Data mining, big data visualization, data integration:
modern approaches to basketball analytics

Marica Manisera – Paola Zuccolotto
University of Brescia, Italy

200 YEARS GHENT UNIVERSITY

Ghent, 24th April 2017

BODaI – Lab
Big&Open Data Innovation Laboratory

UNIVERSITY OF BRESCIA - HEALTH & WEALTH
UNIVERSITAS STUDIORUM BRIXIAE

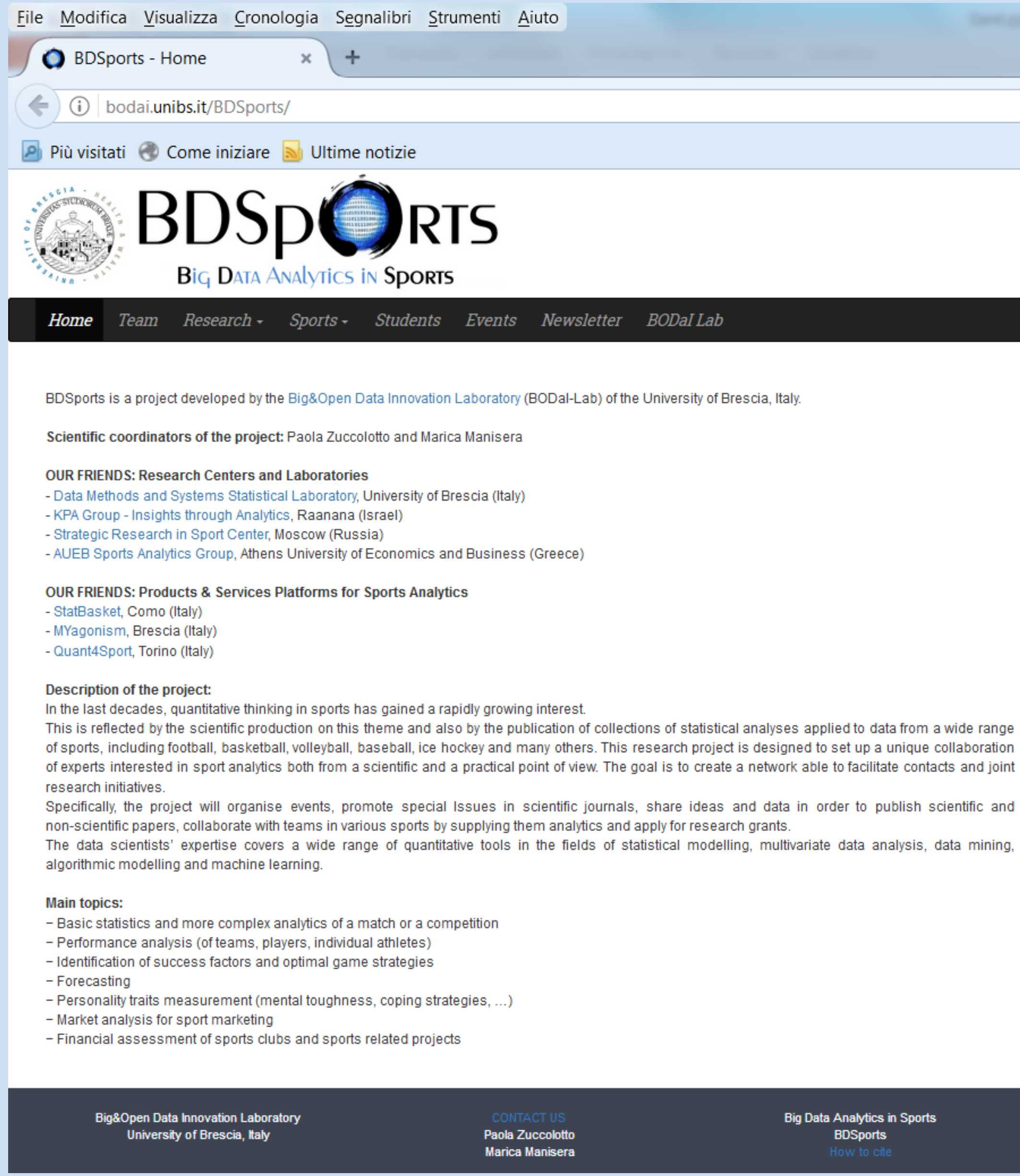
BDSports
Big Data Analytics in Sports

marica.manisera@unibs.it

paola.zuccolotto@unibs.it

BDSports, a network of people interested in Sports Analytics

http://bodai.unibs.it/BDSports/


File Modifica Visualizza Cronologia Segnalibri Strumenti Aiuto

BDSports - Home

bodai.unibs.it/BDSports/

Più visitati Come iniziare Ultime notizie

BDSPORTS
Big Data Analytics in Sports

Home Team Research Sports Students Events Newsletter BODal Lab

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)

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
– Personality traits measurement (mental toughness, coping strategies, ...)
– Market analysis for sport marketing
– Financial assessment of sports clubs and sports related projects

Big&Open Data Innovation Laboratory
University of Brescia, Italy

CONTACT US
Paola Zuccolotto
Marica Manisera

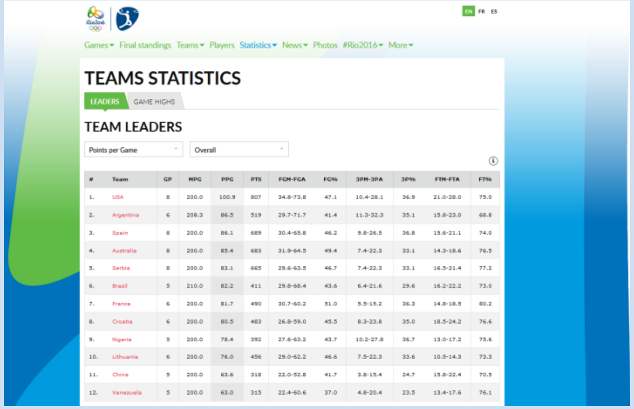
Big Data Analytics in Sports
BDSports
How to cite

Agenda:

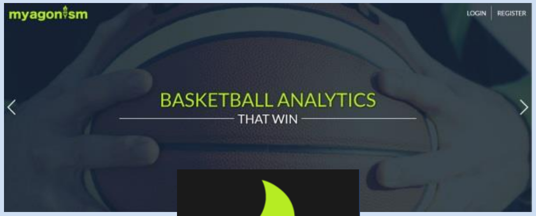
- Basketball analytics: state of the art
- Basketball datasets
- Case studies:
  - CS1: new positions in basketball
  - CS2: scoring probability when shooting under high-pressure conditions
  - CS3: performance variability and teamwork assessment
  - CS4: sensor data analysis
- Concluding remarks

Basketball Analytics

Official Statistics



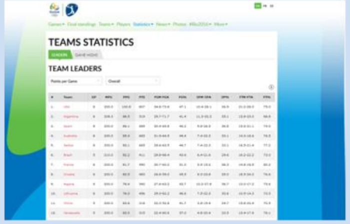
Sport Analytics Services



Scientific Research

Basketball Analytics

Official Statistics



Sport Analytics Services

Scientific Research

Basketball Analytics

Official Statistics

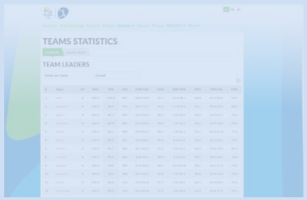
Sport Analytics Services

Scientific Research

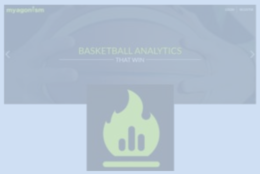
Our analyses often **integrate** machine learning tools and experts' suggestions

Basketball Analytics

Official Statistics

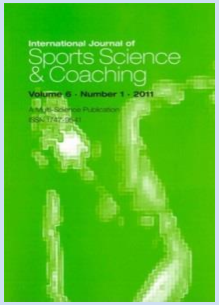


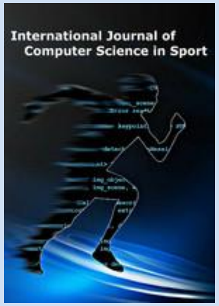
Sport Analytics Services



Scientific Research

Scientific Journals





Special Issues

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).

Data → **Big Data**

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

Splits

TOTAL SPLITS

GAME STATISTICS

PLAYER	GP	GS	MIN	PPG	OFFR	DEFR	RPG	APG	SPG	BPG	TPG	FPG	A/TO	PER
Kevin Durant, SF	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
Stephen Curry, PG	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
Klay Thompson, SG	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
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	17.2
Ian Clark, SG	51	0	14.3	6.5	0.2	1.2	1.4	1.3	0.55	0.12	0.6	1.0	1.9	13.5
Andre Iguodala, SF	54	0	25.6	6.4	0.7	3.2	3.9	3.4	0.93	0.39	0.7	1.2	4.9	13.0
JaVale McGee, C	51	10	9.5	6.2	1.1	2.0	3.0	0.2	0.24	0.69	0.6	1.5	0.4	23.7
Zaza Pachulia, C	44	44	18.8	5.9	2.1	3.9	6.0	2.0	0.95	0.41	1.3	2.4	1.6	15.9
Shaun Livingston, PG	51	1	17.3	5.2	0.4	1.6	1.9	1.7	0.47	0.22	0.8	1.6	2.1	10.1
David West, PF	42	0	11.5	4.1	0.6	2.1	2.7	2.1	0.67	0.45	1.0	1.4	2.1	16.6
Patrick McCaw, PG	46	3	12.4	3.3	0.2	0.8	1.0	1.0	0.35	0.22	0.5	0.8	2.0	7.9
Kevon Looney, SF	43	3	9.3	2.9	0.8	1.7	2.6	0.6	0.30	0.37	0.3	1.4	1.9	14.5
James Michael McAdoo, SF	29	0	9.1	2.8	0.4	1.1	1.5	0.4	0.24	0.52	0.4	0.9	1.0	11.6
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	2.5	5.0
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	0.0	6.0
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
Totals	**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

