

# Basketball data science



MARICA MANISERA – PAOLA ZUCCOLOTTO  
UNIVERSITY OF BRESCIA, Italy

VIENNA, April 13, 2018



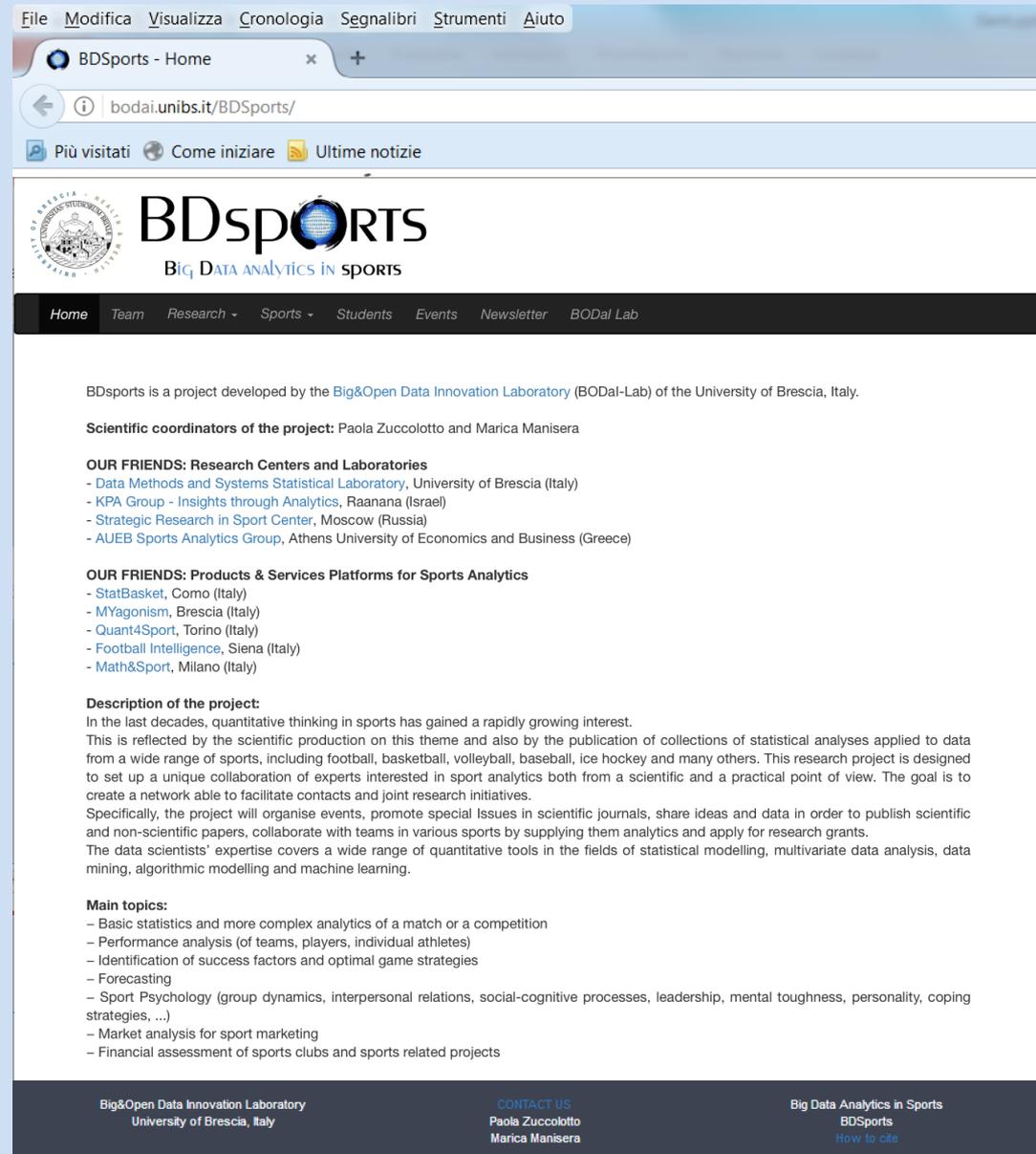
[paola.zuccolotto@unibs.it](mailto:paola.zuccolotto@unibs.it)

[marica.manisera@unibs.it](mailto:marica.manisera@unibs.it)



# BDSports, a network of people interested in Sports Analytics

<http://bodai.unibs.it/bdsports/>



The screenshot shows the BDSports website homepage. The browser address bar displays "bodai.unibs.it/BDSports/". The page features the BDSports logo, which includes the University of Brescia emblem and the text "BDspORTS Big DATA ANALYTICS IN SPORTS". A navigation menu is located below the logo, with options: Home, Team, Research, Sports, Students, Events, Newsletter, and BODal Lab. The main content area contains the following text:

BDSports is a project developed by the [Big&Open Data Innovation Laboratory \(BODal-Lab\)](#) of the University of Brescia, Italy.

**Scientific coordinators of the project:** Paola Zuccolotto and Marica Manisera

**OUR FRIENDS: Research Centers and Laboratories**

- [Data Methods and Systems Statistical Laboratory](#), University of Brescia (Italy)
- [KPA Group - Insights through Analytics](#), Raanana (Israel)
- [Strategic Research in Sport Center](#), Moscow (Russia)
- [AUEB Sports Analytics Group](#), Athens University of Economics and Business (Greece)

**OUR FRIENDS: Products & Services Platforms for Sports Analytics**

- [StatBasket](#), Como (Italy)
- [MYagonism](#), Brescia (Italy)
- [Quant4Sport](#), Torino (Italy)
- [Football Intelligence](#), Siena (Italy)
- [Math&Sport](#), Milano (Italy)

**Description of the project:**

In the last decades, quantitative thinking in sports has gained a rapidly growing interest. This is reflected by the scientific production on this theme and also by the publication of collections of statistical analyses applied to data from a wide range of sports, including football, basketball, volleyball, baseball, ice hockey and many others. This research project is designed to set up a unique collaboration of experts interested in sport analytics both from a scientific and a practical point of view. The goal is to create a network able to facilitate contacts and joint research initiatives. Specifically, the project will organise events, promote special Issues in scientific journals, share ideas and data in order to publish scientific and non-scientific papers, collaborate with teams in various sports by supplying them analytics and apply for research grants. The data scientists' expertise covers a wide range of quantitative tools in the fields of statistical modelling, multivariate data analysis, data mining, algorithmic modelling and machine learning.

**Main topics:**

- Basic statistics and more complex analytics of a match or a competition
- Performance analysis (of teams, players, individual athletes)
- Identification of success factors and optimal game strategies
- Forecasting
- Sport Psychology (group dynamics, interpersonal relations, social-cognitive processes, leadership, mental toughness, personality, coping strategies, ...)
- Market analysis for sport marketing
- Financial assessment of sports clubs and sports related projects

The footer contains contact information for the Big&Open Data Innovation Laboratory at the University of Brescia, Italy, and lists the scientific coordinators: Paola Zuccolotto and Marica Manisera. It also includes the text "Big Data Analytics in Sports BDSports" and a link to "How to use".

- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

## AGENDA:

- Basketball analytics: state of the art
- Basketball datasets
- Case studies:
  -  CS1: new positions in basketball
  -  CS2: scoring probability when shooting under high-pressure conditions
  -  CS3: performance variability and teamwork assessment
  -  CS4: sensor data analysis
- Concluding remarks



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

# BASKETBALL ANALYTICS



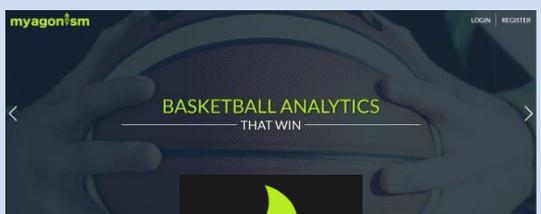
Official  
Statistics

Sport  
Analytics  
Services

Scientific  
Research



#	Team	GP	MPG	PPG	PTS	FGM-FGA	FG%	3PM-3PA	3P%	FTM-FTA	FT%
1.	USA	8	200.0	100.9	807	34.8-73.8	47.1	10.4-28.1	36.9	21.0-28.0	75.0
2.	Argentina	6	208.3	84.5	519	29.7-71.7	41.4	11.3-32.3	35.1	15.8-23.0	68.8
3.	Spain	8	200.0	86.1	689	30.4-68.8	46.2	9.8-26.5	36.8	18.6-21.1	74.0
4.	Australia	8	200.0	85.4	683	31.9-64.3	49.4	7.4-22.3	33.1	14.3-18.6	76.5
5.	Serbia	8	200.0	83.1	660	29.6-62.3	46.7	7.4-22.3	33.1	16.3-21.4	77.2
6.	Brazil	5	210.0	82.2	411	29.6-66.4	45.6	6.4-21.6	29.6	16.2-22.2	73.0
7.	France	4	200.0	81.7	490	30.7-60.3	51.0	9.3-19.3	36.3	14.8-18.8	80.3
8.	Croatia	4	200.0	80.5	483	26.4-59.3	45.5	8.3-23.8	35.0	18.5-24.2	76.4
9.	Nigeria	8	200.0	78.4	392	27.4-62.3	43.7	10.3-27.8	36.7	13.0-17.3	75.4
10.	Lithuania	6	200.0	74.0	456	29.0-62.2	46.6	7.0-22.3	31.6	10.5-14.3	73.3
11.	China	5	200.0	83.6	318	22.0-52.8	41.7	3.8-15.4	24.7	18.8-22.4	70.9
12.	Venezuela	5	200.0	63.0	315	22.4-60.4	37.0	4.8-20.4	23.5	13.4-17.6	76.1

