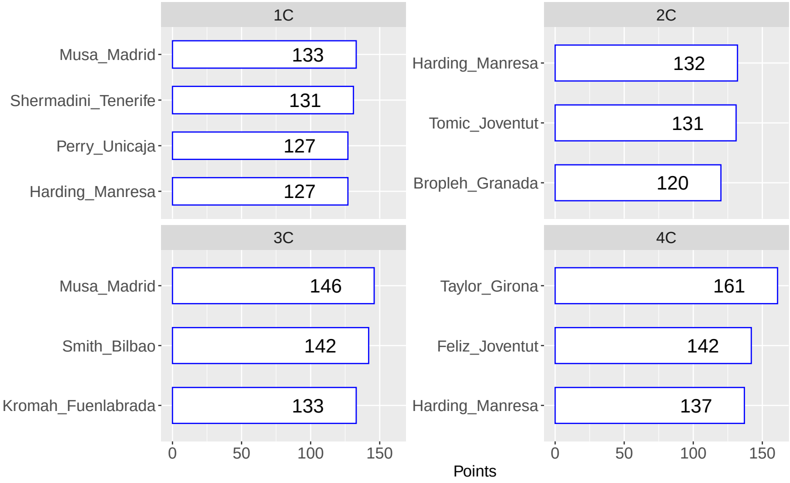
Breakdown by periods

- A basketball game in the ACB is divided in four periods of 10 minutes. If the match ends in a tie, there are successive 5-minute overtimes until the tie is broken at the end of the corresponding overtime.
- In addition the overall players' performance, it is also interesting to know how the player performs by period.

- For instance, to know...
- who the most activated player at the beginning of the first or third quarter is or...
- who the best player at *clutch time* is.

Results

- Top 3 of scorers in each period:



- **Clutch time**: Game situation when the scoring margin is within 5 points with five or fewer minutes remaining in a game.

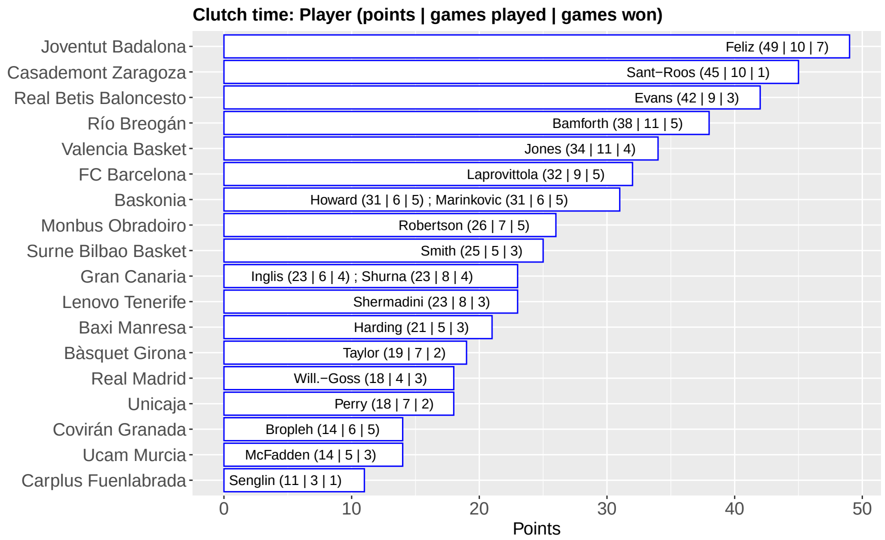
Results

- There have been 76 *clutch time* games (approximately 25%, that is to say, one out of four).

- Top 3 *clutch time* scorers:

player	team	points	games played	games won	percentage victories
Feliz	Joventut	49	10	7	70%
Sant-Roos	Zaragoza	45	10	1	10%
Evans	Betis	42	9	3	33.33%

- Top *clutch time* scorer for each team:



- Extension with other stats. Same analysis as in the NBA.

`https://www.nba.com/stats/players/clutch-traditional`

SEASON: 2022-23 | SEASON TYPE: Regular Season | PER MODE: Totals | SEASON SEGMENT: All Season Segments

Advanced Filters

Per Mode: Totals ×

GLOSSARY SHARE

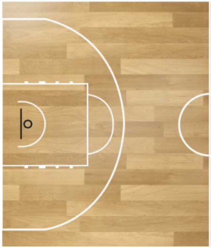
451 Rows • Page 1 of 10

	PLAYER	TEAM	AGE	GP	W	L	MIN	PTS	FGM	FGA	FG%	3PM	3PA	3P%	FTM	FTA	FT%	OREB	DREB	REB	AST	TOV	STL	BLK	PF
1	Reggie Bullock	DAL	32	48	21	27	162.8	39	11	35	31.4	9	30	30.0	8	13	61.5	3	16	19	6	3	3	3	21
2	Bam Adebayo	MIA	25	47	28	19	180.2	90	29	50	58.0	0	2	0.0	32	35	91.4	14	30	44	13	11	12	5	25
3	Anthony Edwards	MIN	21	45	26	19	162.7	121	44	104	42.3	13	39	33.3	20	28	71.4	5	23	28	11	16	6	10	16
3	Buddy Hield	IND	30	45	23	22	148.6	69	23	53	43.4	13	33	39.4	10	11	90.9	7	20	27	4	3	8	4	13

1. Lineups.
2. Breakdown by periods and *clutch time*.
3. **Detailed personal fouls.**
4. Offensive rebounds turned to scoring.
5. Possessions.
6. Offensive efficacy of time-outs.

Detailed personal fouls

- In basketball, a personal foul consists of a physical contact with the opponent in an illegal (non-regulatory) manner.
- If at the moment the foul is made, the attacking player is shooting, that foul is penalized with free throws:
  - 1 free throw if the shot ends in scoring.
  - 2 free throws if not scoring and the foul is within the line of three.
  - 3 free throws if not scoring and the foul is outside the line of three.



- Even if the fouled player is not in the act of shooting, two free throws are also awarded if the opposing team is in *bonus* (5 fouls have been already made in the same period).

- The traditional stat counts the number of fouls committed and received by each player.

- However, we can go a little deeper.

- Idea of analysis: find out how many **free throw fouls** commits and receives every player.

- Added value for a player: He does not only gets fouls, but he manages to get free throw fouls.