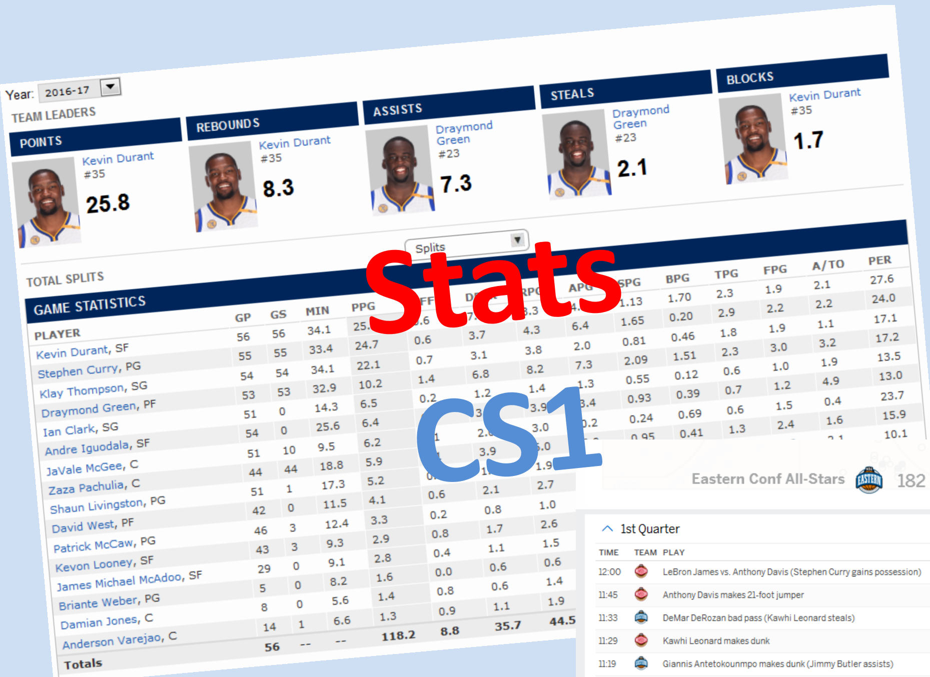
stats.nba.com

www.fiba.com

Leagues

...

Data → **Big Data**

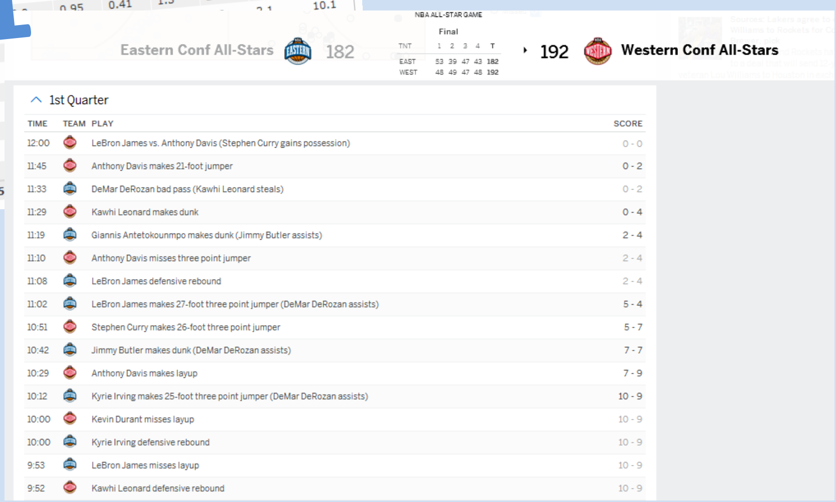


Stats

CS1

play-by-play

CS2 – CS3

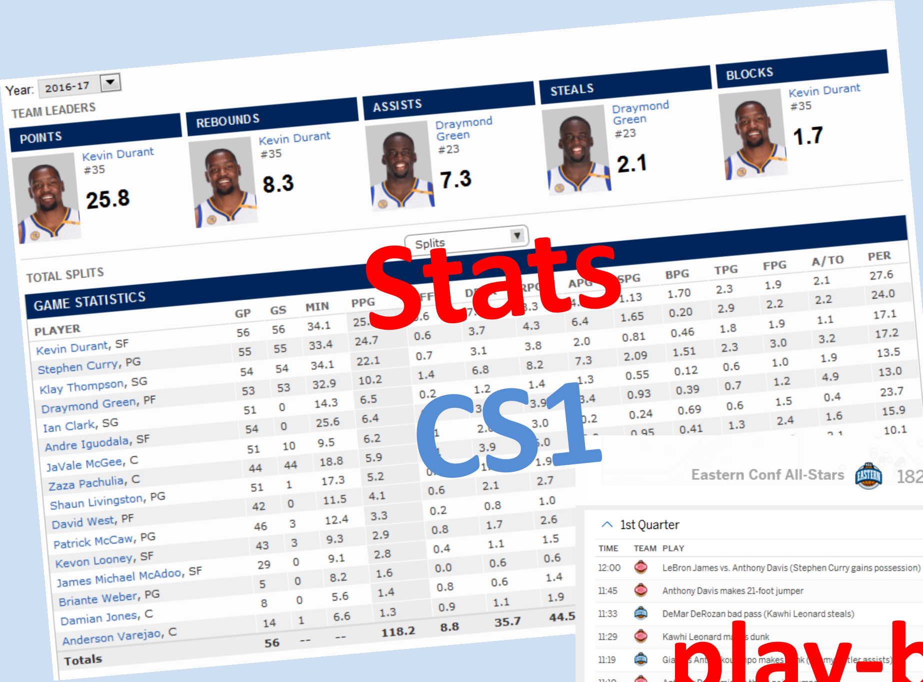

Eastern Conf All-Stars 182 — 192 Western Conf All-Stars

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

- 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

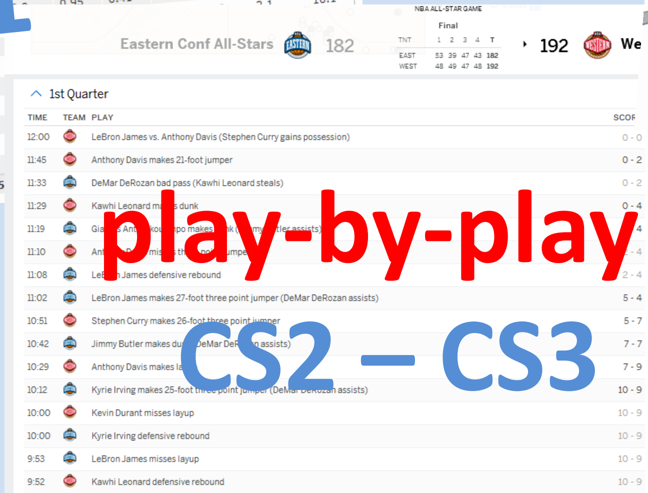
Data → **Big Data**



PLAYER	GP	GS	MIN	PPG	OFF	DEF	RPG	APG	SPG	BPG	TPG	FPG	A/TO	PER
Kevin Durant, SF	56	56	34.1	25.[illegible]	[illegible].6	7.[illegible]	[illegible].3	4.[illegible]	1.13	1.70	2.3	1.9	2.1	27.6
Stephen Curry, PG	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
Klay Thompson, SG	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
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	17.2
Ian Clark, SG	51	0	14.3	6.5	0.2	1.2	1.4	[illegible].3	0.55	0.12	0.6	1.0	1.9	13.5
Andre Iguodala, SF	54	0	25.6	6.4	[illegible]	3[illegible]	3.9	[illegible].4	0.93	0.39	0.7	1.2	4.9	13.0
JaVale McGee, C	51	10	9.5	6.2	[illegible]1	2.[illegible]	3.0	[illegible].2	0.24	0.69	0.6	1.5	0.4	23.7
Zaza Pachulia, C	44	44	18.8	5.9	[illegible]	3.9	[illegible].0	[illegible]	0.95	0.41	1.3	2.4	1.6	15.9
Shaun Livingston, PG	51	1	17.3	5.2	0.[illegible]	1.[illegible]	1.9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.1	10.1
David West, PF	42	0	11.5	4.1	0.6	2.1	2.7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Patrick McCaw, PG	46	3	12.4	3.3	0.2	0.8	1.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Kevon Looney, SF	43	3	9.3	2.9	0.8	1.7	2.6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
James Michael McAdoo, SF	29	0	9.1	2.8	0.4	1.1	1.5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Briante Weber, PG	5	0	8.2	1.6	0.0	0.6	0.6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Damian Jones, C	8	0	5.6	1.4	0.8	0.6	1.4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Anderson Varejao, C	14	1	6.6	1.3	0.9	1.1	1.9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Totals	56	--	--	118.2	8.8	35.7	44.5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Stats
CS1

Sensor Data
CS4



TNT	1	2	3	4	T
EAST	53	39	47	43	182
WEST	48	49	47	48	192

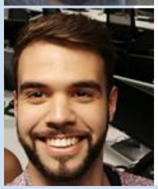
192 We

1st Quarter

TIME	TEAM	PLAY	SCOR
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 ma[illegible]s dunk	0 - 4
11:19		Gia[illegible]s Ant[illegible]kou[illegible]po makes [illegible]k ([illegible]my[illegible]tler assists)	[illegible]4
11:10		Ant[illegible] D[illegible] mis[illegible]s th[illegible] poi[illegible] jumpe[illegible]	[illegible] - 4
11:08		Le[illegible]n 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 du[illegible] DeMar DeR[illegible]n assists)	7 - 7
10:29		Anthony Davis makes la[illegible]	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

CS1: new positions in basketball

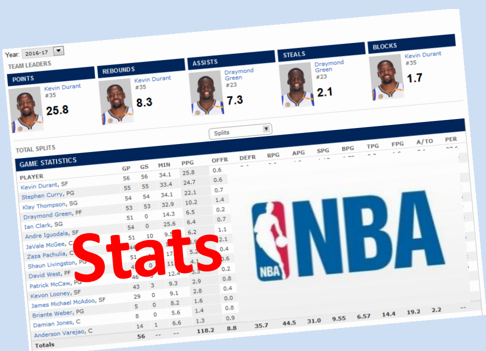



Role revolution: towards a new meaning of positions in basketball

Federico Bianchi[a], Tullio Facchinetti *[a], and Paola Zuccolotto[b]