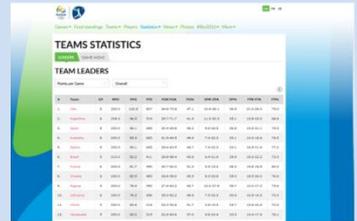


- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

# BASKETBALL ANALYTICS

Official Statistics



Sport Analytics Services



Scientific Research

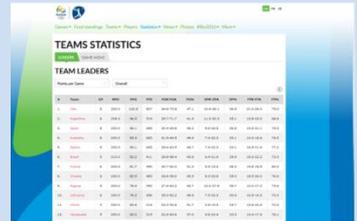


- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

# BASKETBALL ANALYTICS

Official Statistics



Sport Analytics Services



Scientific Research

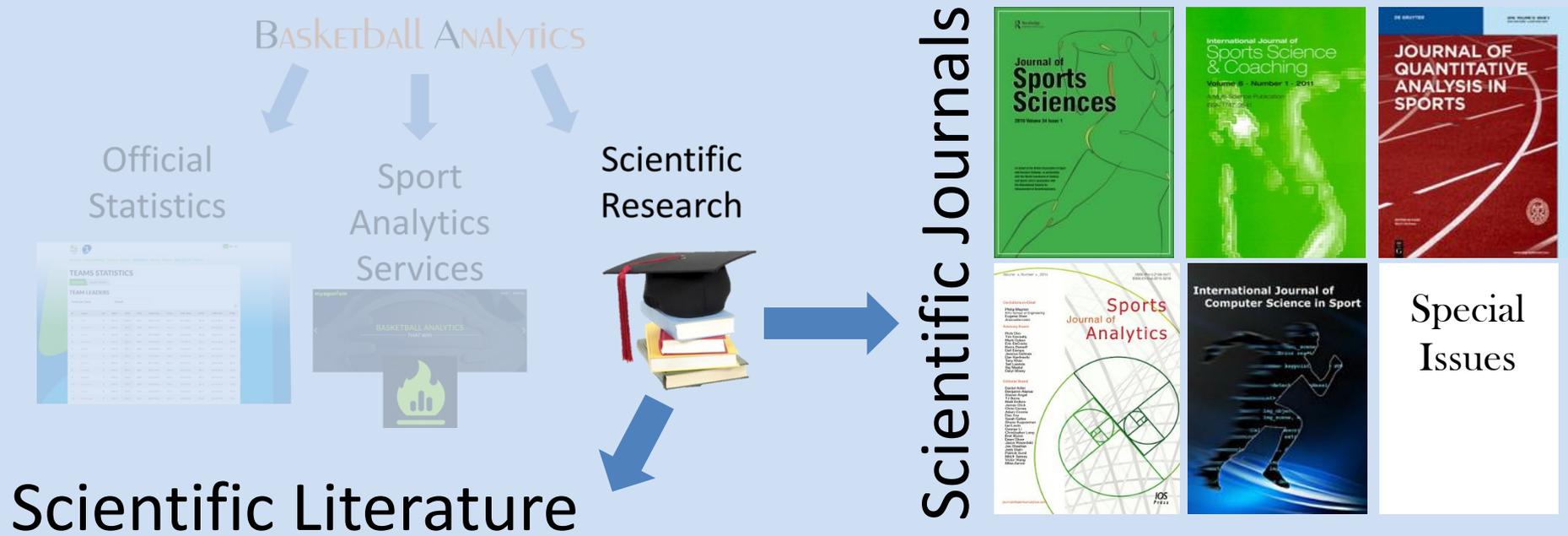


Our analyses often **integrate** machine learning tools and experts' suggestions



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis



## Scientific Literature

As concerns basketball, several statistical techniques have been applied to analyze data with a great variety of different aims, ranging from simply depicting the main features of a game by means of descriptive statistics (Kubatko et al., 2007) to the investigation of more complex problems, such as forecasting the outcomes of a game or a tournament (West et al., 2008; Loeffelholz et al., 2009; Brown et al., 2010; Gupta, 2015; Lopez and Matthews, 2015; Ruiz and Perez-Cruz, 2015; Yuan et al., 2015; Manner, 2016), analysing players' performance (Page et al., 2007; Cooper et al., 2009; Piette et al., 2010; Fearnhead and Taylor, 2011; Ozmen, 2012; Page et al., 2013; Deshpande and Jensen, 2016), studying the network of players' pathways from the in-bounds pass to the basket (Skinner, 2010) and their spatial positioning (Shortridge et al., 2014), or identifying optimal game strategies (Annis et al., 2006).



- Basketball Analytics: state of the art
- **Basketball datasets**
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

**Data**



**Big Data**



- Basketball Analytics: state of the art
- **Basketball datasets**
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

**Data**



**Big Data**

Year: 2016-17

TEAM LEADERS

**POINTS** Kevin Durant #35 **25.8**

**REBOUNDS** Kevin Durant #35 **8.3**

**ASSISTS** Draymond Green #23 **7.3**

**STEALS** Draymond Green #23 **2.1**

**BLOCKS** Kevin Durant #35 **1.7**

Splits

TOTAL SPLITS

GAME STATISTICS	GP	GS	MIN	PPG	OFFR	DEFR	RPG	APG	SPG	BPG	TPG	FPG	A/TO	PER
PLAYER	56	56	34.1	25.8	0.6	7.6	8.3	4.9	1.13	1.70	2.3	1.9	2.1	27.6
Kevin Durant, SF	55	55	33.4	24.7	0.6	3.7	4.3	6.4	1.65	0.20	2.9	2.2	2.2	24.0
Stephen Curry, PG	54	54	34.1	22.1	0.7	3.1	3.8	2.0	0.81	0.46	1.8	1.9	1.1	17.1
Klay Thompson, SG	53	53	32.9	10.2	1.4	6.8	8.2	7.3	2.09	1.51	2.3	3.0	3.2	13.5
Draymond Green, PF	51	0	14.3	6.5	0.2	1.2	1.4	1.3	0.55	0.12	0.6	1.0	1.9	13.0
Ian Clark, SG	54	0	25.6	6.4	0.7	3.2	3.9	3.4	0.93	0.39	0.7	1.2	4.9	23.7
Andre Iguodala, SF	51	10	9.5	6.2	1.1	2.0	3.0	0.2	0.24	0.69	0.6	1.5	0.4	15.9
JaVale McGee, C	51	10	9.5	6.2	2.1	3.9	6.0	2.0	0.95	0.41	1.3	2.4	1.6	15.9
Zaza Pachulia, C	44	44	18.8	5.9	0.4	1.6	1.9	1.7	0.47	0.22	0.8	1.6	2.1	10.1
Shaun Livingston, PG	51	1	17.3	5.2	0.6	2.1	2.7	2.1	0.67	0.45	1.0	1.4	2.1	16.6
David West, PF	42	0	11.5	4.1	0.2	0.8	1.0	1.0	0.35	0.22	0.5	0.8	2.0	7.9
Patrick McCaw, PG	46	3	12.4	3.3	0.2	0.8	1.0	1.0	0.30	0.37	0.3	1.4	1.9	14.5
Kevon Looney, SF	43	3	9.3	2.9	0.8	1.7	2.6	0.6	0.24	0.52	0.4	0.9	1.0	11.6
James Michael McAdoo, SF	29	0	9.1	2.8	0.4	1.1	1.5	0.4	0.24	0.52	0.4	0.6	2.5	5.0
Briante Weber, PG	5	0	8.2	1.6	0.0	0.6	0.6	1.0	0.40	0.20	0.4	0.6	1.1	0.0
Damian Jones, C	8	0	5.6	1.4	0.8	0.6	1.4	0.0	0.13	0.25	0.6	1.1	1.3	9.4
Anderson Varejao, C	14	1	6.6	1.3	0.9	1.1	1.9	0.7	0.21	0.21	0.6	1.1	1.3	9.4
<b>Totals</b>	56	--	--	118.2	8.8	35.7	44.5	31.0	9.55	6.57	14.4	19.2	2.2	--

**Stats**  
**CS1**



[www.espn.com/nba](http://www.espn.com/nba)  
[stats.nba.com](http://stats.nba.com)  
[www.fiba.com](http://www.fiba.com)  
 Leagues

...

- Basketball Analytics: state of the art
- **Basketball datasets**
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

**Data**



**Big Data**

Year: 2016-17

TEAM LEADERS

POINTS	REBOUNDS	ASSISTS	STEALS	BLOCKS
Kevin Durant #35 <b>25.8</b>	Kevin Durant #35 <b>8.3</b>	Draymond Green #23 <b>7.3</b>	Draymond Green #23 <b>2.1</b>	Kevin Durant #35 <b>1.7</b>

TOTAL SPLITS

GAME STATISTICS

PLAYER	GP	GS	MIN	PPG	FF	D	TRP%	APG	SPG	BPG	TPG	FPG	A/TO	PER
Kevin Durant, SF	56	56	34.1	25.8	0.6	3.7	4.3	6.4	1.65	0.20	2.3	1.9	2.1	27.6
Stephen Curry, PG	55	55	33.4	24.7	0.7	3.1	3.8	2.0	0.81	0.46	1.8	1.9	1.1	17.1
Klay Thompson, SG	54	54	34.1	22.1	1.4	6.8	8.2	7.3	2.09	1.51	2.3	3.0	3.2	17.2
Draymond Green, PF	53	53	32.9	10.2	1.4	6.8	8.2	7.3	2.09	1.51	2.3	3.0	3.2	13.5
Ian Clark, SG	54	0	25.6	6.4	0.2	1.2	1.4	1.3	0.55	0.12	0.6	1.0	1.9	13.0
Andre Iguodala, SF	51	10	9.5	6.2	0.1	2.1	3.0	0.0	0.2	0.24	0.69	0.6	1.5	0.4
JaVale McGee, C	44	44	18.8	5.9	0.1	1.1	1.9	0.0	0.05	0.41	1.3	2.4	1.6	15.9
Zaza Pachulia, C	51	1	17.3	5.2	0.1	1.1	1.9	0.0	0.05	0.41	1.3	2.4	1.6	15.9
Shaun Livingston, PG	42	0	11.5	4.1	0.6	2.1	2.7	0.0	0.05	0.41	1.3	2.4	1.6	15.9
David West, PF	46	3	12.4	3.3	0.2	0.8	1.0	0.0	0.05	0.41	1.3	2.4	1.6	15.9
Patrick McCaw, PG	43	3	9.3	2.9	0.8	1.7	2.6	0.0	0.05	0.41	1.3	2.4	1.6	15.9
Kevon Looney, SF	29	0	9.1	2.8	0.4	1.1	1.5	0.0	0.05	0.41	1.3	2.4	1.6	15.9
James Michael McAdoo, SF	5	0	8.2	1.6	0.0	0.6	0.6	0.0	0.05	0.41	1.3	2.4	1.6	15.9
Briante Weber, PG	8	0	5.6	1.4	0.8	0.6	1.4	0.0	0.05	0.41	1.3	2.4	1.6	15.9
Damian Jones, C	14	1	6.6	1.3	0.9	1.1	1.9	0.0	0.05	0.41	1.3	2.4	1.6	15.9
Anderson Varejao, C	14	1	6.6	1.3	0.9	1.1	1.9	0.0	0.05	0.41	1.3	2.4	1.6	15.9
<b>Totals</b>	<b>56</b>	<b>--</b>	<b>--</b>	<b>118.2</b>	<b>8.8</b>	<b>35.7</b>	<b>44.5</b>							

**Stats CS1**

**play-by-play**  
**CS2 – CS3**



NBA ALL-STAR GAME

Final

Eastern Conf All-Stars 182 vs Western Conf All-Stars 192

1st Quarter

TIME	TEAM	PLAY	SCORE
12:00	LeBron James vs. Anthony Davis	(Stephen Curry gains possession)	0 - 0
11:45	Anthony Davis	makes 21-foot jumper	0 - 2
11:33	DeMar DeRozan	bad pass (Kawhi Leonard steals)	0 - 2
11:29	Kawhi Leonard	makes dunk	0 - 4
11:19	Giannis Antetokounmpo	makes dunk (Jimmy Butler assists)	2 - 4
11:10	Anthony Davis	misses three point jumper	2 - 4
11:08	LeBron James	defensive rebound	2 - 4
11:02	LeBron James	makes 27-foot three point jumper (DeMar DeRozan assists)	5 - 4
10:51	Stephen Curry	makes 26-foot three point jumper	5 - 7
10:42	Jimmy Butler	makes dunk (DeMar DeRozan assists)	7 - 7
10:29	Anthony Davis	makes layup	7 - 9
10:12	Kyrie Irving	makes 25-foot three point jumper (DeMar DeRozan assists)	10 - 9
10:00	Kevin Durant	misses layup	10 - 9
10:00	Kyrie Irving	defensive rebound	10 - 9
9:53	LeBron James	misses layup	10 - 9
9:52	Kawhi Leonard	defensive rebound	10 - 9

```

string input;
int length, in;
double dInput;
bool again = true;

while (again) {
    in = -1;
    again = false;
    getline(cin, input);
    system("cls");
    stringstream(input) >> dInput;
    length = input.length();
    if (length < 4) {
        again = true;
    }
}

```

**WEB SCRAPING SOFTWARE**



- Basketball Analytics: state of the art
- **Basketball datasets**
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

**Data**



**Big Data**

Year: 2016-17

TEAM LEADERS

POINTS	REBOUNDS	ASSISTS	STEALS	BLOCKS
Kevin Durant #35: 25.8	Kevin Durant #35: 8.3	Draymond Green #23: 7.3	Draymond Green #23: 2.1	Kevin Durant #35: 1.7