- Find out a negative aspect of a player: He does not only commits fouls, but these are free throw fouls.

Clarification

- For some players there may be a slight deviation between the number of total fouls calculated in this analysis and the one on the ACB website, due to some inconsistencies in the description of the data, that I have not been able to solve for all cases.

589	2C	00:54	Abromaitis	Falta Personal (1TL)	39	33	2	103361	Lenovo Tenerife
590	2C	00:54	Bursac	Falta Recibida	39	33	2	103361	Bàsquet Girona
591	2C	00:46	Abromaitis	Falta Personal (2TL)	39	33	2	103361	Lenovo Tenerife
592	2C	00:46	Vila	Falta Recibida	39	33	2	103361	Bàsquet Girona
593	2C	00:46	Vila	Tiro Libre fallado	39	33	2	103361	Bàsquet Girona
594	2C	00:46	Vila	Tiro Libre anotado	39	34	2	103361	Bàsquet Girona

period	time_point	player	action	local_score	visitor_score	day	game_code	team
2C	07:50	Poirier	Falta Personal	24	21	1	103350	Real Madrid
2C	07:50	Gasol	Falta Recibida	24	21	1	103350	Bàsquet Girona
2C	07:49	Gasol	Tiro Libre anotado	25	21	1	103350	Bàsquet Girona
2C	07:49	Gasol	Tiro Libre anotado	26	21	1	103350	Bàsquet Girona

Results

- Top 3 of players with **most free throw fouls committed**:

team	player	n$^{\underline{r}}$ of FT fouls committed and percentage	n$^{\underline{r}}$ of FT given	type of FT	amount
Bilbao	Sulejmanovic	55 (58.5%)	97	1TL	15
				2TL	38
				3TL	2
Zaragoza	Hlinason	54 (57.4%)	97	1TL	12
				2TL	41
				3TL	1
Girona	Miletic	52 (66.7%)	88	1TL	16
				2TL	36
Breogán	Happ	52 (52.5%)	89	1TL	16
				2TL	35
				3TL	1

- Player(s) with **most 1, 2 and 3 free throw fouls committed**:

team	player	type of FT	amount
Breogán	Happ	1TL	16
Girona	Miletic	1TL	16
Murcia	Sakho	2TL	45
Baskonia	Costello	3TL	4
Valencia	Radebaugh	3TL	4
Gran Canaria	Inglis	3TL	4
Madrid	Causeur	3TL	4

- Top 3 of players with **most free throw fouls received**:

team	player	nº of FT fouls received and percentage	nº of FT achieved	type of FT	amount
Tenerife	Shermadini	96 (63.2%)	173	1TL	19
				2TL	77
Madrid	Musa	87 (66.4%)	155	1TL	19
				2TL	68
Gran Canaria	Balcerowski	70 (72.9%)	124	1TL	16
				2TL	54

- Player(s) with **most 1, 2 and 3 free throw fouls received**:

team	player	type of FT	amount
Madrid	Tavares	1TL	21
Tenerife	Shermadini	2TL	77
Baskonia	Howard	3TL	9
Bilbao	Andersson	3TL	9

- Further analysis with attack fouls.

- Top 3 of players with **most attack fouls committed**:

team	player	amount
Tenerife	Shermadini	20
Obradoiro	Blazevic	20
Unicaja	Ejim	15

- Top 3 of players with **most attack fouls received**:

team	player	amount
Gran Canaria	Albicy	26
Obradoiro	Westermann	20
Unicaja	Alberto Díaz	18

- This allows us to complete the official stats.

FALTAS RECIBIDAS			Ver lista
1	GIORGI SHERMADINI	LENOVO TENERIFE	4,9
2	Ante Tomic	Joventut Badalona	4,3
3	Dzanan Musa	Real Madrid	4,1
4	Kassius Robertson	Monbus Obradoiro	4,0
5	Jasiel Rivero	Valencia Basket	4,0

FALTAS COMETIDAS			Ver lista
1	MAREK BLAZEVIC	MONBUS OBRADOIRO	3,1
2	Juan Fernández	Carplus Fuenlabrada	3,1
3	Guillem Vives	Joventut Badalona	3,1
4	Ethan Happ	Río Breogán	2,9
5	Tryggvi Hlinason	Casademont Zaragoza	2,9

1. Lineups.
2. Breakdown by periods and *clutch time*.
3. Detailed personal fouls.
4. **Offensive rebounds turned to scoring.**
5. Possessions.
6. Offensive efficacy of time-outs.

Offensive rebounds turned to scoring

- When a team is attacking and misses the shot, can continue attacking if grabbing the rebound $\rightarrow$ offensive rebounds stat.
- Grabbing offensive rebounds is considered as one of the most important aspects of the game.
- The traditional stat counts the number of offensive rebounds that every player has grabbed.

- However, we can go here also a little deeper.

- Idea of analysis: find out the outcome of the offensive rebound, that is to say, **how many offensive rebounds end in scoring**.

- Added value for a player: He not only grabs the offensive rebound, but he also scores it.

Results

- Teams ordered by points scored after offensive rebounds. In red, the three with the best percentage.

team	offensive rebounds	scored offensive rebounds	percentage	scored points
Barcelona	375	209	55.73	466
Manresa	362	212	58.56	461
Betis	405	210	51.85	457
Joventut	366	208	56.83	447
Murcia	408	203	49.75	436
Granada	389	202	51.93	429
Breogán	394	207	52.54	429
Valencia	345	194	56.23	427
Madrid	359	201	55.99	426