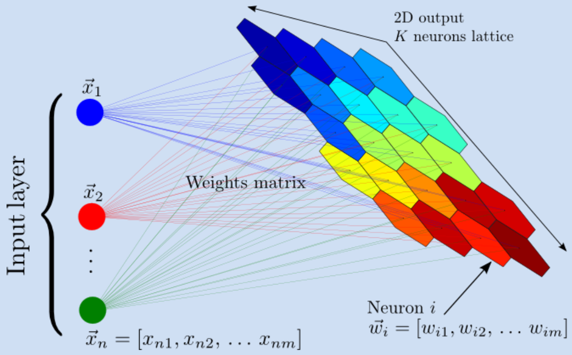
[a] *University of Pavia, via Ferrata, 1, 27100 Pavia, Italy*
[b] *University of Brescia, c.da S. Chiara 50, 25122 Brescia, Italy*

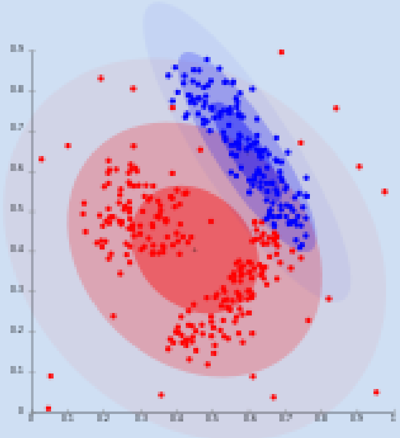
(*submitted*)



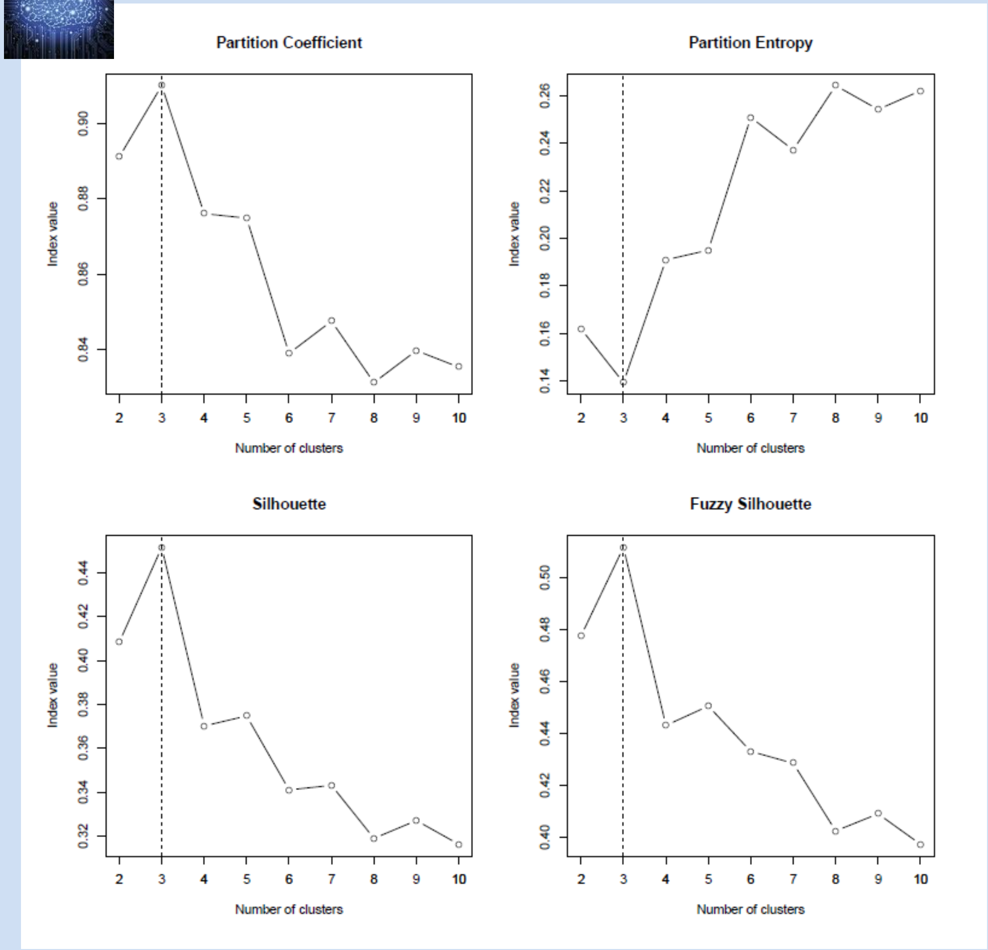
MOTIVATION: The existing positions - often defined a long time ago - tend to reflect traditional points of view about the game and sometimes they are no longer well-suited to the new concepts arisen with the evolution of the way of playing.

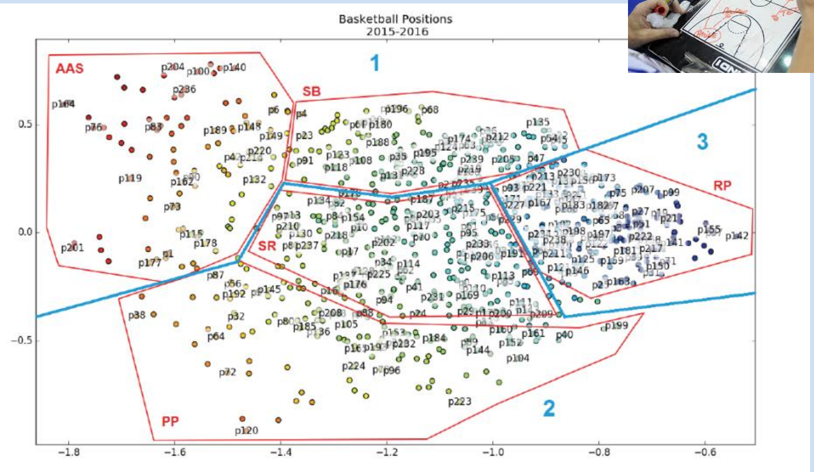
AIM: describing new roles of players during the game, by means of the analysis of players' performance statistics with data mining and machine learning tools.





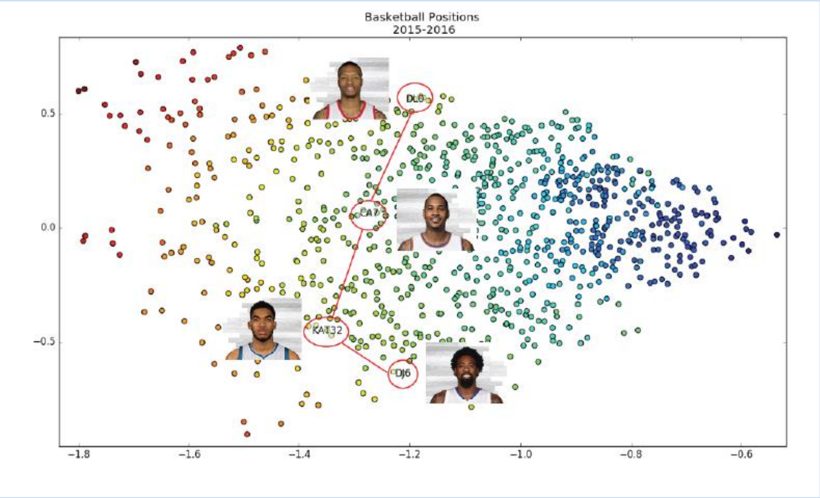
- «Key-players» training set → 7-dimensional **SOM**
- clusterization of the SOM output layer into a proper number of groups by means of a **fuzzy clustering algorithm**

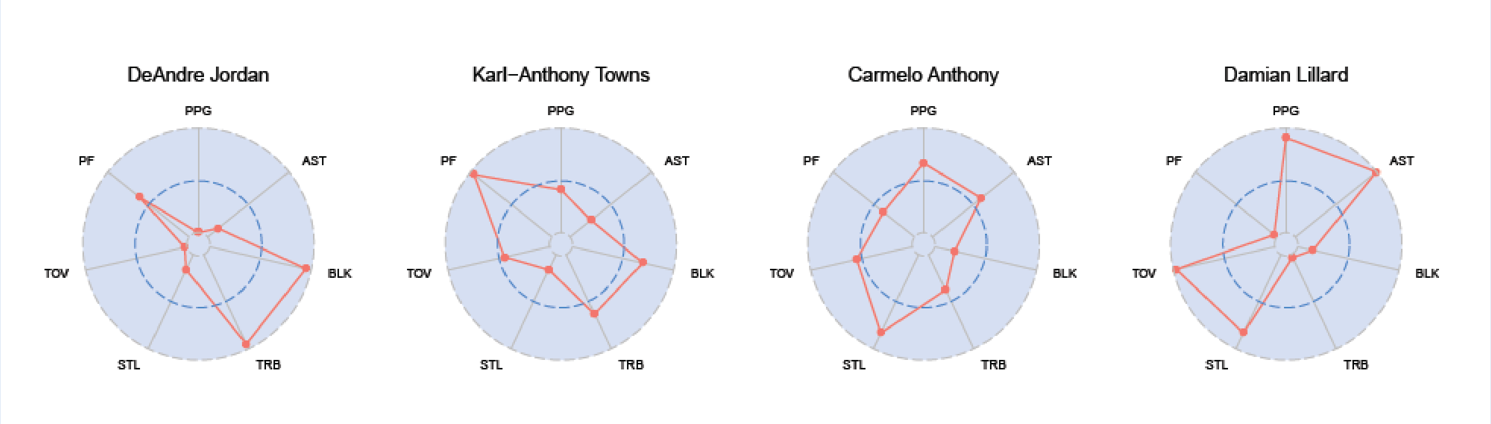




Position	Short	Players
All-Around All Stars	AAS	LeBron James (LBJ23), Kevind Durant (KD35), Paul George (PG13), Stephen Curry (SC30), James Harden (JH13)
Scoring Backcourt	SB	Kobe Bryant (KB24), Damian Lillard (DL0), Kyrie Irving (KI2), Dwyane Wade (DW3)
Scoring Rebounder	SR	Marc Gasol (MG33), Carmelo 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





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

(*submitted*)

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.