TOTAL SPLITS

GAME STATISTICS

PLAYER	GP	GS	MIN	PPG	FF	D	TRP%	APG	SPG	BPG	TPG	FPG	A/TO	PER
Kevin Durant, SF	56	56	34.1	25.8	0.6	3.7	4.3	6.4	1.65	0.20	2.9	2.2	2.1	24.0
Stephen Curry, PG	55	55	33.4	24.7	0.7	3.1	3.8	2.0	0.81	0.46	1.8	1.9	1.1	17.1
Klay Thompson, SG	54	54	34.1	22.1	1.4	6.8	8.2	7.3	2.09	1.51	2.3	3.0	3.2	17.2
Draymond Green, PF	53	53	32.9	10.2	1.2	1.4	1.3	0.55	0.12	0.6	1.0	1.9	1.1	17.1
Ian Clark, SG	51	0	14.3	6.5	0.2	3.0	3.9	3.4	0.93	0.39	0.7	1.2	4.9	13.0
Andre Iguodala, SF	54	0	25.6	6.4	0.2	2.1	3.0	0.2	0.24	0.69	0.6	1.5	0.4	23.7
JaVale McGee, C	44	44	18.8	5.9	0.1	1.1	1.9	0.0	0.45	0.41	1.3	2.4	1.6	15.9
Zaza Pachulia, C	51	1	17.3	5.2	0.1	1.1	2.7	0.0	0.0	0.0	0.0	0.0	0.0	10.1
Shaun Livingston, PG	42	0	11.5	4.1	0.6	2.1	2.7	0.0	0.0	0.0	0.0	0.0	0.0	10.1
David West, PF	46	3	12.4	3.3	0.2	0.8	1.0	0.0	0.0	0.0	0.0	0.0	0.0	10.1
Patrick McCaw, PG	43	3	9.3	2.9	0.8	1.7	2.6	0.0	0.0	0.0	0.0	0.0	0.0	10.1
Kevon Looney, SF	29	0	9.1	2.8	0.4	1.1	1.5	0.0	0.0	0.0	0.0	0.0	0.0	10.1
James Michael McAdoo, SF	5	0	8.2	1.6	0.0	0.6	0.6	0.0	0.0	0.0	0.0	0.0	0.0	10.1
Briante Weber, PG	8	0	5.6	1.4	0.8	0.6	1.4	0.0	0.0	0.0	0.0	0.0	0.0	10.1
Damian Jones, C	14	1	6.6	1.3	0.9	1.1	1.9	0.0	0.0	0.0	0.0	0.0	0.0	10.1
Anderson Varejao, C	14	1	6.6	1.3	0.9	1.1	1.9	0.0	0.0	0.0	0.0	0.0	0.0	10.1
<b>Totals</b>	<b>56</b>	<b>--</b>	<b>--</b>	<b>118.2</b>	<b>8.8</b>	<b>35.7</b>	<b>44.5</b>							

**Stats CS1**



**Sensor Data CS4**

Eastern Conf All-Stars 182

Final

TIME	TEAM	PLAY	SCORE
12:00	LeBron James vs. Anthony Davis	(Stephen Curry gains possession)	0-0
12:45	Anthony Davis	makes 21-foot jumper	0-2
13:33	DeMar DeRozan	bad pass (Kawhi Leonard steals)	0-2
13:29	Kawhi Leonard	misses dunk	0-4
13:19	Giannis Antetokounmpo	misses layup (Stephen Curry assists)	0-4
13:10	Anthony Davis	makes 21-foot jumper	0-4
13:08	LeBron James	defensive rebound	2-4
12:02	LeBron James	makes 27-foot three point jumper (DeMar DeRozan assists)	5-4
10:51	Stephen Curry	makes 26-foot three point jumper	5-7
10:42	Jimmy Butler	makes dunk (DeMar DeRozan assists)	7-7
10:29	Anthony Davis	makes layup	7-9
10:12	Kyrie Irving	makes 25-foot three point jumper (DeMar DeRozan assists)	10-9
10:00	Kevin Durant	misses layup	10-9
10:00	Kyrie Irving	defensive rebound	10-9
9:53	LeBron James	misses layup	10-9
9:52	Kawhi Leonard	defensive rebound	10-9

**play-by-play CS2 - CS3**



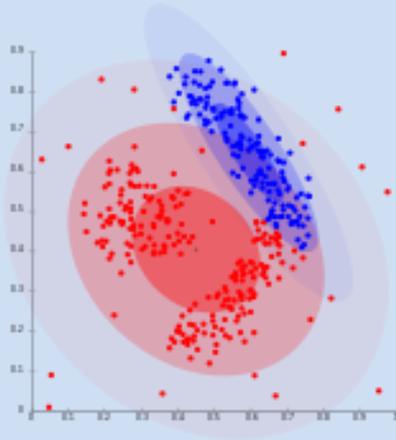
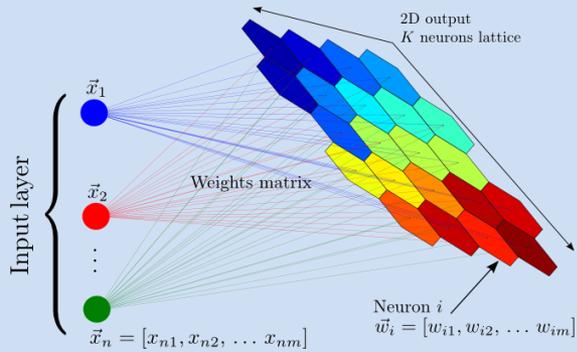
**BDsports**  
BIG DATA ANALYTICS IN SPORTS

MARICA MANISERA  
PAOLA ZUCCOLOTTO – UNIVERSITY OF BRESCIA, ITALY



- Basketball Analytics: state of the art
- Basketball datasets
- **CS1: new positions in basketball**

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis



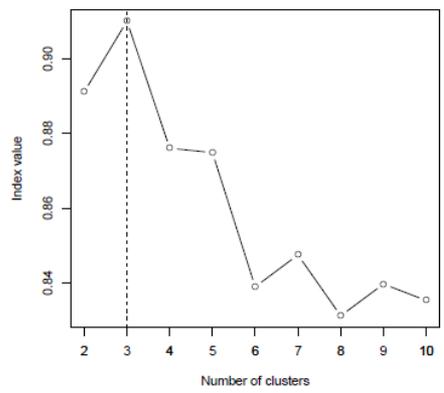
- «Key-players» training set  
→ 7-dimensional **SOM**
- clusterization of the SOM output layer into a proper number of groups by means of a **fuzzy clustering algorithm**

- Basketball Analytics: state of the art
- Basketball datasets
- **CS1: new positions in basketball**

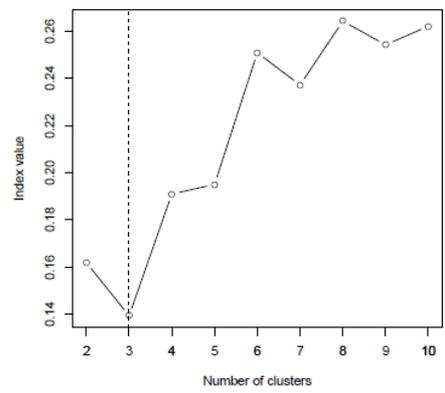
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis



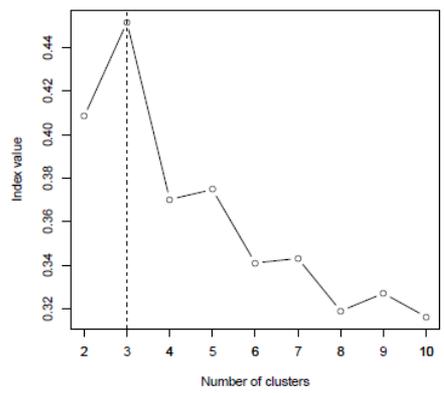
Partition Coefficient



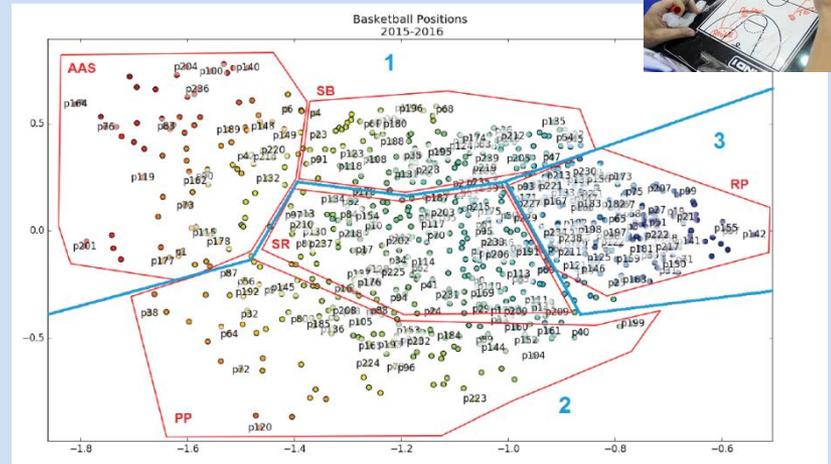
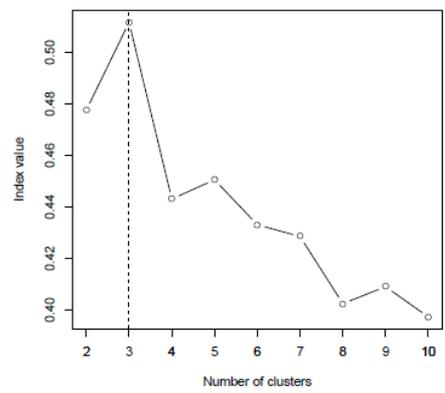
Partition Entropy



Silhouette



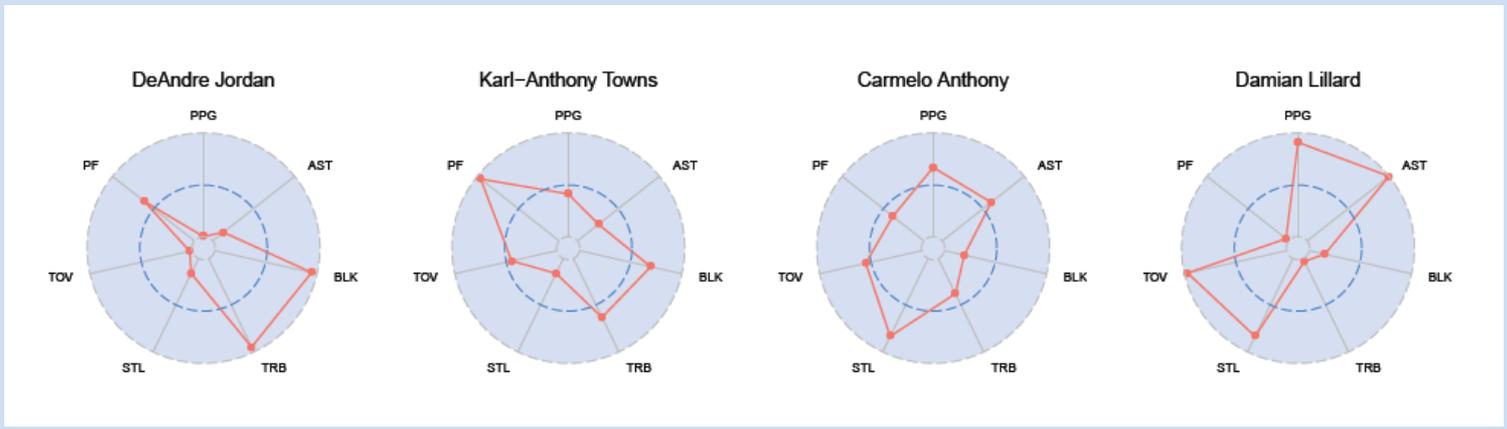
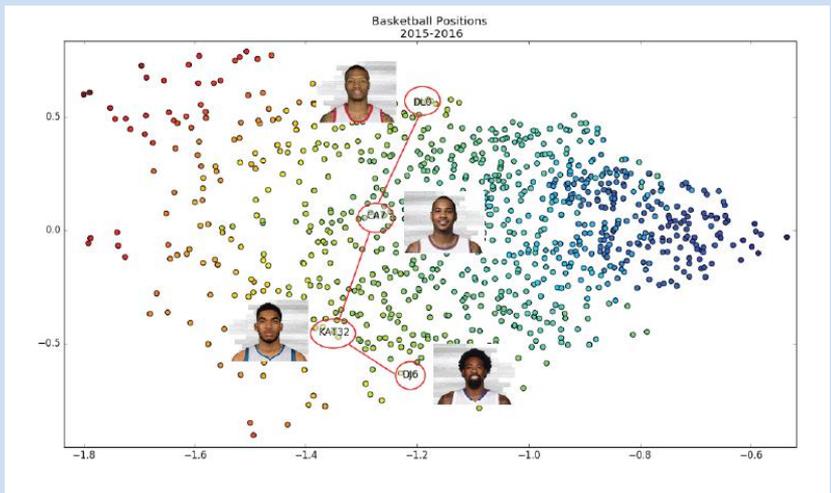
Fuzzy Silhouette



Position	Short	Players
All-Around All Stars	AAS	LeBron James (LBJ23), Kevin Durant (KD35), Paul George (PG13), Stephen Curry (SC30), James Harden (JH13)
Scoring Backcourt	SB	Kobe Bryant (KB24), Damian Lillard (DL0), Kyrie Irving (KI2), Dwyane Wade (DW3)
Scoring Rebounder	SR	Marc Gasol (MG33), Carmelo Anthony (CA7), Karl-Anthony Towns (KAT32), Anthony Davis (AD23), Blake Griffin (BG32)
Paint Protector	PP	DeAndre Jordan (DJ6), Andrew Bogut (AB12), Steven Adams (SA12), Kenneth Faried (KF35)
Role Players	RP	Marco Belinelli (MB3), J.J. Redick (JJ4), Harrison Barnes (HB40), Jabari Parker (JP12), Avery Bradley (AB0)

- Basketball Analytics: state of the art
- Basketball datasets
- **CS1: new positions in basketball**

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis



# CS2: scoring probability when shooting under high-pressure conditions

Big data analytics to model scoring probability in basketball: the effect of shooting under high-pressure conditions

Paola Zuccolotto, Marica Manisera and Marco Sandri



International Journal of Sports Science & Coaching



**MOTIVATION:** Basketball players have often to face high-pressure game conditions. To be aware of the overall and personal reactions to these situations is of primary importance to coaches.