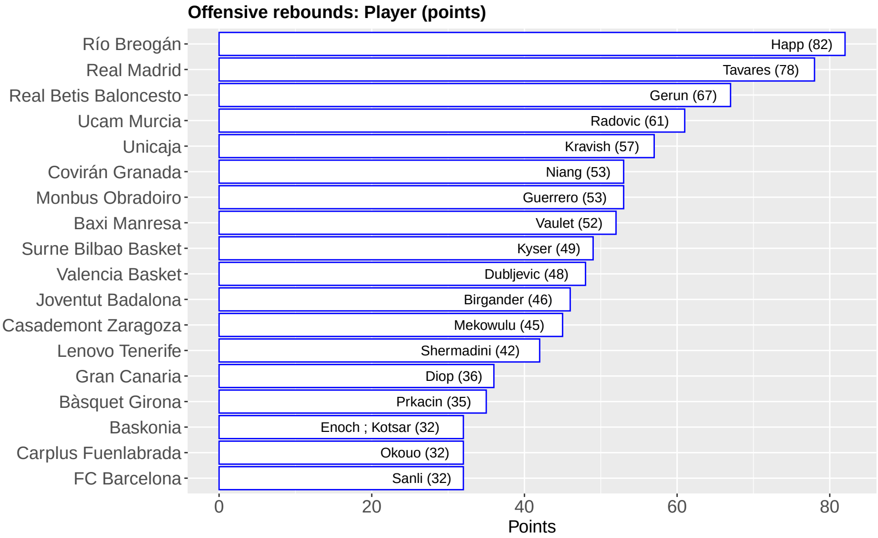
- Continues in the next slide...

team	offensive rebounds	scored offensive rebounds	percentage	scored points
Gran Canaria	327	193	59.02	409
Unicaja	337	189	56.08	409
Obradoiro	336	174	51.79	382
Bilbao	355	182	51.27	375
Baskonia	323	163	50.46	363
Zaragoza	337	179	53.12	361
Girona	341	168	49.27	358
Fuenlabrada	330	167	50.61	346
Tenerife	304	160	52.63	345

- This allows us to complete the official stats.
- **Note:** There is a slight deviation between the total data in this analysis and those on the ACB website because I am not counting team rebounds.

		PAR	MIN	PT	TIROS DE 3			TIROS DE 2			TIROS LIBRES			REBOTES		
		Jug	Jug		Con	Int	%	Con	Int	%	Con	Int	%	Ofe	Def	Tot
1	Real Betis Baloncesto	34	6950:00	2.659	308	965	31,9%	628	1.266	49,6%	479	658	72,8%	413	814	1.227
2	UCAM Murcia	34	6825:00	2.714	318	1.000	31,8%	642	1.340	47,9%	476	622	76,5%	409	868	1.277
3	Río Breogán	34	6875:00	2.620	266	825	32,2%	722	1.402	51,5%	378	572	66,1%	397	790	1.187
4	Coviran Granada	34	6825:00	2.604	262	854	30,7%	678	1.324	51,2%	462	664	69,6%	392	849	1.241
5	Joventut Badalona	36	7250:00	2.948	285	838	34,0%	778	1.453	53,5%	537	689	77,9%	388	987	1.375
6	Barça	35	7000:00	2.979	342	922	37,1%	724	1.269	57,1%	505	655	77,1%	381	897	1.278
7	Real Madrid	35	7000:00	3.100	345	919	37,5%	744	1.321	56,3%	577	751	76,8%	365	968	1.333
8	BAXI Manresa	34	6850:00	2.853	287	861	33,3%	767	1.448	53,0%	458	618	74,1%	364	825	1.189
9	Valencia Basket	35	7050:00	2.916	351	1.011	34,7%	673	1.224	55,0%	517	665	77,7%	356	813	1.169
10	Surne Bilbao Basket	34	6800:00	2.587	288	869	33,1%	628	1.273	49,3%	467	630	74,1%	355	830	1.185
11	Unicaja	35	7050:00	3.041	355	990	35,9%	769	1.342	57,3%	438	616	71,1%	348	882	1.230
12	Casademont Zaragoza	34	6950:00	2.675	241	744	32,4%	737	1.401	52,6%	478	697	68,6%	348	928	1.276
13	Bàsquet Girona	34	6875:00	2.604	243	817	29,7%	718	1.387	51,8%	439	615	71,4%	345	891	1.236
14	Cazoo Baskonia	36	7275:00	3.295	444	1.098	40,4%	718	1.242	57,8%	527	698	75,5%	342	928	1.270
15	Dreamland Gran Canaria	35	7050:00	2.905	360	973	37,0%	654	1.203	54,4%	517	718	72,0%	340	838	1.178
16	Monbus Obradoiro	34	6850:00	2.694	345	953	36,2%	574	1.143	50,2%	511	670	76,3%	337	838	1.175
17	Carplus Fuenlabrada	34	6800:00	2.589	228	715	31,9%	711	1.388	51,2%	483	675	71,6%	330	781	1.111
18	Lenovo Tenerife	35	7075:00	2.893	325	908	35,8%	686	1.189	57,7%	546	695	78,6%	315	864	1.179

- Top team scorer after grabbing an offensive rebound:



- This allows us to complete the official stats.

		PAR	MIN	PT	TIROS DE 3			TIROS DE 2			TIROS LIBRES			REBOTES		
		Jug	Jug		Con	Int	%	Con	Int	%	Con	Int	%	Ofe	Def	Tot
1	Ethan Happ	34	828	427	0	0	0,0%	188	337	55,8%	51	93	54,8%	106	162	268
2	Volodymyr Gerun	29	695	242	0	2	0,0%	89	153	58,2%	64	109	58,7%	95	96	191
3	Nemanja Radovic	33	746	344	9	25	36,0%	138	258	53,5%	41	54	75,9%	74	116	190
4	Bojan Dubljevic	30	673	312	42	117	35,9%	60	111	54,1%	66	76	86,8%	72	129	201
5	Tyson Pérez	28	698	354	25	82	30,5%	101	170	59,4%	77	100	77,0%	69	146	215
6	Edy Tavares	29	591	303	1	1	100,0%	122	176	69,3%	56	77	72,7%	67	109	176
7	Emir Sulejmanovic	34	707	248	24	71	33,8%	75	139	54,0%	26	49	53,1%	66	140	206
8	Vincent Poirier	26	472	219	3	4	75,0%	78	149	52,3%	54	73	74,0%	66	59	125
9	Tryggvi Hlinason	33	642	243	0	1	0,0%	97	124	78,2%	49	86	57,0%	65	100	165
10	Khalifa Diop	34	684	263	0	2	0,0%	110	187	58,8%	43	70	61,4%	63	107	170
11	David Kravish	34	655	285	10	26	38,5%	117	202	57,9%	21	35	60,0%	63	124	187
12	Simon Birgander	33	560	220	0	1	0,0%	89	165	53,9%	42	50	84,0%	62	102	164
13	Ante Tomic	34	786	443	1	3	33,3%	174	323	53,9%	92	128	71,9%	61	149	210
14	Thomas Scrubb	33	974	383	41	108	38,0%	107	195	54,9%	46	61	75,4%	61	121	182
15	Michale Kyser	34	674	260	4	18	22,2%	99	180	55,0%	50	69	72,5%	60	65	125

1. Lineups.
2. Breakdown by periods and *clutch time*.
3. Detailed personal fouls.
4. Offensive rebounds turned to scoring.
5. **Possessions.**
6. Offensive efficacy of time-outs.