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)

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

Data Mining Tools:

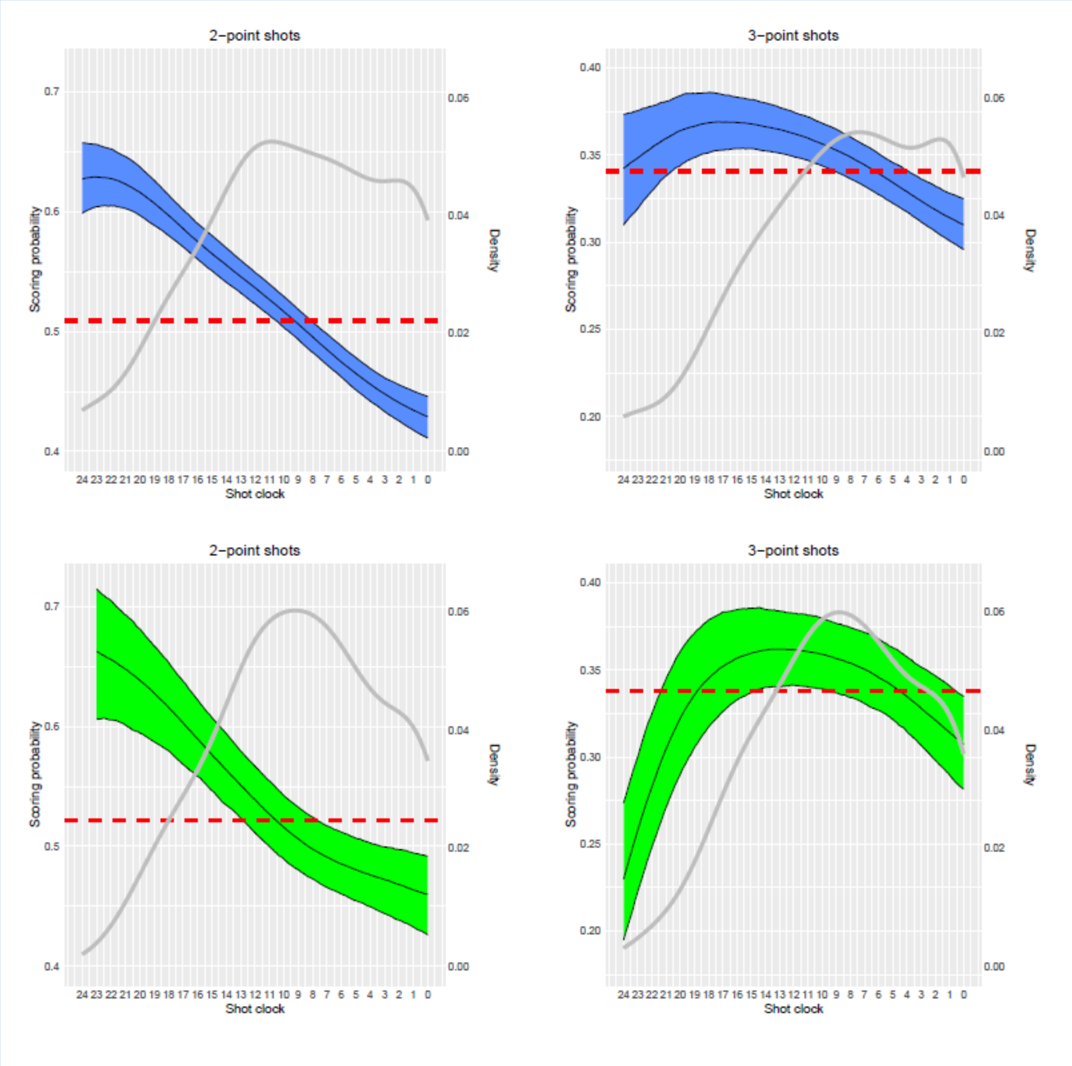
- **univariate** non parametric regressions via kernel smoothing on the dependent variable Made (assuming values 1 and 0 according to whether, respectively, the attempted shot scored a basket or not)
- 1000 bootstrap samples of size *nboot* = 5000 and *nboot* = 1000 for the dataset A2Ita and Rio16, respectively.

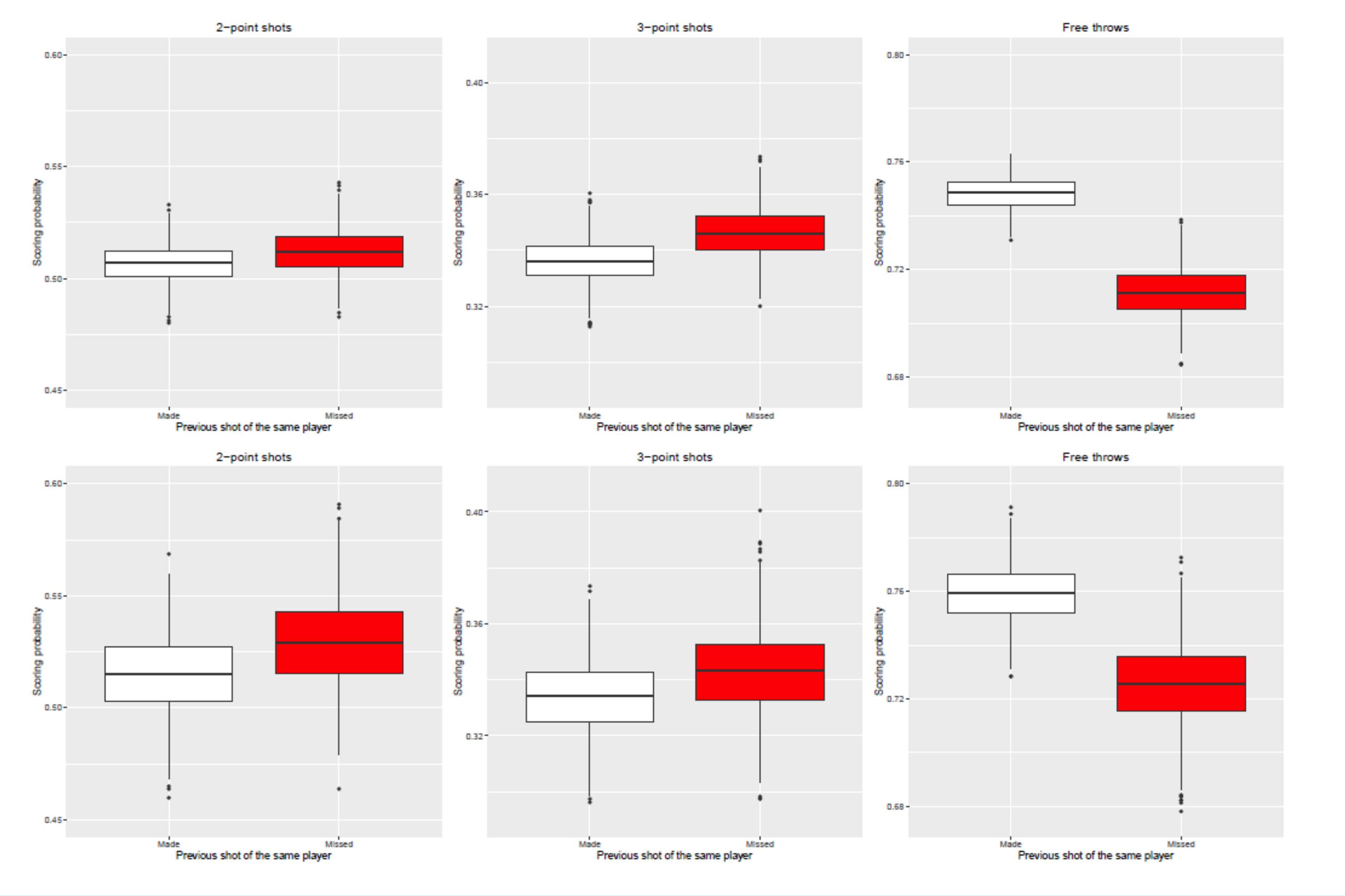
→ **few univariate relationships detected** - Just Shot.Clok and Miss.Pl

- 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

LEGA NAZIONALE PALLACANESTRO

Rio2016

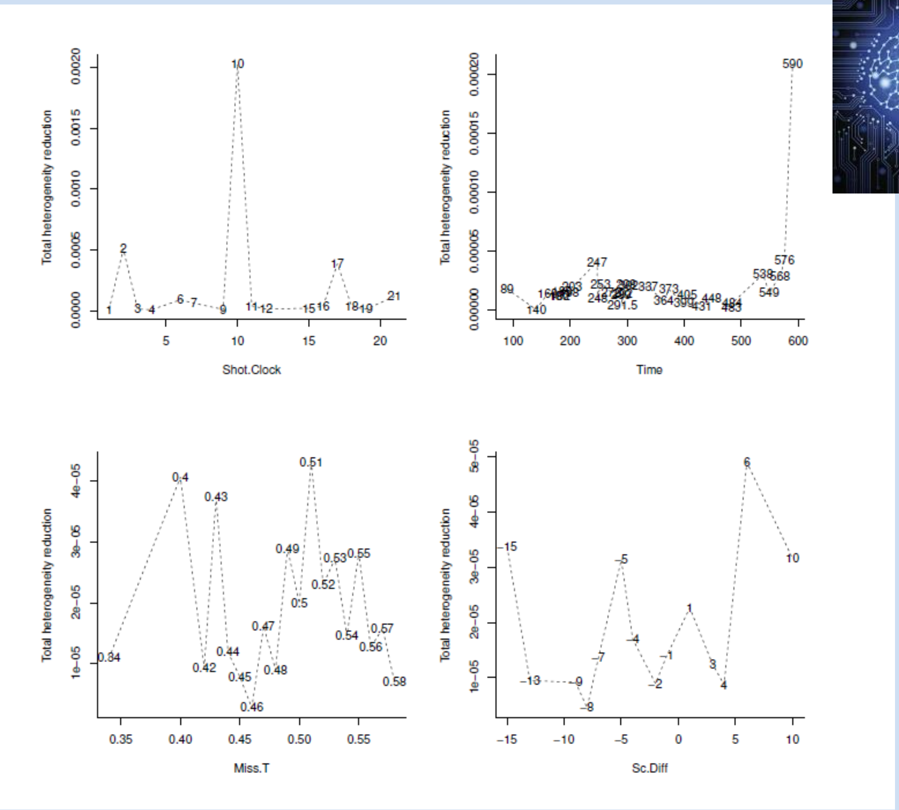




2–point shots
3–point shots
Free throws
Scoring probability
Made
Missed
Previous shot of the same player
0.60
0.55
0.50
0.45
0.40
0.36
0.32
0.80
0.76
0.72
0.68