**AIM:** To develop a model describing the impact of some high-pressure game situations on the probability of scoring and to assess players' personal reactions.

- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

## HIGH-PRESSURE GAME SITUATIONS:



- when the shot clock is going to expire (SHOT.CLOCK)
- when the score difference with respect to the opponent is small (Sc.DIFF)
- when the team, for some reason, has globally performed bad during the match, up to the considered moment (MISS.T)
- when the player missed the previous shot (MISS.PL)
- the time to the end of quarter (TIME)
- type of action (POSS.TYPE, 24'' or 14'' extratime)



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis



Dataset	A2ITA	RIO16
Competition	Championship - regular season	Olympic Tournament
Period	2015, 4th Oct - 2016, 23rd Apr	2016, 6th - 21st Aug
Gender	Male	Male
Number of matches	480	38
Number of teams	32	12
Number of players	438	144
Number of 2-point shots	33682 (48.3%, 50.9% Made)	3101 (47.9%, 52.2% Made)
Number of 3-point shots	21163 (30.4%, 34.1% Made)	1780 (27.5%, 33.8% Made)
Number of free throws	14843 (21.3%, 73.5% Made)	1589 (24.6%, 74.8% Made)

69688

6470



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

## DATA MINING TOOLS:

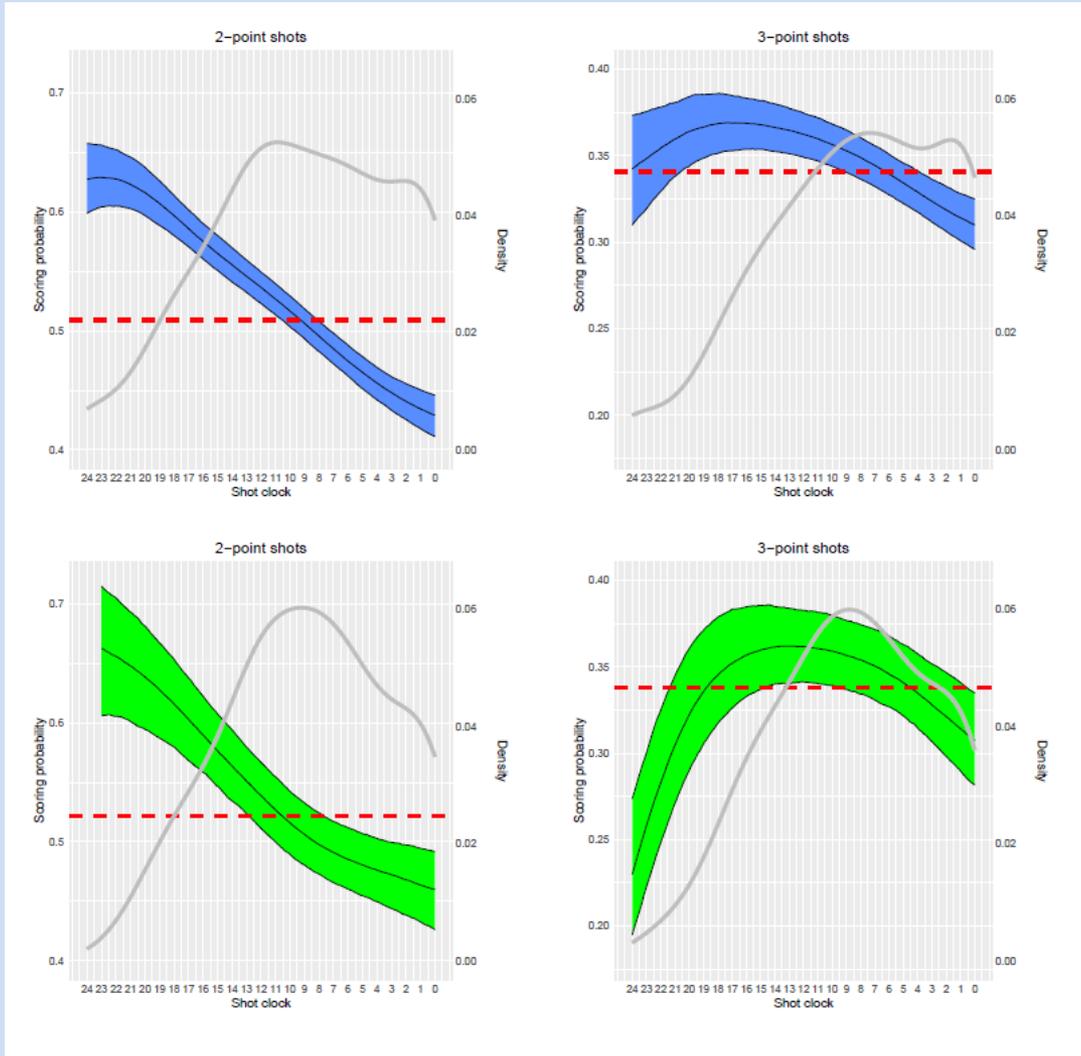
- **univariate** non parametric regressions via kernel smoothing on the dependent variable MADE (assuming values 1 and 0 according to whether, respectively, the attempted shot scored a basket or not)
- 1000 bootstrap samples of size  $n_{boot} = 5000$  and  $n_{boot} = 1000$  for the dataset A2ITA and RIO16, respectively.

➔ few univariate relationships detected - Just SHOT.CLOCK and MISS.PL

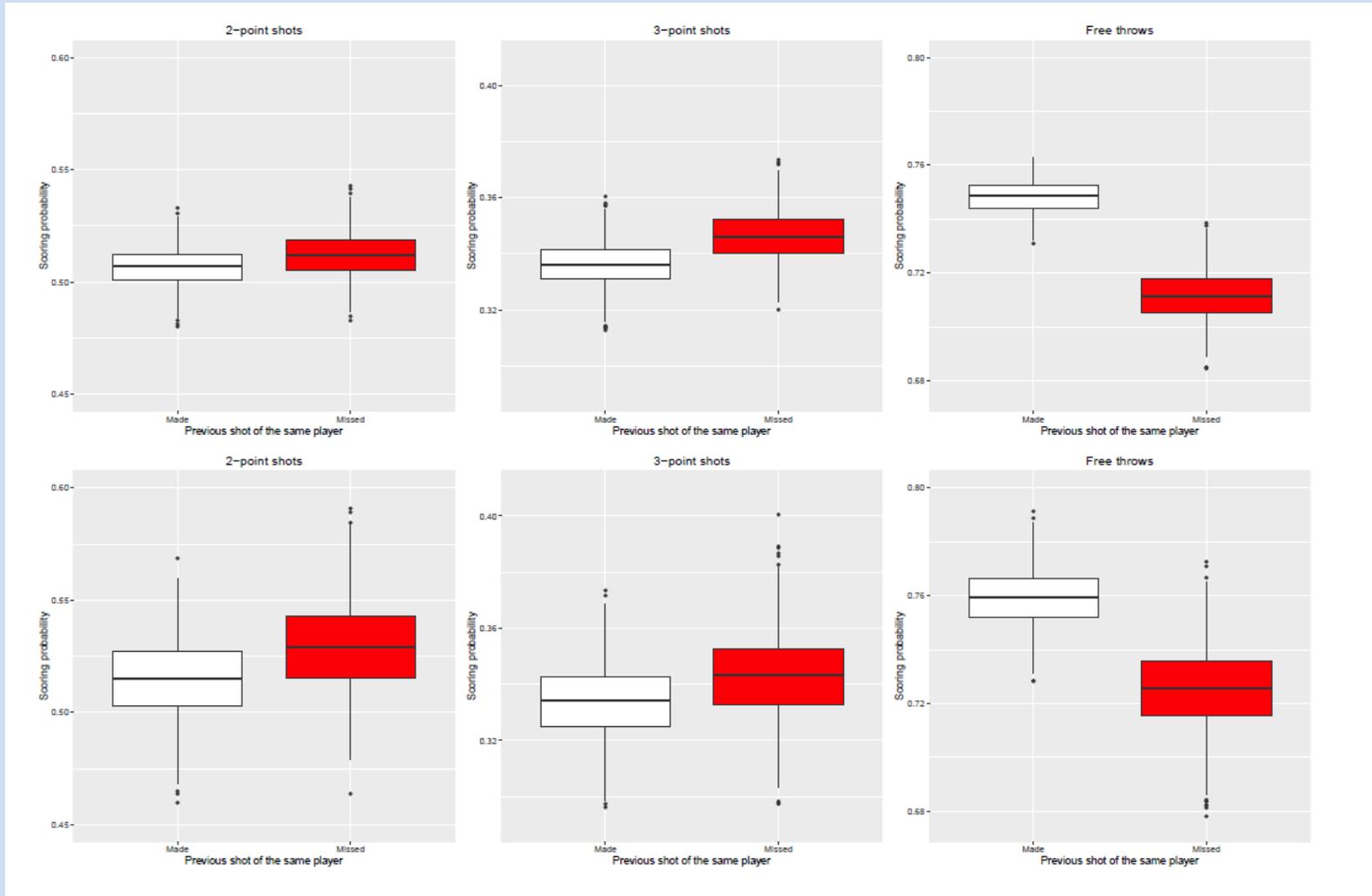


- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

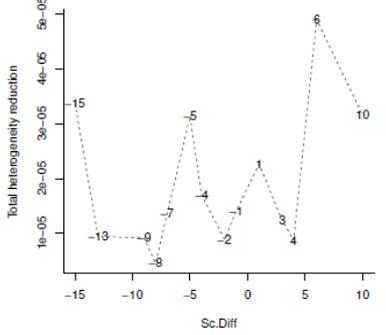
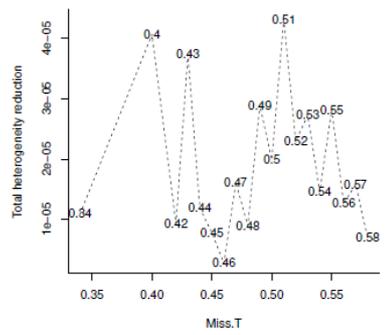
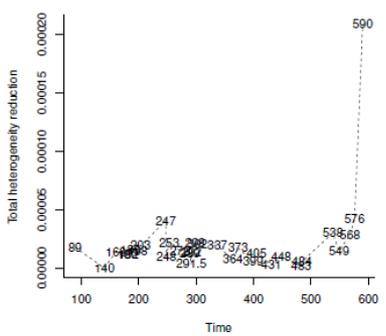
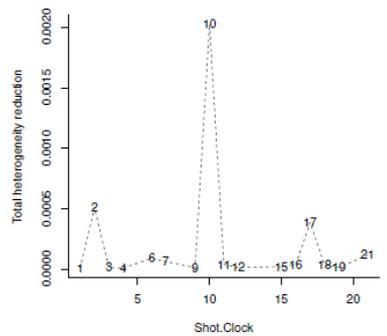
## DATA MINING TOOLS:

- **CART** (Classification And Regression Trees), algorithm able to deal with multivariate complex relationships, also detecting interactions among predictors
- we transform numerical into categorical covariates in order to improve interpretability → combination of the results of a **machine learning procedure** and **experts' suggestions**
- pruning



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

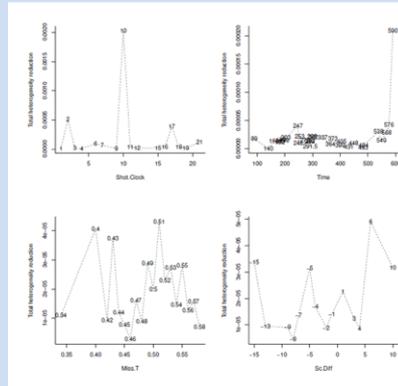


focus on:

- (1) the last 1-2 seconds of possession, very close to the shot clock buzzer sounding,
- (2) games where the score difference is low, for example, between -4 and 4,
- (3) the last 1-2 minutes of each quarter (especially the final quarter)

- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis



focus on:

- (1) the last 1-2 seconds of possession, very close to the shot clock buzzer sounding,
- (2) games where the score difference is low, for example, between -4 and 4,
- (3) the last 1-2 minutes of each quarter (especially the final quarter)

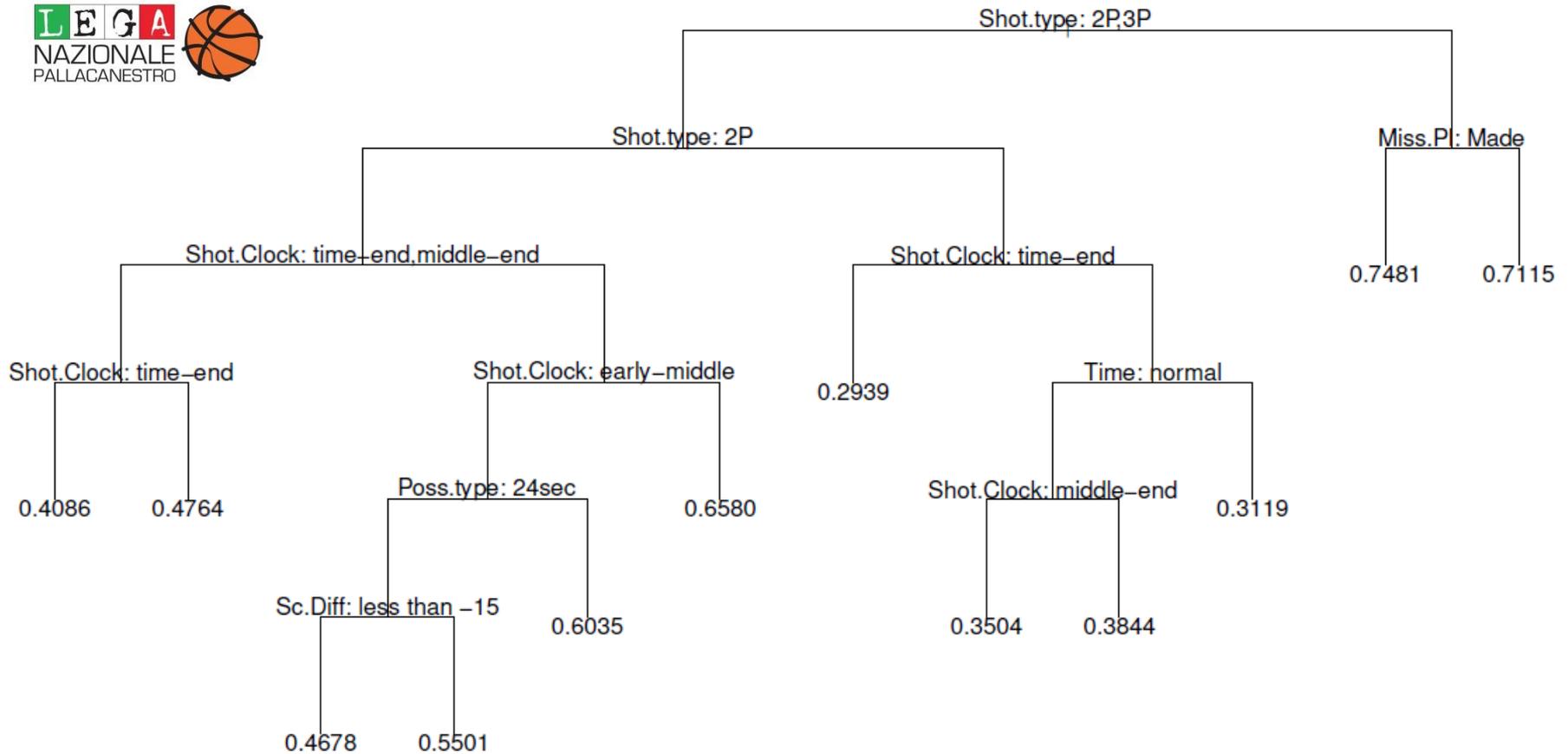
SHOT.CLOCK	<p>early: <math>SHOT.CLOCK &gt; 17</math></p> <p>early-middle: <math>10 &lt; SHOT.CLOCK \leq 17</math></p> <p>middle-end: <math>2 &lt; SHOT.CLOCK \leq 10</math></p> <p>time-end: <math>SHOT.CLOCK \leq 2</math></p>
TIME	<p>normal: <math>TIME \leq 500</math></p> <p>quarter-end: <math>TIME &gt; 500</math></p>
MISS.T	<p>Bad: <math>MISS.T \leq 0.44</math> (25th percentile)</p> <p>Medium: <math>0.44 &lt; MISS.T \leq 0.56</math></p> <p>Good: <math>MISS.T &gt; 0.56</math> (75th percentile)</p>
SC.DIFF	<p>less than -15: <math>SC.DIFF \leq -15</math></p> <p>between -15 and -5: <math>-15 &lt; SC.DIFF \leq -5</math></p> <p>between -5 and 1: <math>-5 &lt; SC.DIFF \leq 1</math></p> <p>between 1 and 6: <math>1 &lt; SC.DIFF \leq 6</math></p> <p>more than 6: <math>SC.DIFF &gt; 6</math></p>



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

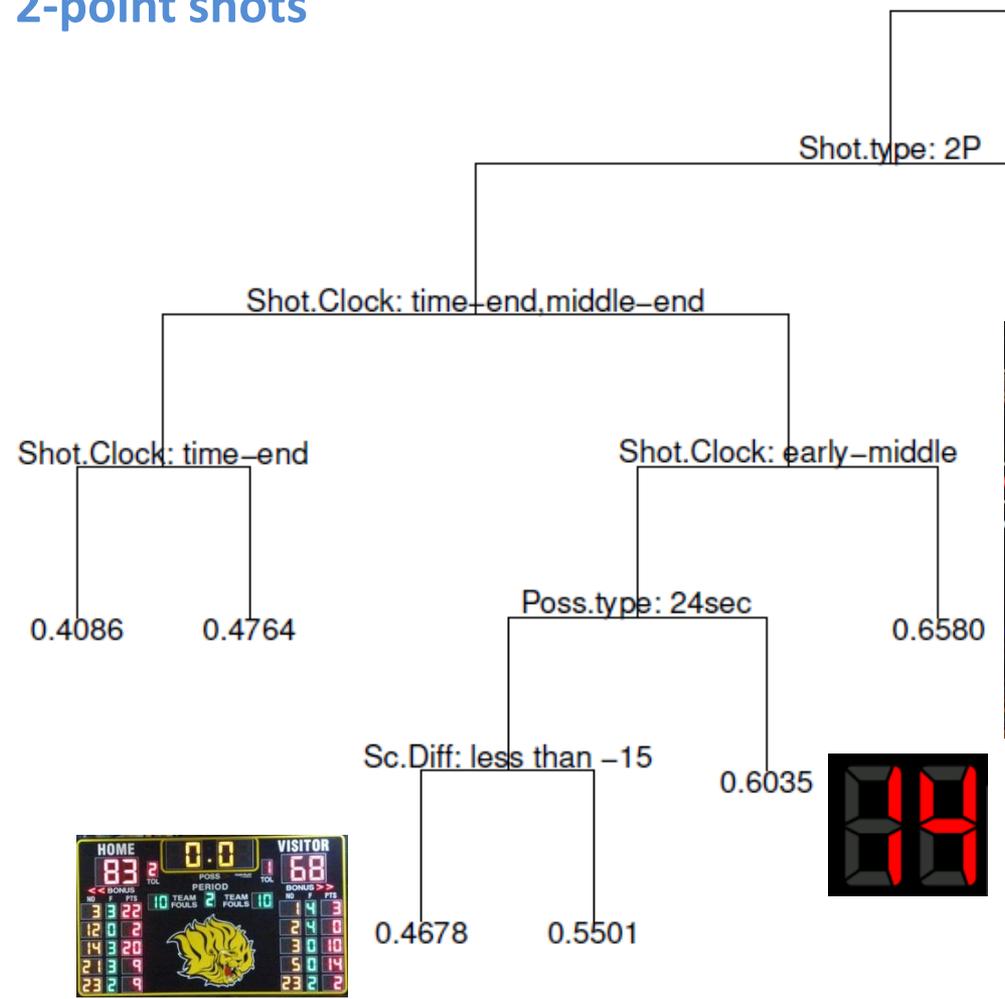
## Very similar results with Rio 2016 data



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

## 2-point shots



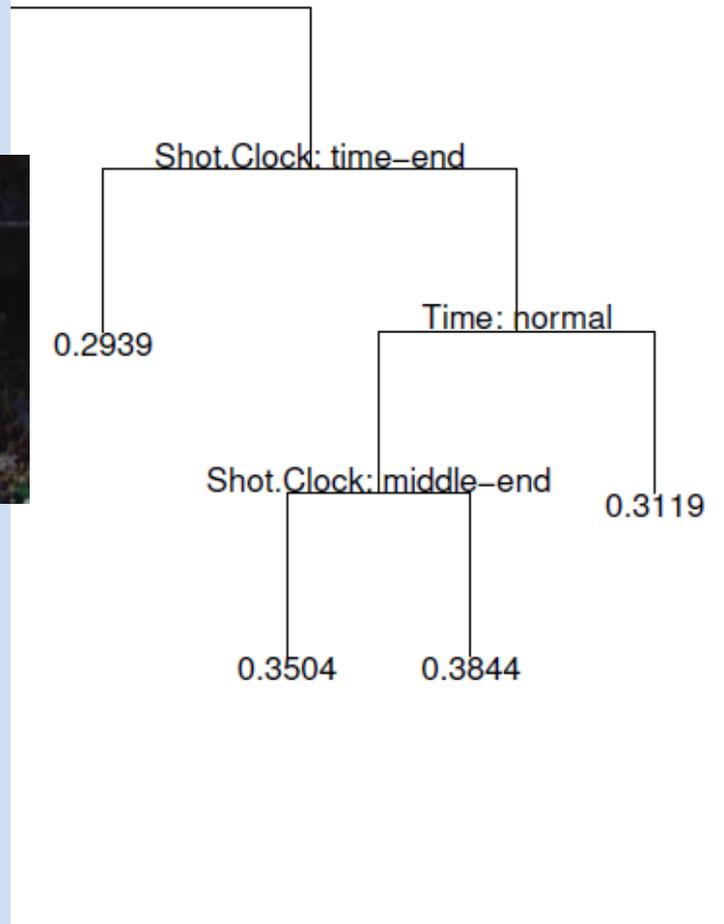
HOME				VISITOR			
PTS	REB	AST	TOV	PTS	REB	AST	TOV
83	2			68	1		
3	3	22	10	10	2	10	10
12	0	2		2	4	3	
14	3	20		3	0	10	
2	3	9		2	0	14	
23	2	9		2	2	2	