- In basketball, the concept of possession is defined as the time that the team has the ball and is attacking.

- The possession ends when the defense gets the ball (and then, starts its possession).

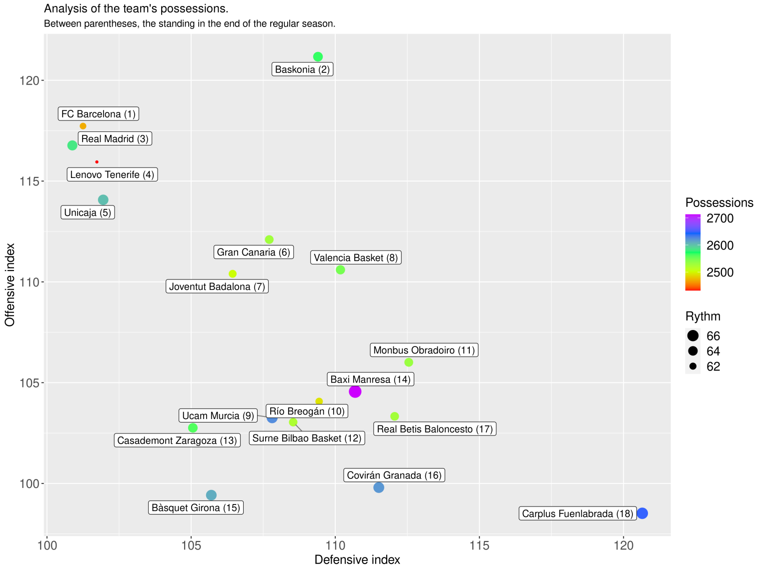
- Importance: The variables are normalized according a common criterion.

- It allows us to compare leagues, games, teams or players.

Team variables related to possessions

- Offensive rating: points in favor / nº of possessions attacking.
- Defensive rating: points against / nº of possessions defending.
- Pace: nº of possessions in 40 minutes.

Results


Analysis of the team's possessions.
Between parentheses, the standing in the end of the regular season.
Baskonia (2)
FC Barcelona (1)
Real Madrid (3)
Lenovo Tenerife (4)
Unicaja (5)
Gran Canaria (6)
Valencia Basket (8)
Joventut Badalona (7)
Monbus Obradoiro (11)
Baxi Manresa (14)
Ucam Murcia (9)
Río Breogán (10)
Real Betis Baloncesto (17)
Casademont Zaragoza (13)
Surne Bilbao Basket (12)
Covirán Granada (16)
Bàsquet Girona (15)
Carplus Fuenlabrada (18)
Offensive index
Defensive index
120
115
110
105
100
100
105
110
115
120
Possessions
2700
2600
2500
Rythm
66
64
62

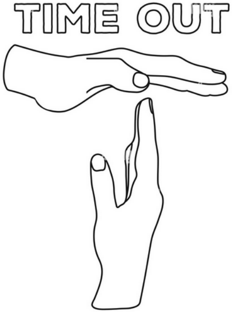
1. Lineups.
2. Breakdown by periods and *clutch time*.
3. Detailed personal fouls.
4. Offensive rebounds turned to scoring.
5. Possessions.
6. **Offensive efficacy of time-outs.**

Offensive efficacy of time-outs

- The work of coaches is more difficult to analyze coldly.
- Usual conclusion: if the team wins, the players are very good. If the team loses, the coach is very bad.

- To add objectivity to the opinion about the coaches we can try to study how they intervene during the matches.

- A direct intervention is through time-outs.



- Coaches use time-outs to guide their players in both attack and defense.
- When a coach calls a time-out, it is because he has to attack in the next play.
- Idea of analysis: to find out if the attacking play proposed by the coach ends in scoring (at least one point).

Results

- Coaches ordered by most points scored after a time-out. In red, the three with the best percentage.

team	coach	games	requested time-outs	successful time-outs	percentage	scored points
Murcia	S. Alonso	34	141	60	42.55	133
Valencia	Á. Mumbrú	33	108	55	50.93	128
Baskonia	J. Peñarroya	34	111	56	50.45	124
Granada	P. Pin	34	137	56	40.88	122
Obradoiro	M. Fernández	34	126	56	44.44	119
Breogán	V. Mirsic	34	139	52	37.41	107
Barcelona	S. Jasikevicius	34	107	44	41.12	101
Joventut	C. Duran	34	110	46	41.82	97
Betis	L. Casimiro	34	108	44	40.74	95
Zaragoza	P. Fisac	30	104	43	41.35	91
Bilbao	J. Ponsarnau	34	104	35	33.65	82

- Continues in the next slide...

team	coach	games	requested time-outs	successful time-outs	percentage	scored points
Gran Canaria	J. Lakovic	30	91	33	36.26	77
Manresa	P. Martínez	33	97	34	35.05	74
Girona	A.G. Reneses	34	119	35	29.41	74
Madrid	C. Mateo	34	80	31	38.75	71
Tenerife	T. Vidorreta	31	71	30	42.25	65
Unicaja	I. Navarro	33	87	32	36.78	64
Fuenlabrada	J.L Pichel	14	57	21	36.84	45
Fuenlabrada	Ó. Quintana	17	76	20	26.32	45
Fuenlabrada	J.M. Raventós	3	13	4	30.77	9
Zaragoza	M. Schiller	4	17	5	29.41	9
Gran Canaria	V. García	4	10	3	30	9

- **Curiosity**: The coach of Tenerife, Txus Vidorreta, did not requested a time-out for two consecutive games (days 8 and 9).

Coaches who did not requested a time-out				
day	team	coach	game	result
6	Unicaja	I. Navarro	Unicaja-Betis	106-60
8	Tenerife	T. Vidorreta	Tenerife-Valencia	94-78
9	Tenerife	T. Vidorreta	Breogán-Tenerife	61-86
17	Tenerife	T. Vidorreta	Tenerife-Betis	88-64
33	Valencia	Á. Mumbrú	Fuenlabrada-Valencia	75-93
34	Manresa	P. Martínez	Manresa-Gran Canaria	105-75

Predictive statistics

Forecasting the future players' performance

- The goal of the study is to provide predictions of the future players' performance, based on their previous activity.