Data Mining Tools:

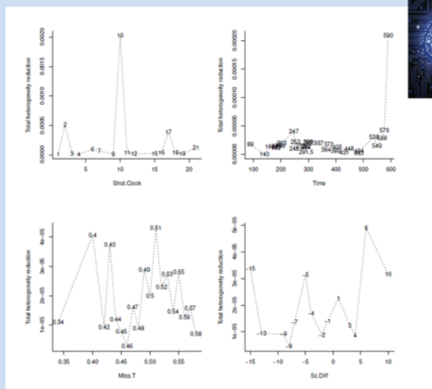
- **CART** (Classification And Regression Trees), algorithm able to deal with multivariate complex relationships, also detecting interactions among predictors
- we transform numerical into categorical covariates in order to improve interpretability → combination of the results of a **machine learning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



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)

Marica Manisera
Paola Zuccolotto
– University of Brescia, Italy

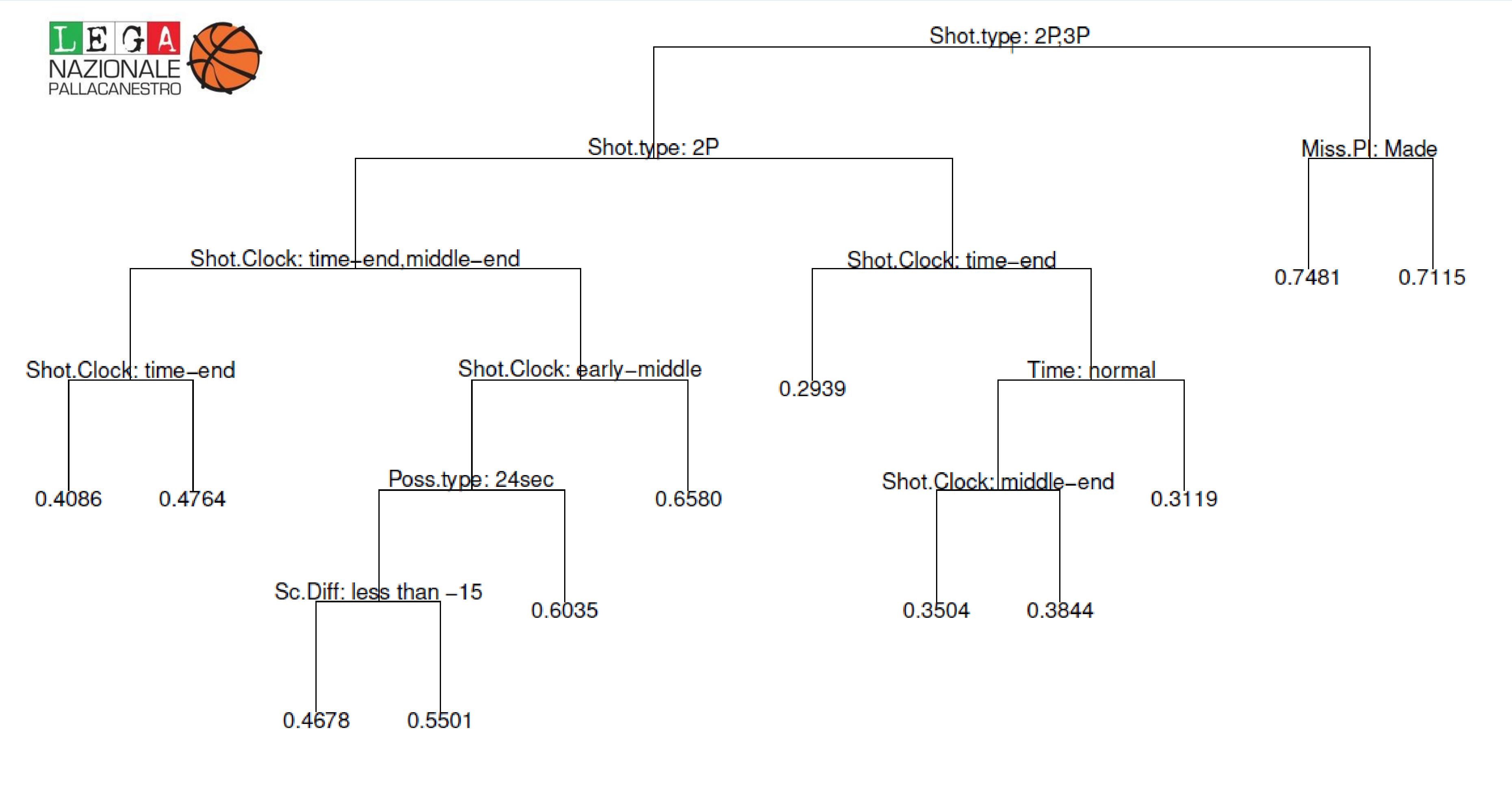


focus on:

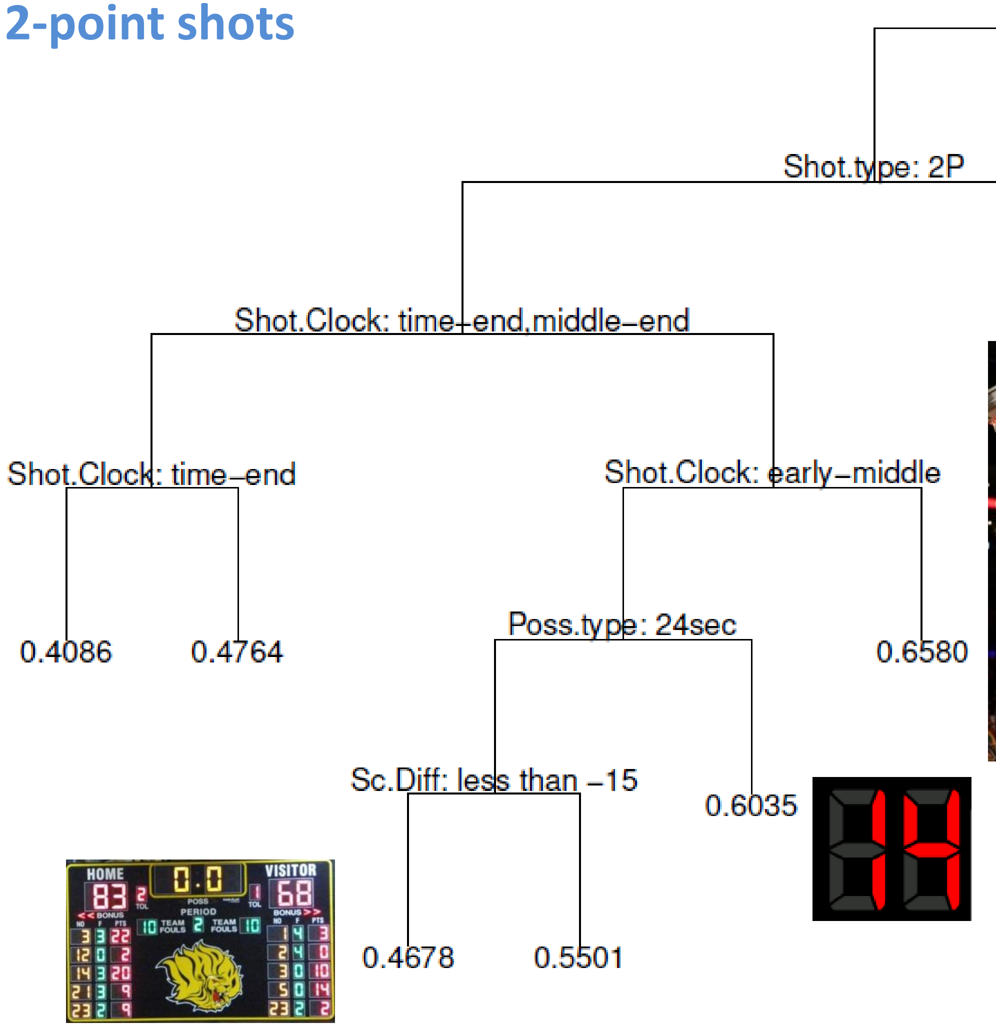
(1) the last 1-2 seconds of possession, very close to the shot clock buzzer sounding,
(2) games where the score difference is low, for example, between -4 and 4,
(3) the last 1-2 minutes of each quarter (especially the final quarter)