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

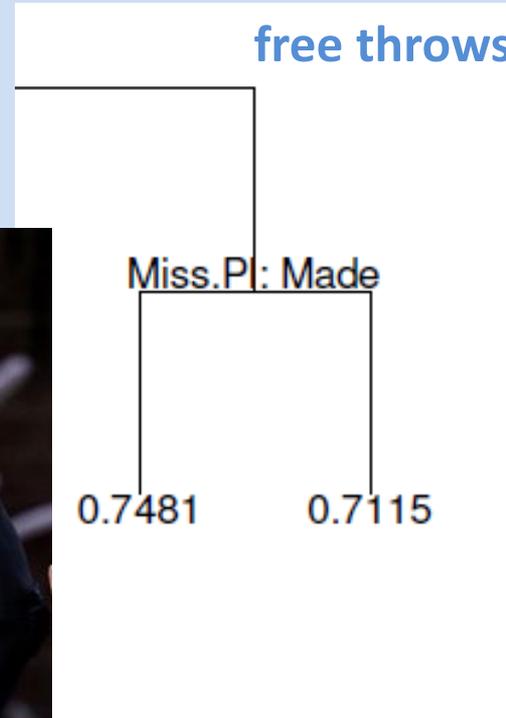
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

## 3-point shots



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

## NEW SHOOTING PERFORMANCE MEASURE:

Takes into account that shots attempted in different moments have different scoring probabilities

Performance of Player  $i$

$$P_i(T) = \text{av}_{j \in J_T} (x_{ij} - \pi_{ij})$$

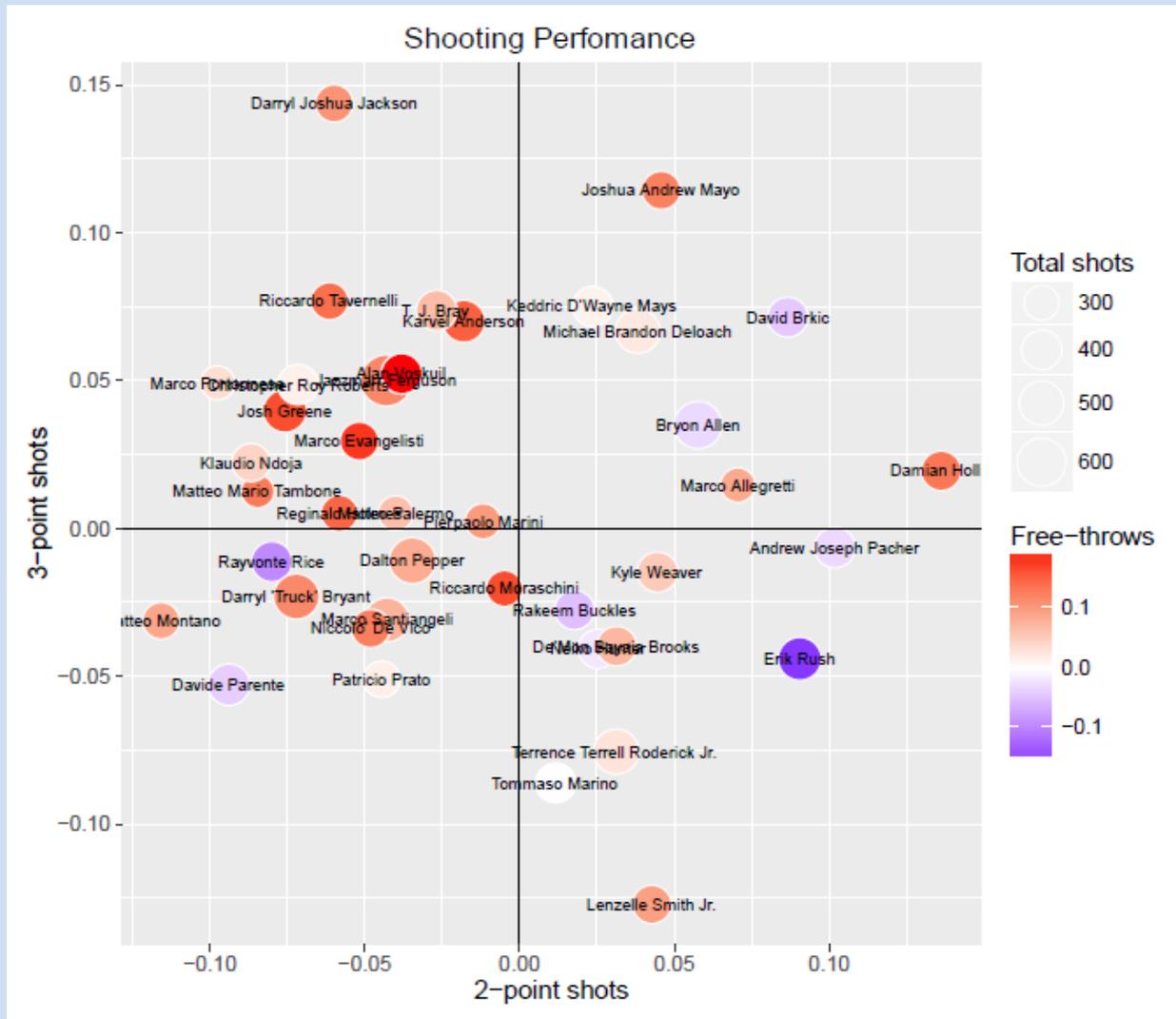
for shot type  $T$   
(2P, 3P, FT)

$j$ -th shot made (1)  
or missed (0)

scoring probability  
of  $j$ -th shot  
according to CART

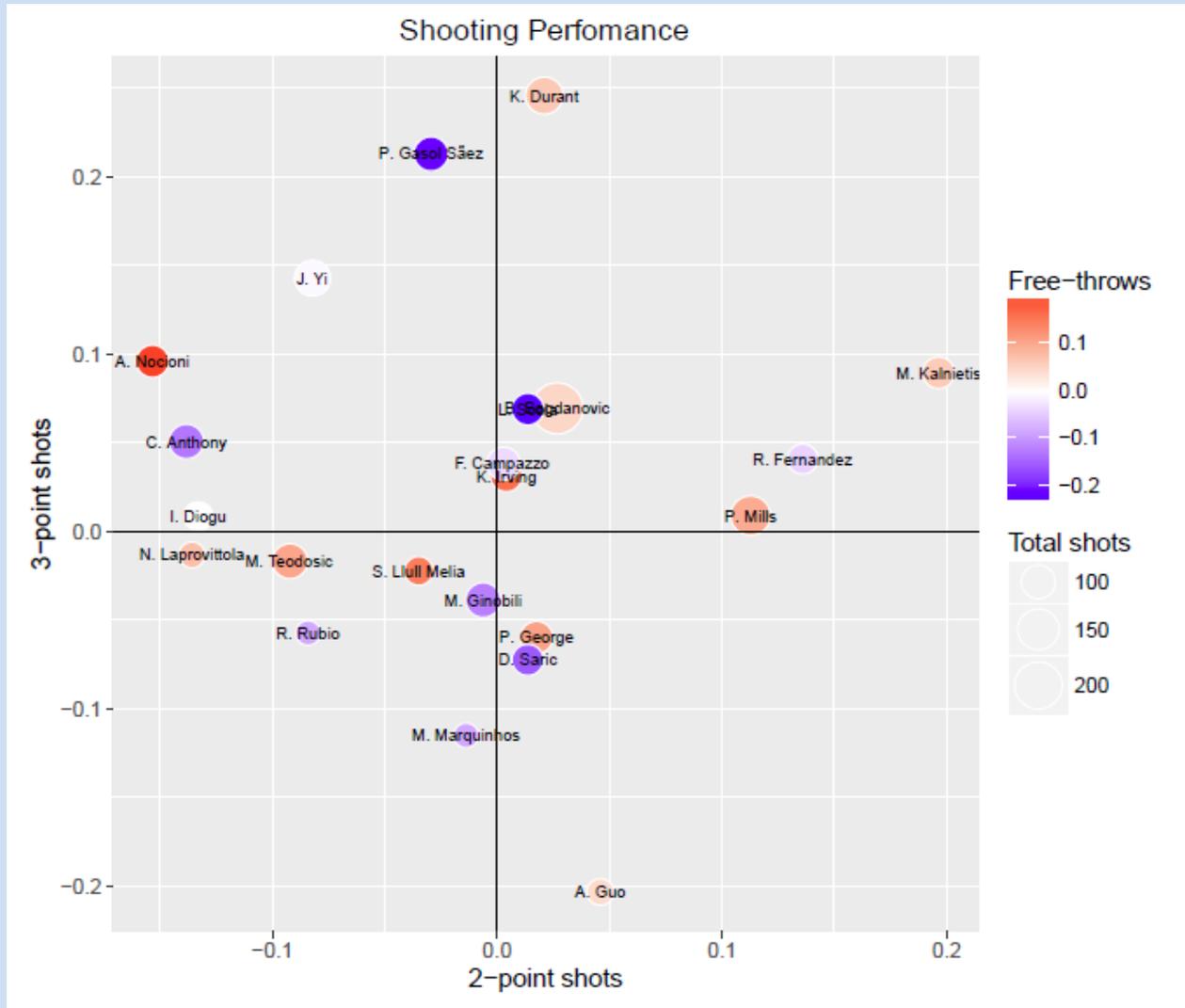
- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

## FURTHER RESEARCH:

according to psychological studies, some athletes view the competitive situations as challenging, and others perceive the same situations as stressful and anxiety-provoking. For this reason, it may be difficult to statistically detect stressful situations from large datasets including several players, as the overall average performance may remain unchanged as a response to some players improving their performance and some others getting worse.

→ Analysis of single players' reactions to stressful game situations (**propension to shot and variation in scoring probability**)

→ **Integration** with psychological studies



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- **CS3: performance variability and teamwork**
- CS4: sensor data analysis



# CS3: performance variability and teamwork

Markov Switching performance modelling for team network analysis in basketball *(in progress)*

Paola Zuccolotto, Marco Sandri, Marica Manisera and Rodolfo Metulini



**MOTIVATION:** Psychological studies have pointed out that typical performance is but one attribute of performance, but other aspects should be taken into account, in particular performance variability.

**AIM:** Assessment of players' shooting performance variability and investigation of its relationships with the team composition.

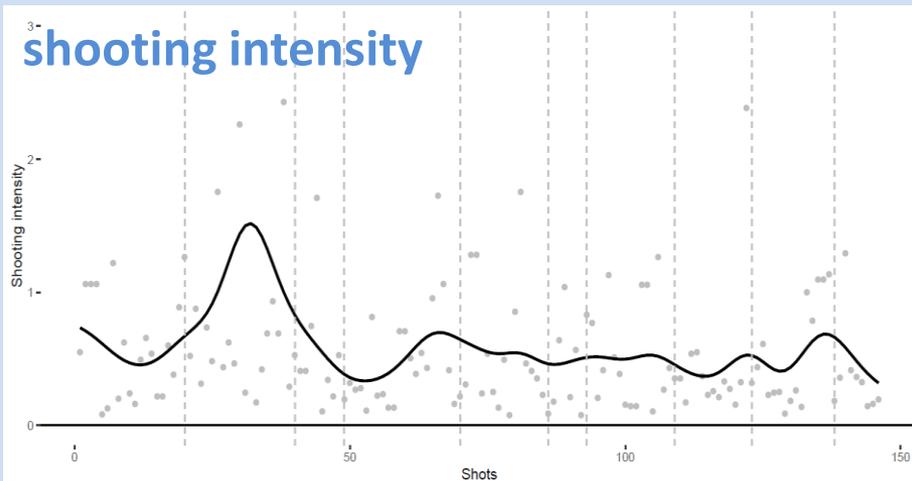


- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- **CS3: performance variability and teamwork**
- CS4: sensor data analysis