- The statistical approach is based on combining the archetypoid analysis with sparse functional data.

- In professional sports, it is very common that data are sparse and irregular.
    - They are sparse because most players do not play a lot of time in the same league.
    - They are irregular because each player plays a different number of seasons.

Example of available data (time data):

- Rows: Players.
- Columns: Age (years).
- Cells: Value of the stat of interest.

	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Claver, V.	NA	NA	NA	1.9	6.5	10.6	9.7	13	7.3	NA	NA	NA	NA	6.8	6.8	NA	NA	NA	NA	NA	NA
Gasol, P.	NA	NA	NA	4.0	5.7	14.5	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Perasovic, V.	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	24.2	18.1	17.9	17.4	13.7	NA	20.6	19	21.2
Rubio, R.	2.6	8.3	13.9	14.7	11.5	9.0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Sabonis, A.	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	26.7	24.7	31.1	25.3	28.6	34.2	NA	NA	NA	NA	NA

Methodology (I): Archetypoids

- The analysis of archetypes allows us to identify in a data set those activity patterns that stand out notably from the rest.
- These patterns are called archetypes, since they are models that serve to exemplify a certain behavior.

- The analysis of archetypoids is an extension of the archetype analysis so that the archetypes correspond with specific observations of the data set $\rightarrow$ archetypoids.

- In basketball (and any other sport), these archetypoids are players who have a very specific performance (for better or for worse).

- Each individual of the database is represented as a combination of archetypoids.

Mathematical definition

- Let $\mathbf{X_{n \times m}}$ be a multivariate data matrix (n individuals, m variables).
- The archetypoid analysis aims to identify k archetypoids by computing the matrices α and β ($n \times k$) that minimize the following residuals squared sum (RSS):

$$RSS = \sum_{i=1}^{n} \|\mathbf{x}_i - \sum_{j=1}^{k} \alpha_{ij}\mathbf{z}_j\|^2 = \sum_{i=1}^{n} \|\mathbf{x}_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl}\mathbf{x}_l\|^2$$

under the constraints:

1) $\sum_{l=1}^{n} \beta_{jl} = 1$ with $\sum_{l=1}^{n} \beta_{jl} = 1$ and $\boxed{\beta_{jl} \in \{0, 1\}}$ $\Longrightarrow$ $\mathbf{z}_j = \sum_{l=1}^{n} \beta_{jl}\mathbf{x}_l$

2) $\sum_{j=1}^{k} \alpha_{ij} = 1$ with $\alpha_{ij} \geq 0$ $\Longrightarrow$ $\hat{\mathbf{x}}_i = \sum_{j=1}^{k} \alpha_{ij}\mathbf{z}_j$

- Ideally, the archetypoids are located in the frontiers of the data cloud.

Philosophical approximation to the archetypoids, or...

THE STORY OF TWO COACHES NAMED
PLATO AND ARISTOTLE

- Plato (427-347 a. C.) already proposed an abstract idea of archetype when he defined it as an original model of any expression of reality.
- What's beauty?

What's beauty?

Is it an ideal of perfection and symmetry?

Is it a paradigm of brilliance and splendor?

- Aristotle (384-322 a. C.) was more pragmatic.
- What's beauty? Stop messing...

Beauty is...

AN ARCHETYPE FOR PLATO:

AN ARCHETYPOID FOR ARISTOTLE:

Method of analogues

- In order to refine predictions, we use the so-called *method of analogues*.

- The idea is to find players related to the one of interest and then use their documented activity to obtain the predictions.

- We know how other players already performed, so we can use their information to gain an approximate idea about the future performances of others.

- In order to find related players, we use the archetypoid analysis.

Methodology (II): Functional data

- A time series is a sequence of data ordered in time.
- The functional data analysis allows us to analyze time series data.
- The idea of this work is to consider the sequence of data for every player as a continuous function, instead of as individual observations.

- Each datum is a function, $x_i(t)$, that represents the metric value of player i for a certain age t.

- The archetypoid analysis was adapted to the functional field by Irene Epifanio.

- In order to obtain the forecasts, we use the *Regularized Optimization for Prediction and Estimation with Sparse data* (ROPES) method, developed by Alexander Dokumentov and Rob Hyndman.

- It makes predictions with bidimensional spare functional data.

- It provides prediction intervals, which are very important because they give a measure of uncertainty associated with the predictions.

Results: Luka Doncic

- Foreign players have an increasing impact in the NBA (Nikola Jokic, Giannis Antetokounmpo, Luka Doncic...).
- The identification of future international stars has gained great importance.

- Luka Doncic played in the ACB for Real Madrid between the 2014-2015 and 2017-2018 seasons (between the ages of 15 and 18).

- In the 2018 NBA draft, he was selected number 3, which indicates the high expectations that were set for him (and that are being met).

- We proposed this study in 2018.

- Our proposal was to use this methodology to forecast Doncic's evolution in future seasons, in terms of his ACB rating.

- This forecast could be used as an approximation of his possible performance in the NBA.

Stat of interest

- In the ACB and other European leagues, the *Performance Index Rating (PIR)* is used to evaluate the players' performance.

- It is a unique value for each player that results from adding their positive actions and subtracting the negative ones:

  - Sum: points + rebounds + assists + steals + blocks in favor + received fouls.

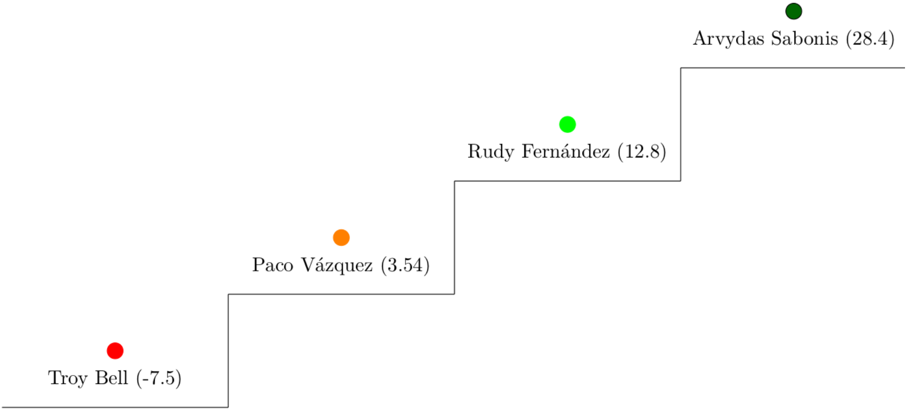
  - Subtraction: missed field goals - missed free throws - turnovers - blocks against - committed fouls.