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

Very similar results with Rio 2016 data

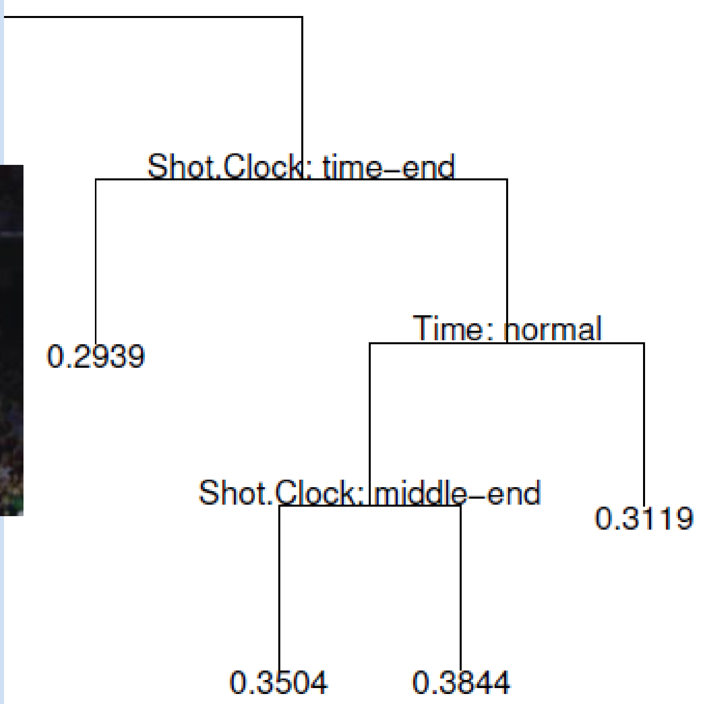


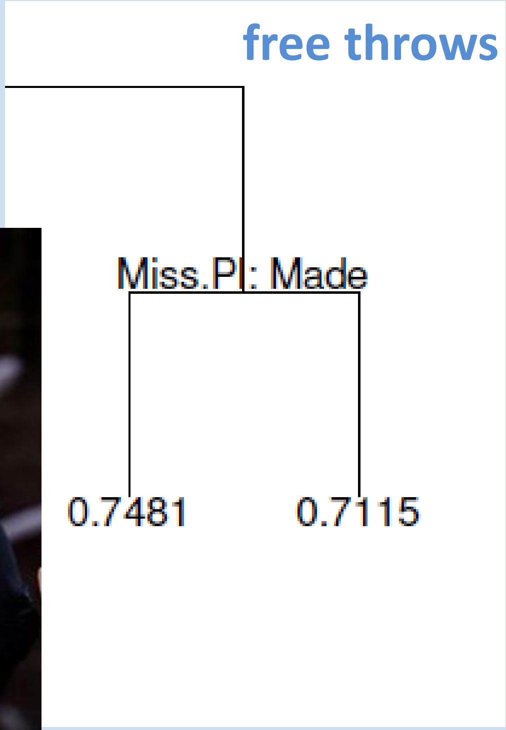
2-point shots



Marica Manisera
Paola Zuccolotto
– University of Brescia, Italy

3-point shots





free throws
Miss.Pl: Made
0.7481
0.7115

New Shooting Performance Measure:

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

Performance of Player i

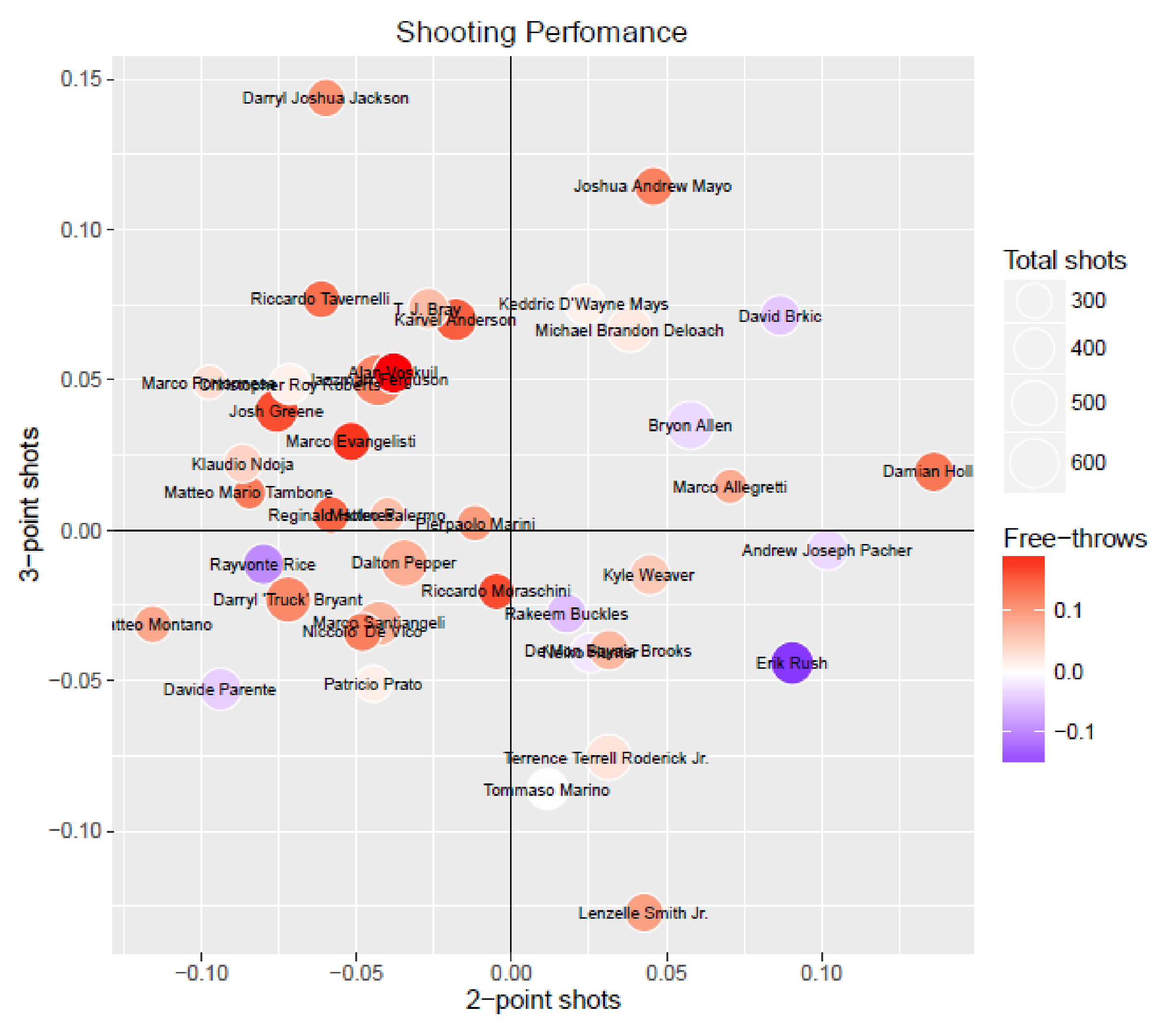
$$P_i(T) = av_{j \in J_T}\left(x_{ij} - \pi_{ij}\right)$$

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

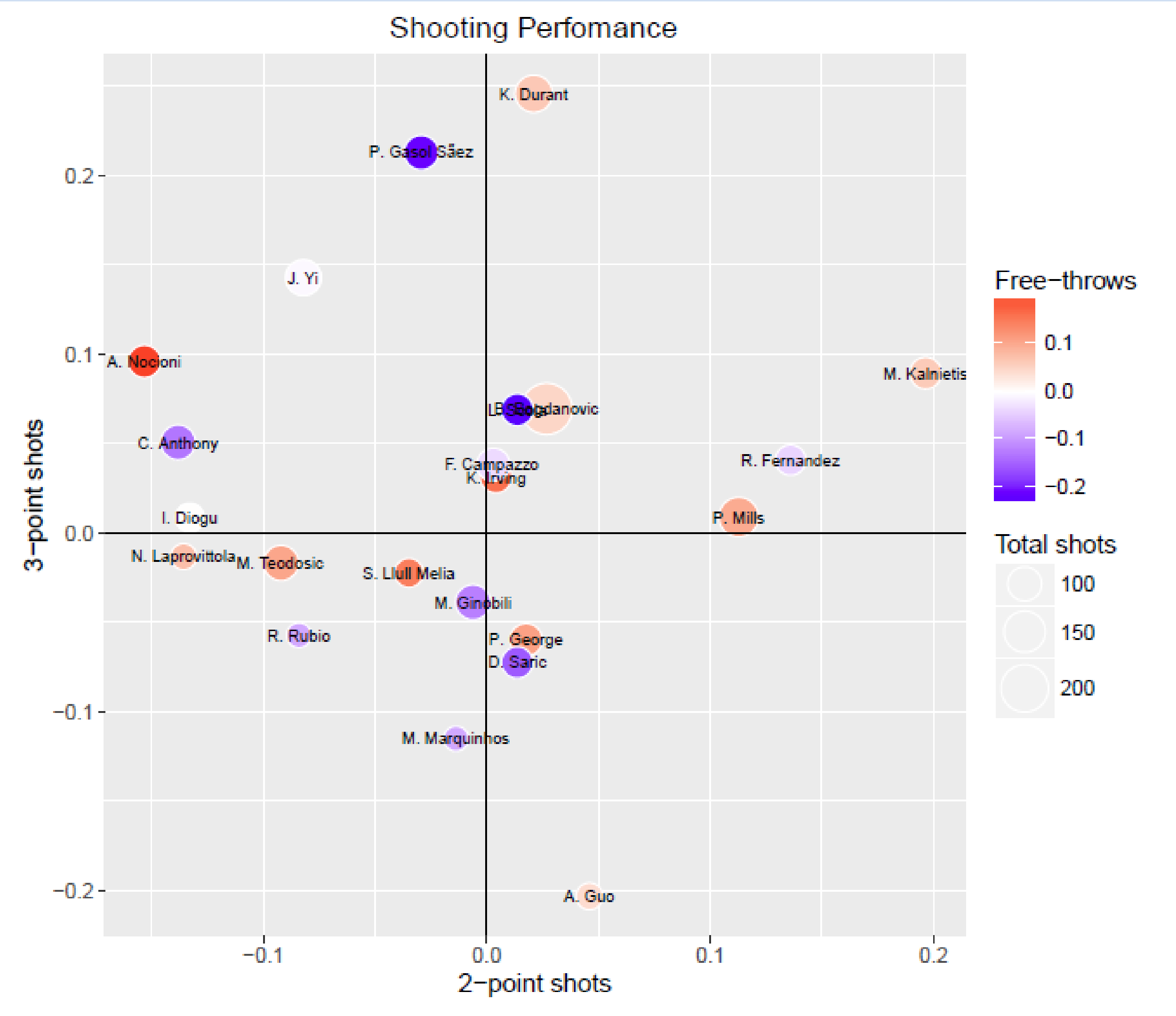
j-th shot made (1) or missed (0)

scoring probability of j-th shot according to CART

LEGA NAZIONALE PALLACANESTRO



Rio2016



Shooting Perfomance
3−point shots
2−point shots
0.2
0.1
0.0
−0.1
−0.2
−0.1
0.0
0.1
0.2
K. Durant
P. Gasol Sáez
J. Yi
A. Nocioni
M. Kalnietis
Bogdanovic
C. Anthony
F. Campazzo
K. Irving
R. Fernandez
I. Diogu
P. Mills
N. Laprovittola
M. Teodosic
S. Llull Melia
M. Ginobili
R. Rubio
P. George
D. Saric
M. Marquinhos
A. Guo
Free−throws
0.1
0.0
−0.1
−0.2
Total shots
100
150
200