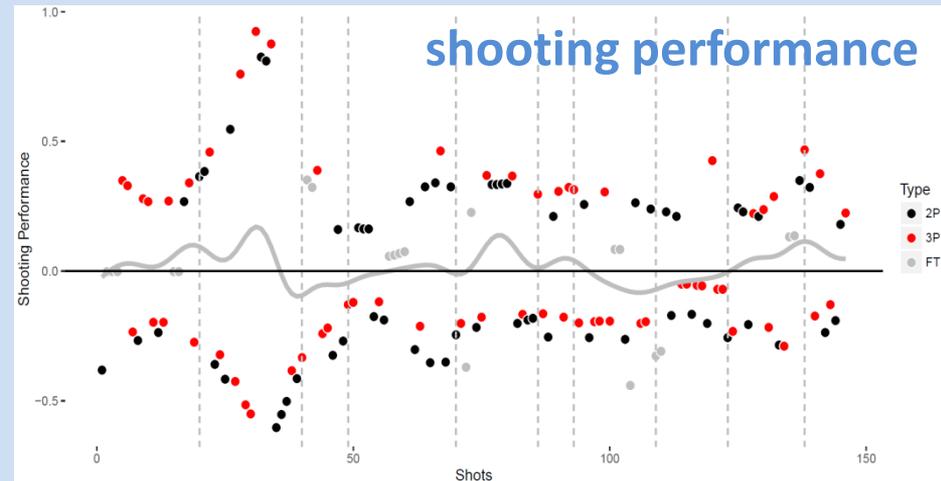
# PERFORMANCE VARIABILITY:

- Definition of a performance index based on the % of attempted shots that scored a basket and on the shooting intensity



$$\tilde{\phi}_{ij} = \frac{1}{t_{ij}}$$

$$\phi_{ij} = \frac{\tilde{\phi}_{ij}}{\phi^{(m_{ij})}}$$



$$E_{ij} = x_{ij} - p_{ij}$$

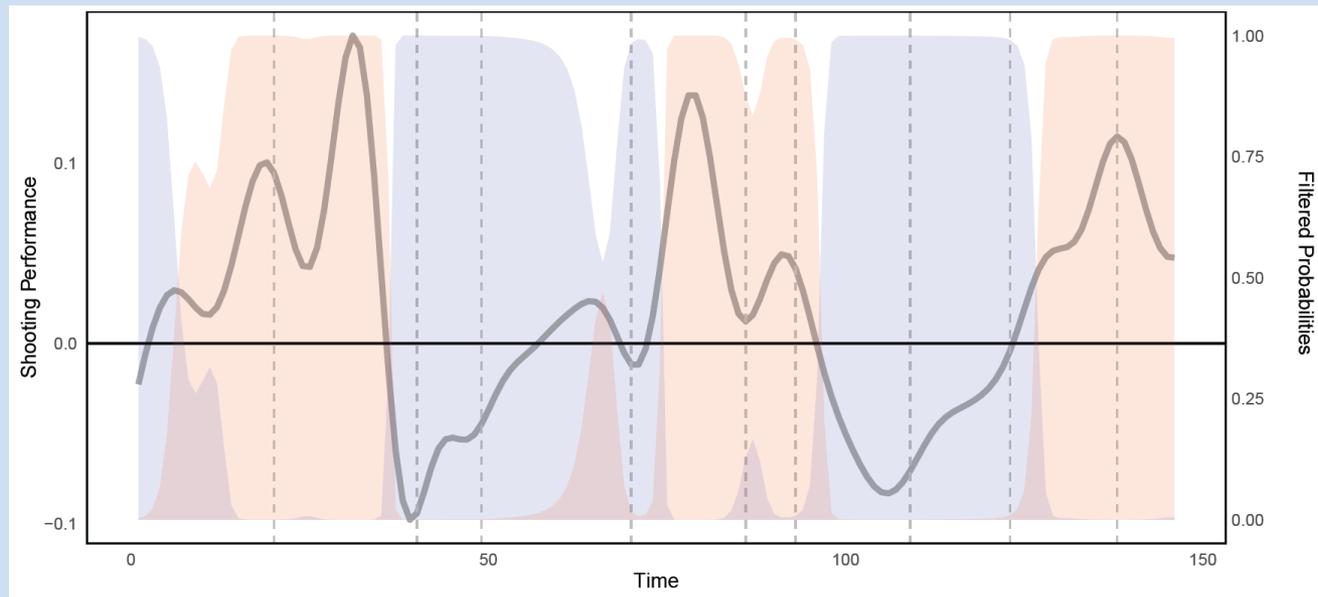
$$\psi_{ij} = \hat{\phi}_{ij} E_{ij}$$



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

# PERFORMANCE VARIABILITY:

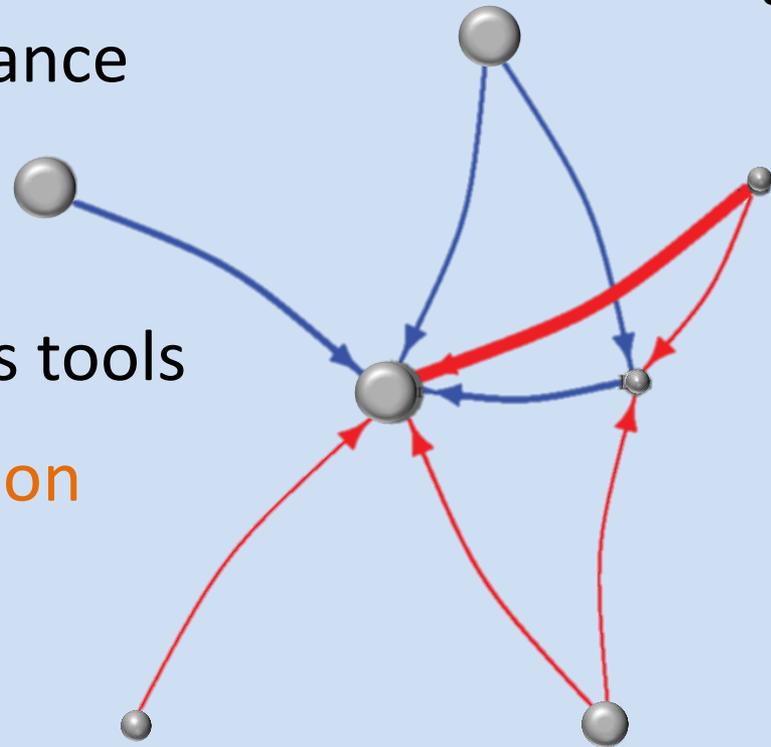
- Fit **Markov Switching models** to the shooting performance index, in order to detect the (significant) presence of periods of good and bad performance



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

## TEAMWORK ASSESSMENT:

- determine influence of each teammate on the regime of good and bad performance
- display the significant relationships by means of graphical network analysis tools
- predict the best substitution at a given time



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis



# CS4: sensor data analysis

Space-Time Analysis of Movements in Basketball



using Sensor Data

Marica Manisera <sup>\*1</sup>, Rodolfo Metulini <sup>†1</sup>, and Paola Zuccolotto <sup>‡1</sup>

<sup>1</sup>University of Brescia - Department of Economics and Management, Contrada Santa Chiara, 50, 25122, Brescia, Italy

MathSport International 2017



SIS2017



Sensor  
Data  
CS4



**Aim:** A first approach to sensor data analysis in basketball (visualization tools, cluster analysis, future challenges)

In collaboration with MYagonism



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- **CS4: sensor data analysis**



# CS4: sensor data analysis

Modelling the dynamic pattern of surface area in basketball and its effects on team performance



SIS 2018



COMPSTAT 2018

23rd International Conference on  
COMPUTATIONAL STATISTICS (COMPSTAT 2018)  
28-31 August 2018, Iasi, Romania



**AIM:** to study

- structural changes in the surface area
- associations between regimes and game variables
- the relation between the regime probabilities and the scored points

Sensor  
Data  
CS4



**BDsports**

Big DATA ANALYTICS IN SPORTS

MARICA MANISERA  
PAOLA ZUCCOLOTTO

– UNIVERSITY OF BRESCIA, ITALY

- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- **CS4: sensor data analysis**

# VISUALIZATION TOOLS

## Spatio-Temporal Movements in Team Sports: A Visualization approach using Motion Charts.

Rodolfo Metulini <sup>(1)</sup>

(1) Department of Economics and Management, University of Brescia, Contrada Santa Chiara, 50, 25122

Brescia BS, Italy,

[rodolfo.metulini@unibs.it](mailto:rodolfo.metulini@unibs.it).

A tool to display data recorded by tracking systems producing spatio-temporal traces of player trajectories with high definition and frequency

<https://www.youtube.com/watch?v=aejyrDnqYVY>



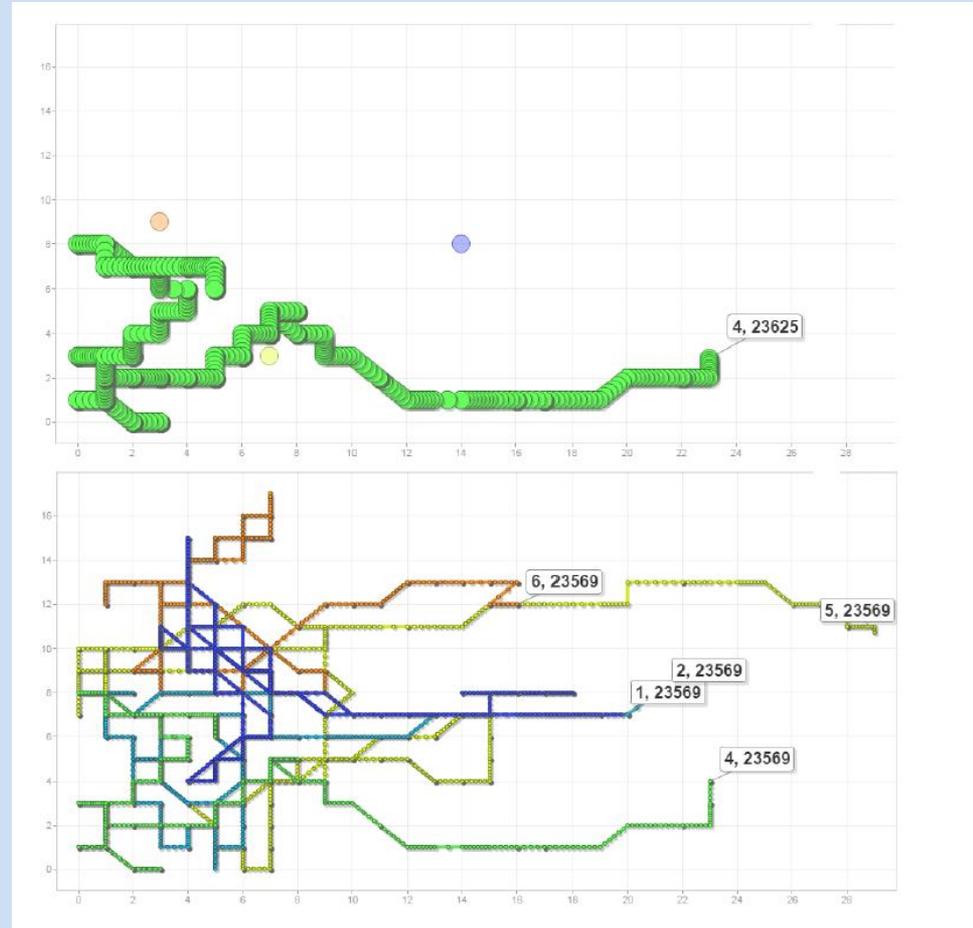
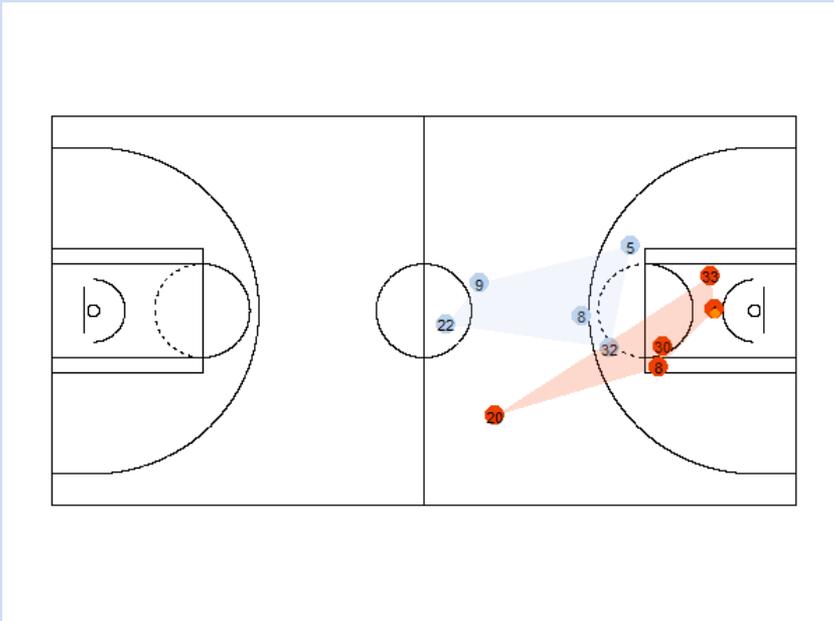
- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- **CS4: sensor data analysis**