Procedure

- **Database**: Average ratings of all ACB players from the 1985-1986 season to the 2017-2018.

1. Express the rating data for all players as functions in an orthonormal basis.
2. Apply archetypoid analysis to the coefficients of that base.

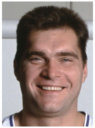
- 4 players (archetypoids) representative of different types of performance were obtained (rating in parentheses):



- Troy Bell: He represents those players with an insignificant performance or with a low presence in the league.
- Paco Vázquez: He represents those mid-level players with a regular presence on their teams.
- Rudy Fernández: He represents those very high level players.
- Arvydas Sabonis: He represents those players considered as super stars, the highest level.

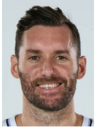
3. Obtain the players analogous to Doncic using the analogous method.

	Archetypoids and similarity percentage			
	P. Vázquez	A. Sabonis	T. Bell	R. Fernández
L. Doncic	0	54%	0	46%



54%

+



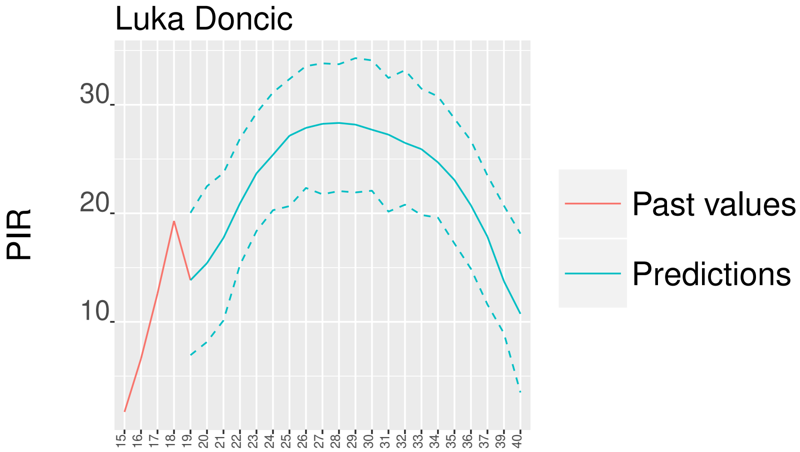
46%

=

- The players that Doncic can be similar to are those who have a percentage of similarity with Sabonis greater than that of Doncic.

4. Apply ROPES to the set of those analogous players (including Doncic).

- Forecast made in 2018 about Doncic's future performance:



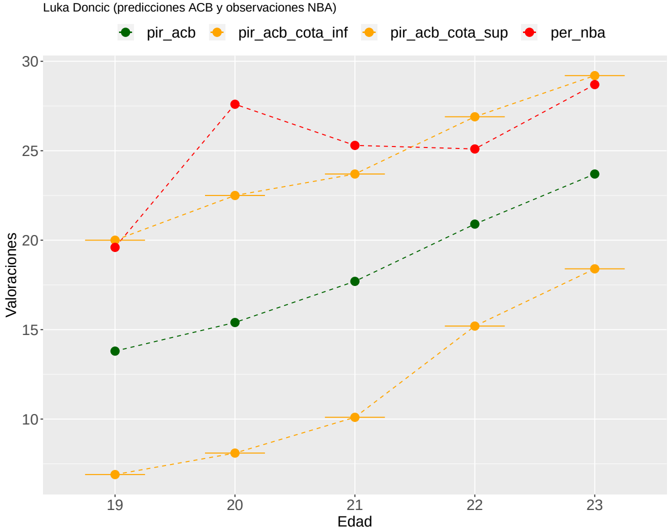
- Over the years, we can compare the predictions with the observations. Doncic has been playing in the NBA since he was 19 years old.

- The NBA stat that we can consider closest to the ACB's PIR is the *Player Efficiency Rating (PER)*[4].

[4] https://en.wikipedia.org/wiki/Player_efficiency_rating

- PER, like PIR, tries to reduce all player's contributions to a single number.

- But they don't use the same formula. Therefore, this comparison can only be done in an approximate way.



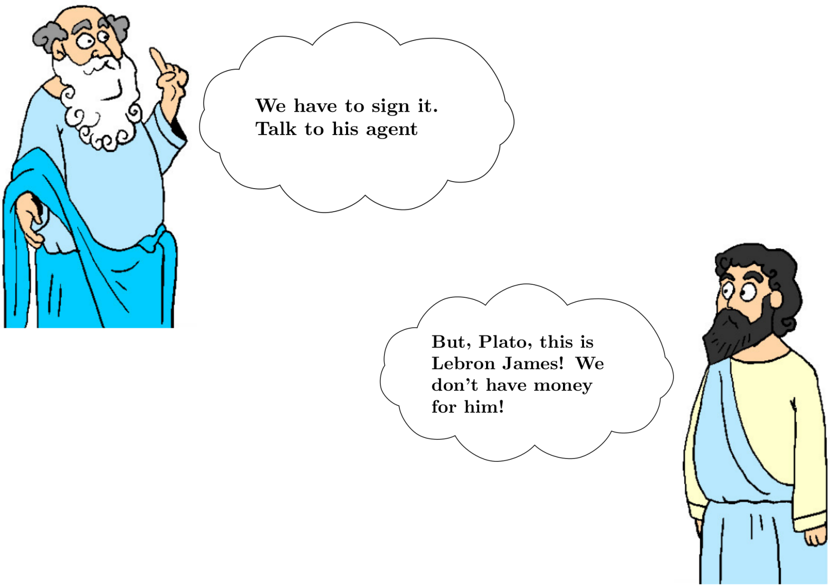
- Green: Point PIR prediction.
- Orange: Upper and lower bound of the PIR prediction intervals.
- Red: PER observed value.