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

CS3: performance variability and teamwork

Markov Switching performance modelling for team network analysis in basketball

Paola Zuccolotto, Marica Manisera and Marco Sandri

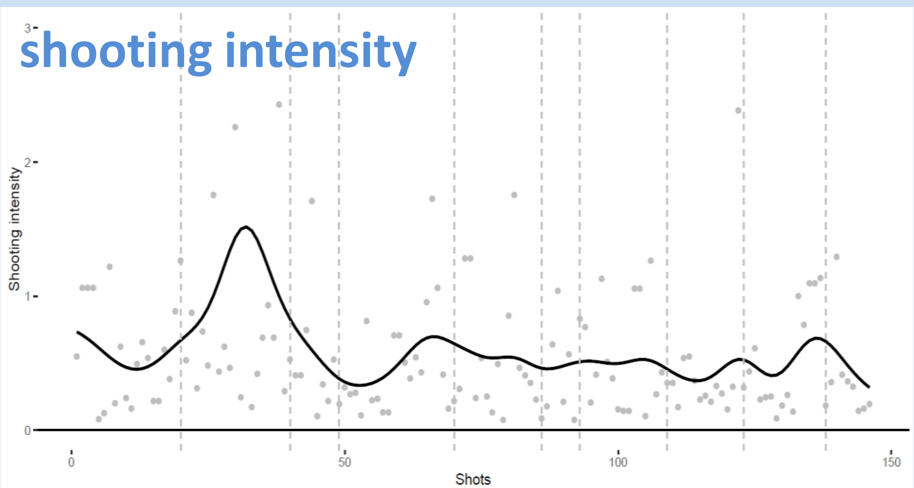
(*in progress*)

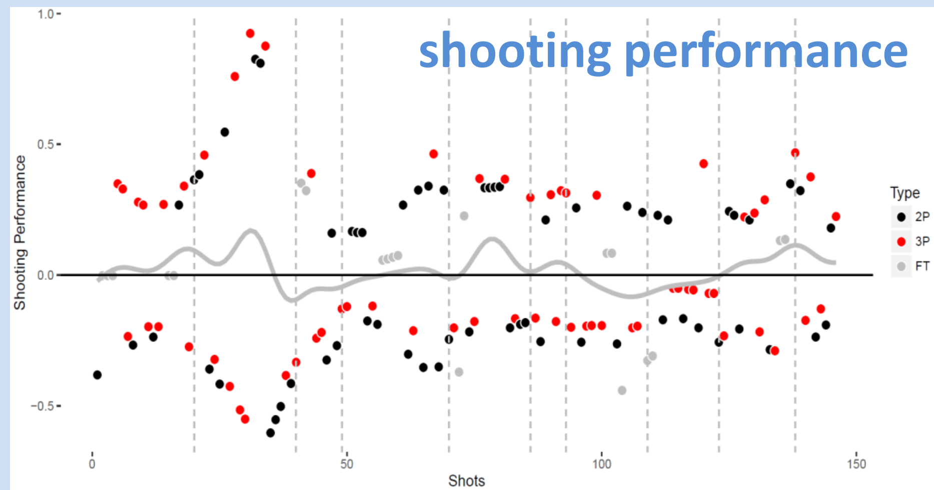
MOTIVATION: Psychological studies have pointed out that typical performance is but one attribute of performance, but other aspects should be taken into account, in particular performance variability.

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

Performance Variability:

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





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

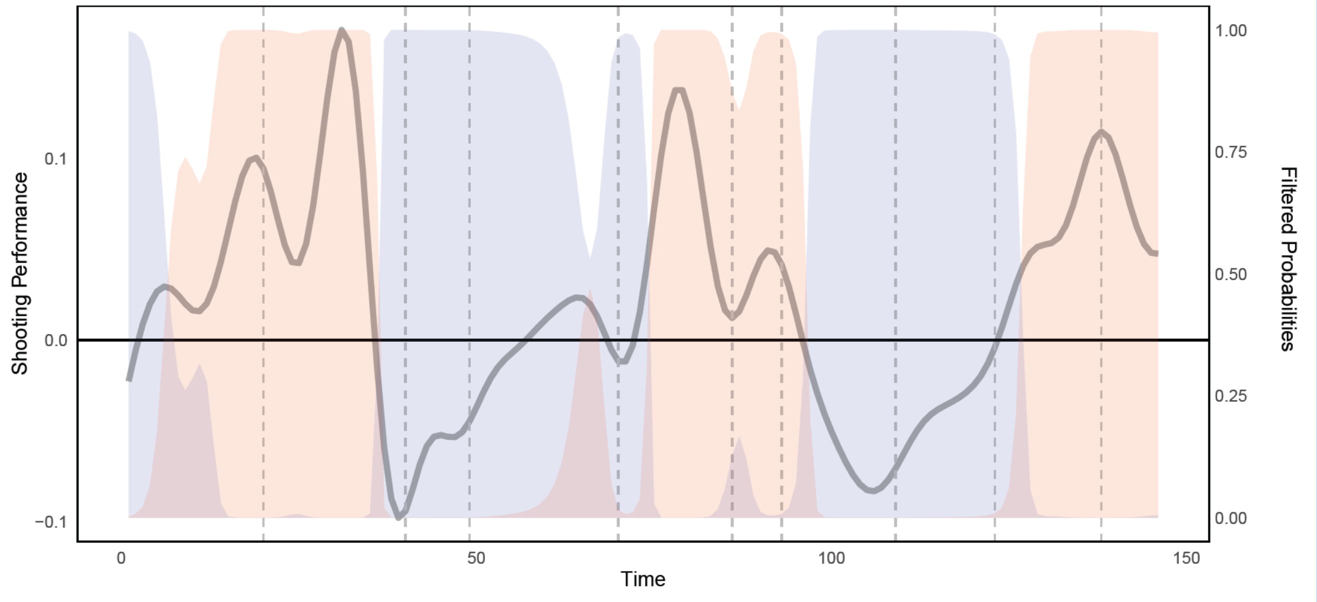
$$\phi_{ij} = \frac{\tilde{\phi}_{ij}}{\phi(m_{ij})}$$

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

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

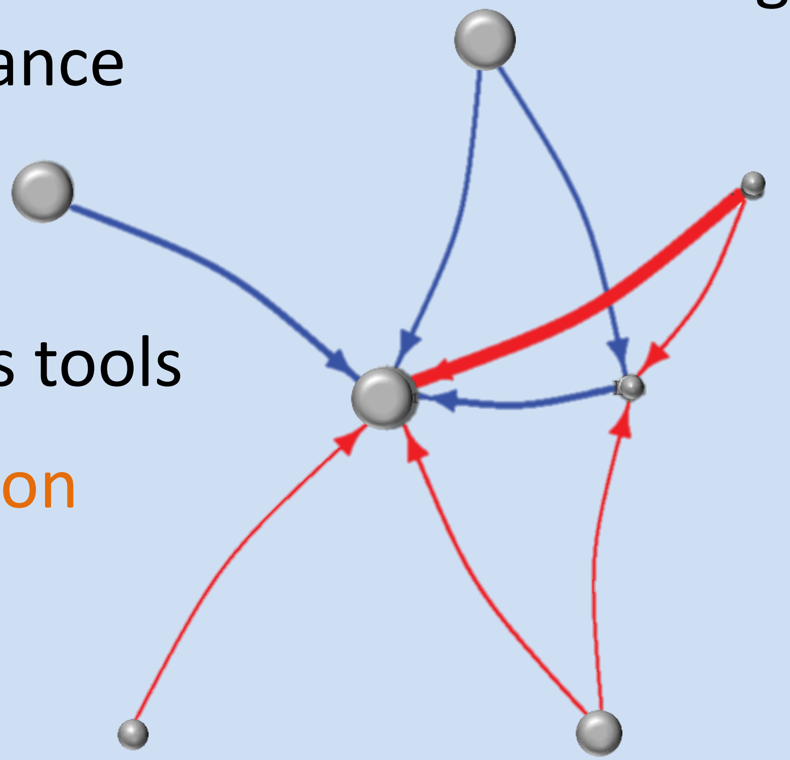
Performance Variability:

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



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



CS4: sensor data analysis

Space-Time Analysis of Movements in Basketball using Sensor Data

Marica Manisera *[1], Rodolfo Metulini †[1], and Paola Zuccolotto ‡[1]

[1]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

Visualization Tools

Department of Economics and Management
University of Brescia

SEMINAR

"Sensor Analytics in basketball: analyzing spatio-temporal movements of players around the court"

Rodolfo Metulini
Università degli Studi di Brescia

Chair: Prof.ssa Paola Zuccolotto

March 15, 2017 – h. 11.00
Sala Biblioteca, S. Faustino building

Global Positioning Systems (GPS) are nowadays intensively used in Sport Science as they capture the trajectories of players and /or the ball, sometimes together with play-by-play recording the time of match events, with the aim of infer to supply coaches, experts and analysts with useful information in addition to traditional statistics. To find any regularities and synchronizations in players' trajectories, and to study their relationship with team's performance, however, is a complex task, because of the strong interdependencies among players in the court and because of external factors that can influence players. To this aim, a variety of methods has been proposed in Sport Science literature, which borrow from the disciplines of Machine Learning, Network and Complex Systems, Geographical Information Systems, Computer Vision and Statistics. In this seminar, with an application to basketball, I propose a methodological approach that can be generalized to other team sports. I first demonstrate the usefulness of a visual tool approach in order to extract preliminar insights from trajectories, then, I use data mining techniques such as Cluster Analysis and Multidimensional Scaling to decompose the game into homogeneous phases in terms of spatial relations. To conclude, I present specific research questions, such as: i) who is the most influencing player of the team? ii) how much each player influences the others? iii) how much trajectories are determined by trajectories of other players and by external factors? where the adoption of methods traditionally used in Spatial Statistics and Spatial Econometrics could have a potential. In this regard, the seminar is also intended as a `platform" to launch new research challenges and to search for collaboration.