# Visualization Tools

**James P. Curley**

Curley Social Neurobiology Lab website  
(Psychology Department and Center for Integrative Animal Behavior, Columbia University, New York City)



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

# CONVEX Hulls ANALYSIS

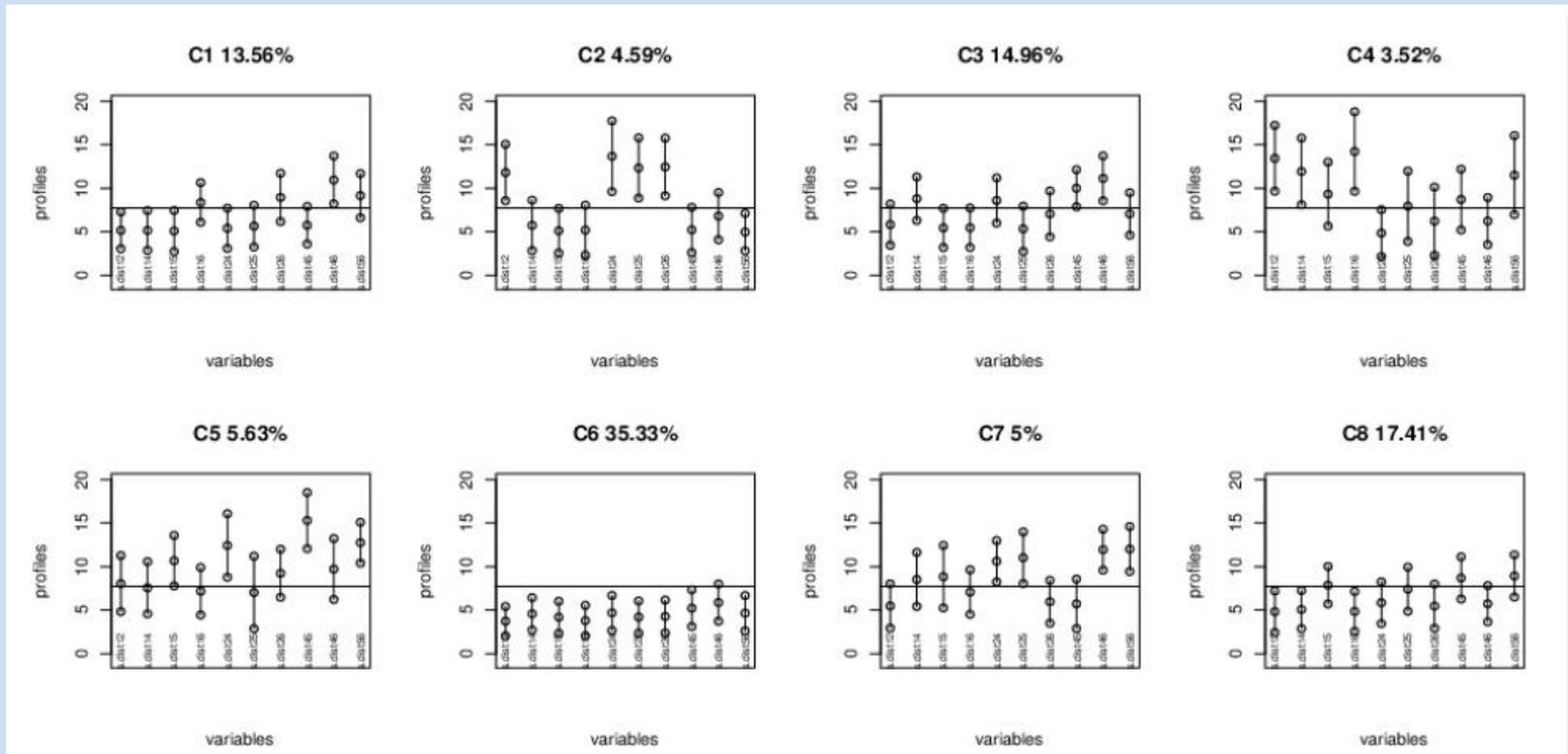
	Average distance		Convex hull area	
	attack	defence	attack	defence
Min	5.418	2.709	11.000	4.500
1st Qu.	7.689	3.942	32.000	12.500
Median	8.745	4.696	56.000	18.500
Mean	8.426	5.548	52.460	32.660
3rd Qu.	9.455	5.611	68.500	27.500
Max	10.260	11.640	99.500	133.500



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- **CS4: sensor data analysis**

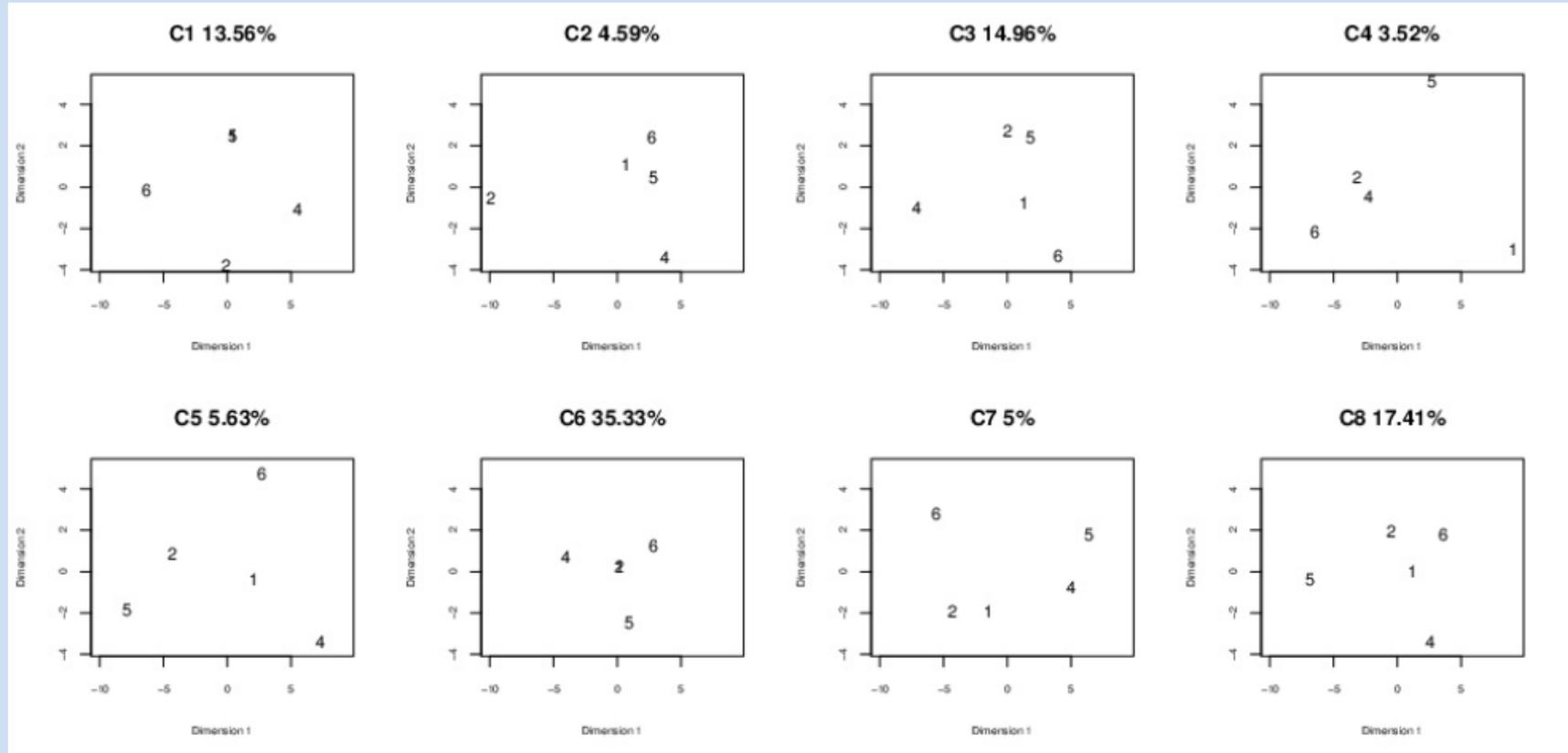
# CLUSTER ANALYSIS



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- **CS4: sensor data analysis**

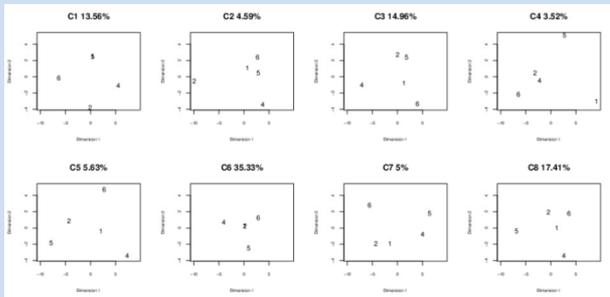
# CLUSTER ANALYSIS + MULTIDIMENSIONAL SCALING



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- **CS4: sensor data analysis**

# CLUSTER ANALYSIS + MULTIDIMENSIONAL SCALING



D	A
14.12	85.88
14.09	85.91
26.07	73.93
15.38	84.62
57.98	42.02
85.07	14.93
39.53	60.47
39.08	60.92

NA	1	2	3	4	5	6	7	8
1	0	10.71	23.53	47.83	0	20.83	31.25	20.23
2	0.77	0	9.15	0	1.85	2.08	8.33	2.89
3	31.54	42.86	0	8.7	44.44	20.83	18.75	20.23
4	6.15	3.57	1.96	0	7.41	0	10.42	1.16
5	0.77	3.57	16.99	17.39	0	0	16.67	8.09
6	27.69	7.14	18.95	0	1.85	0	0	43.93
7	15.38	21.43	3.92	4.35	18.52	0	0	3.47
8	17.69	10.71	25.49	21.74	25.93	56.25	14.58	0



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- **CS4: sensor data analysis**

## FUTURE CHALLENGES:

- **Integration** with play-by-play data
- **Integration** with video and match analysis
- **Integration** with body metrics (body physiology tracking via “smart clothing” and/or body measurements)
- **Integration** with qualitative assessments
- Network analysis tools
- Spatio-temporal statistical models
- Addition of the other team’s data
- Addition of the ball’s position



- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

# Concluding ...

## TRUE:

- If people keep thinking that Statistics is merely PPG, AST, REB, ...
- If people don't learn how Stats have to be interpreted (*"Do not put your faith in what statistics say until you have carefully considered what they do not say."* W. W. Watt)

MARC GASOL SAYS: 'STATS ARE KILLING THE GAME OF BASKETBALL'



Stats have always been important to players, coaches, the media, and fans; this year in particular, we've been closely watching Russell Westbrook as he made triple-double history. Memphis Grizzlies center Marc Gasol made history as well, becoming the first center to record 300 assists, 100 threes and 100 blocks in a season, but he doesn't want to discuss stats, in fact, he says they're killing the game.

Gasol was asked about point guard Mike Conley's breakout season statistically and initially responded with this take:

“We've got 43 wins. If we win (tonight), we'll have 44. That's the only number you guys (media) should care about,” Gasol said. “Stats are great, but wins and losses matter. Stats are killing the game of basketball. Basketball is a subjective game. A lot of things happen that you cannot measure in stats. Different things matter. To me, the most important things in basketball are not measured by stats.”

## FALSE:

- If **modern approaches** to basketball analytics are used
- If we are able to **integrate** analytics and technical experience
- If we are able to spread the **culture of Statistics**



# REFERENCES

Download a (regularly updated) list of references at  
<http://bodai.unibs.it/bdsports/basketball.htm>

# THANK YOU