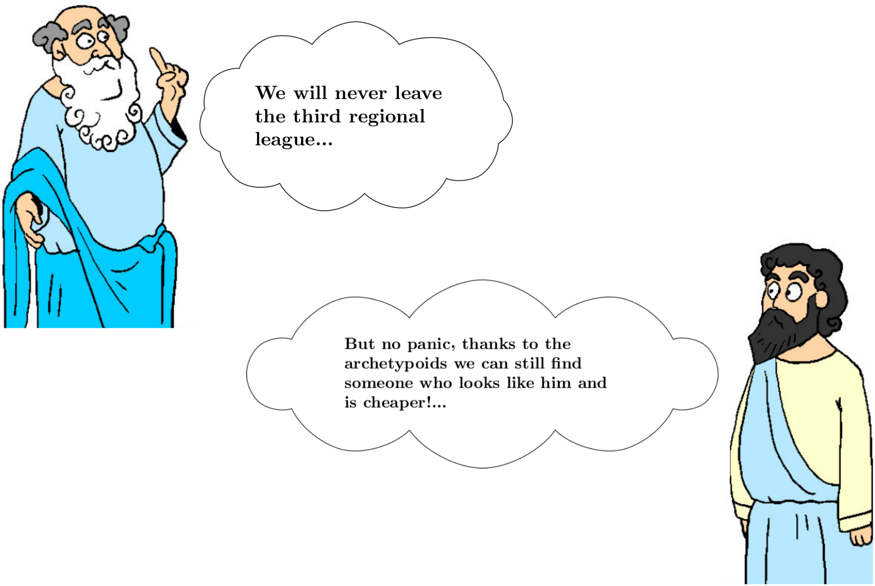
- The PER obtained by Doncic is similar to the upper bound of prediction.
- In three out of five cases, it is within the prediction interval. In the other two, it surpasses it.

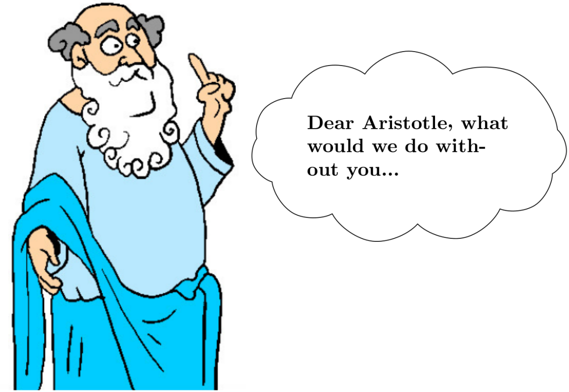
BUT LET'S GO BACK TO OUR COACHES...

Dear Aristotle, you have convinced me.

I have applied the analysis of archetypoids and I have come out as the best player...


We have to sign it. Talk to his agent
But, Plato, this is Lebron James! We don't have money for him!


We will never leave the third regional league...
But no panic, thanks to the archetypoids we can still find someone who looks like him and is cheaper!...


Dear Aristotle, what would we do without you...

The analysis of archetypoids applied to basketball allows us to:

The analysis of archetypoids applied to basketball allows us to:

1. Identify representative patterns of performance different from the rest, embodied in specific players (archetypoids).

The analysis of archetypoids applied to basketball allows us to:

1. Identify representative patterns of performance different from the rest, embodied in specific players (archetypoids).

2. Relate the other players to the archetypoids based on a similar performance, and, consequently, help in the search for players with high performance, but more affordable in terms of economic requirements.

The analysis of archetypoids applied to basketball allows us to:

1. Identify representative patterns of performance different from the rest, embodied in specific players (archetypoids).

2. Relate the other players to the archetypoids based on a similar performance, and, consequently, help in the search for players with high performance, but more affordable in terms of economic requirements.

3. Combining it with functional data, give a forecast of player development.

Publication

Received: 11 January 2018 | Revised: 30 July 2019 | Accepted: 5 August 2019
DOI: 10.1002/sam.11436

RESEARCH ARTICLE

WILEY

Forecasting basketball players' performance using sparse functional data

Guillermo Vinué[1] | **Irene Epifanio**[2]

[1]Department of Statistics and O.R., University of Valencia, Burjassot, Spain
[2]Departamento de Matemàtiques and Institut de Matemàtiques i Aplicacions de Castelló, Universitat Jaume I, Castelló, Spain

Correspondence
Guillermo Vinué, Department of Statistics and O.R., University of Valencia, 46100 Burjassot, Spain.
Email: guillermo.vinue@uv.es

Funding information
Ministerio de Ciencia, Innovación y Universidades (AEI/FEDER, UE), Grant DPI2017-87333-R; Universitat Jaume I, UJI-B2017-13

Abstract
Statistics and analytic methods are becoming increasingly important in basketball. In particular, predicting players' performance using past observations is a considerable challenge. The purpose of this study is to forecast the future behavior of basketball players. The available data are sparse functional data, which are very common in sports. So far, however, no forecasting method designed for sparse functional data has been used in sports. A methodology based on two methods to handle sparse and irregular data, together with the analogous method and functional archetypoid analysis is proposed. Results in comparison with traditional methods show that our approach is competitive and additionally provides prediction intervals. The methodology can also be used in other sports when sparse longitudinal data are available.

KEYWORDS
archetypal analysis, basketball, forecasting, functional data analysis, functional sparse data

- Impact factor: 1.396 ; Position: 50/124 = 0.4 (Statistics & Probability) ; Base: JCR (2019) ; Cites: 17 ; https://doi.org/10.1089/big.2018.0124

Conclusions

- Pioneering studies in European and Spanish basketball. Placing Spanish basketball at the level of the NBA also at the forefront of statistical analysis.

- Data treatment: Proposal for improvement through visualizations of the traditional stats and accomplishment of new analyses with play-by-play data, not available until now in Spanish basketball.

- Research line: Publications in both scientific and dissemination journals. Contribution to the free code software.

- Complement, not replace, the technical knowledge of sports directors and coaches.

- Extension to any sport, both male and female.

- Expansion of journalistic coverage. More information for anyone interested in basketball.

- Real importance of having all the possible knowledge. Example of Kevin de Bruyne.

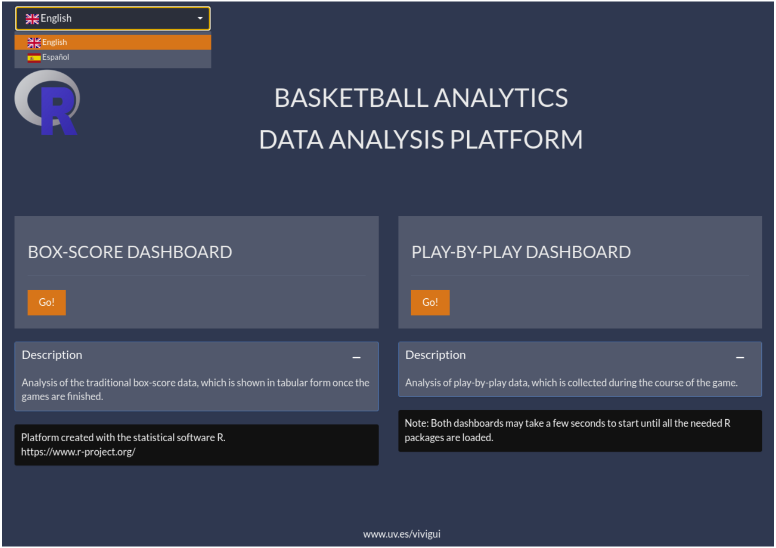
Alberto P. Sierra alberpsierra
Actualizado a 17 de abril de 2021 11:22 CEST

INGLATERRA

Kevin De Bruyne lidera la revolución de los contratos

Kevin de Bruyne usó un estudio de Big Data sobre su juego en el Manchester City para demostrar su valor al club de manera tangible. Ha renovado por un 30% más.

Work in progress



English

Season

2022-2023

Home page

Operations

Lineups

Possessions

Timeouts

Periods

Fouls

Offensive rebounds

Play-by-play dashboard.

This dashboard allows the user to visualize and analyze the play-by-play data coming from the regular season games of the Spanish ACB basketball league.

Direct access to every tab:

Lineups
Possessions
Timeouts
Periods
Fouls
Offensive rebounds
Software

Author:

Guillermo Vinue

guillermovinue@gmail.com

CRAN.R-project.org/package=BAwiR

www.uv.es/vivigui

Thank you for using this app.

Team

Valencia Basket

Player of interest

Alexander

Position to combine

Base

Player(s) to combine

Evans Harper

Combination

Pareja

Period

Todos

Rival(s)

Todos

Get results

Games Summary Players

Show 5 **entries** **Search:**

day	game	period	lineup	time_seconds	plus_minus	avg	score	result
2	Valencia Basket - Joventut Badalona	1C	Alexander, Harper	64	1	0.016	101 - 97	victoria
2	Valencia Basket - Joventut Badalona	1C-2C	Alexander, Harper	190	2	0.011	101 - 97	victoria
2	Valencia Basket - Joventut Badalona	2C	Alexander, Harper	179	4	0.022	101 - 97	victoria
2	Valencia Basket - Joventut Badalona	4C	Alexander, Harper	137	1	0.007	101 - 97	victoria
2	Valencia Basket - Joventut Badalona	4C	Alexander, Harper	51	-1	-0.02	101 - 97	victoria

Showing 1 to 5 of 109 entries

Previous 1 2 3 4 5 … 22 Next

Basketball Big Data Platform for Box-Score and Play-By-Play Data

G. Vinué, PhD

Spain

Abstract

This is the second part of a research diptych devoted to improve the basketball data management in Spain. The Spanish ACB (Association of Basketball Clubs, acronym in Spanish for the male professional league) is the top European national competition. It attracts most of the best foreign players out of the NBA (National Basketball Association, acronym for the male professional league in USA) and accelerates the development of Spanish players that ultimately contribute to the successes of the Spanish national team. However, this excellence in sportive terms is not reciprocated by an advanced treatment of the data generated by teams and players, the so called *statistics*. On the contrary, the statistics management is still very rudimentary. A first paper published in this journal introduced in 2020 the first open web application for interactive visualization of the box-score data, those data provided once the game is finished. By following the same inspiration, this new research wishes to present the work done with one-step further advanced data, namely, play-by-play data, that is provided as the game develops. This type of data allows us to extract a deeper knowledge of basketball performance. A novel data platform encompassing the visualization of both box-score and play-by-play data is available at.... The R package presented in the first paper is updated with new functions.

References I

A. Cutler and L. Breiman.
Archetypal Analysis.
Technometrics, 36(4):338–347, 1994.
http://dx.doi.org/10.2307/1269949.

A. Dokumentov and R. J. Hyndman.
Low-dimensional decomposition, smoothing and forecasting of sparse functional data.
http://robjhyndman.com/papers/ROPES.pdf. Working paper, 1-31, 2016.

I. Epifanio.
Functional archetype and archetypoid analysis.
Computational Statistics & Data Analysis, 104:24–34, 2016.
https://doi.org/10.1016/j.csda.2016.06.007.

S. Munzert, C. Rubba, P. Meissner, and D. Nyhuis.
Automated Data Collection with R. A Practical Guide to Web Scraping and Text Mining.
John Wiley & Sons Ltd, 2015.

G. Vinué.
A Web Application for Interactive Visualization of European Basketball Data.
Big Data, 8(1):70–86, 2020.
https://doi.org/10.1089/big.2018.0124.

References II

G. Vinué and I. Epifanio.
Archetypoid Analysis for Sports Analytics.
Data Mining and Knowledge Discovery, 31(6):1643–1677, 2017.
https://doi.org/10.1007/s10618-017-0514-1.

G. Vinué and I. Epifanio.
Forecasting basketball players' performance using sparse functional data.
Statistical Analysis and Data Mining: The ASA Data Sci Journal, 12(6):534–547, 2019.
https://doi.org/10.1002/sam.11436.

G. Vinué, I. Epifanio, and S. Alemany.
Archetypoids: A new approach to define representative archetypal data.
Computational Statistics & Data Analysis, 87:102–15, 2015.
https://doi.org/10.1016/j.csda.2015.01.018.

H. Wickham et al.
Welcome to the tidyverse.
Journal of Open Source Software, 4(43):1686, 2019.
10.21105/joss.01686.

THANK FOR THE ATTENTION

guillermovinue@gmail.com

www.uv.es/vivigui
www.sportsandvisualdata.com