Acknowledgements: The research is carried out thanks to sensor data made available by MyAgonism.

Via S. Faustino 74/b, 25122 Brescia – Italy
Tel 00390302988742 Fax 00390302988703
email: paqricerca.dem@unibs.it

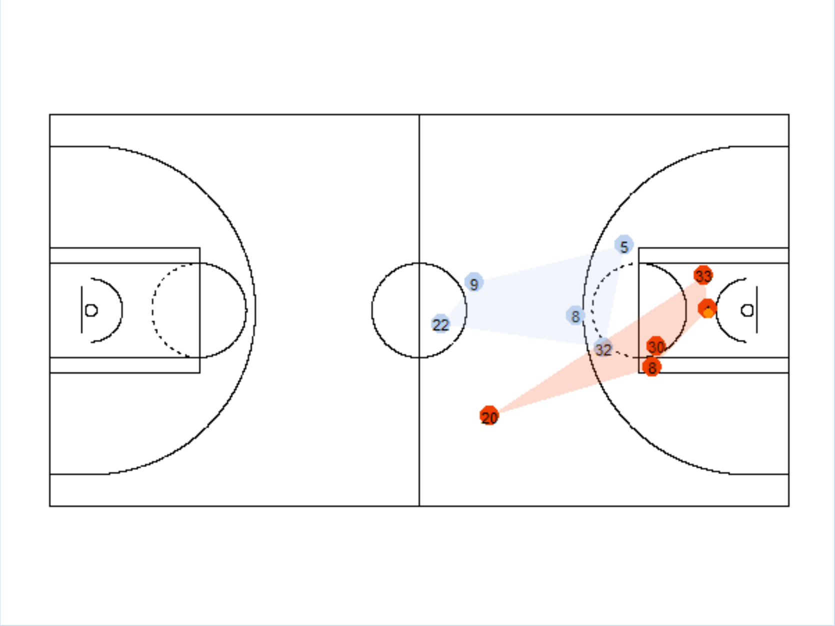
Spatio-Temporal Movements in Team Sports: A Visualization approach using Motion Charts.

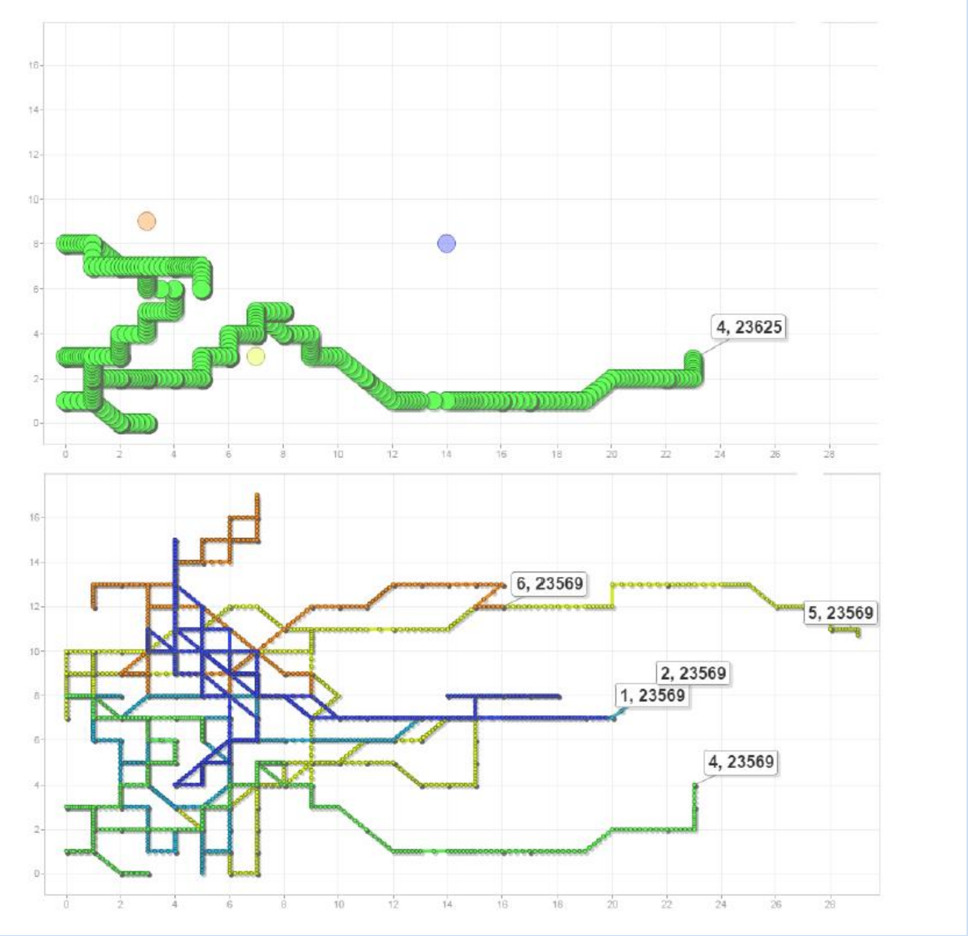
Rodolfo Metulini (1)
(1) *Department of Economics and Management, University of Brescia, Contrada Santa Chiara, 50, 25122 Brescia BS, Italy,*
rodolfo.metulini@unibs.it.

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

Visualization Tools

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

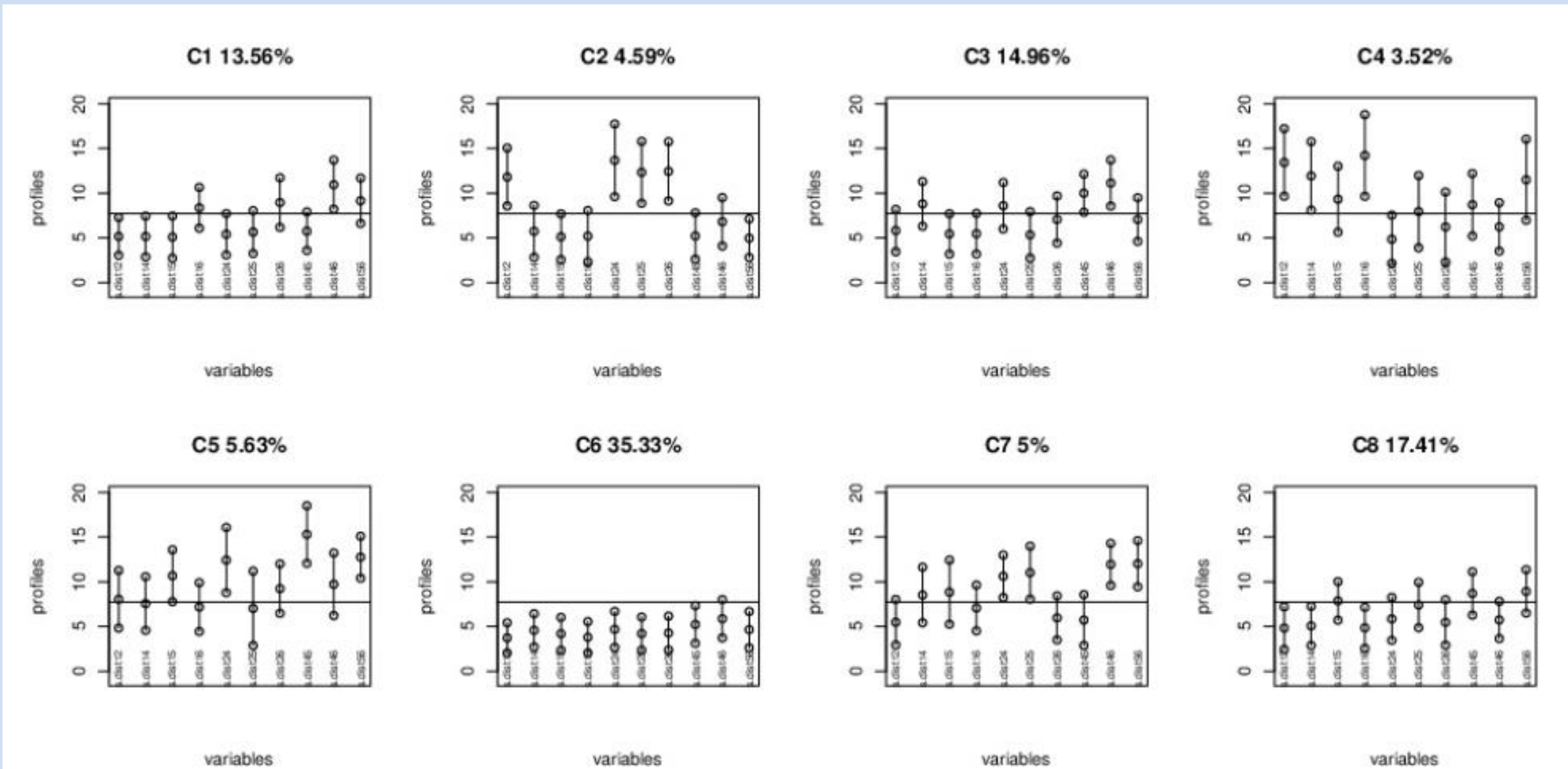




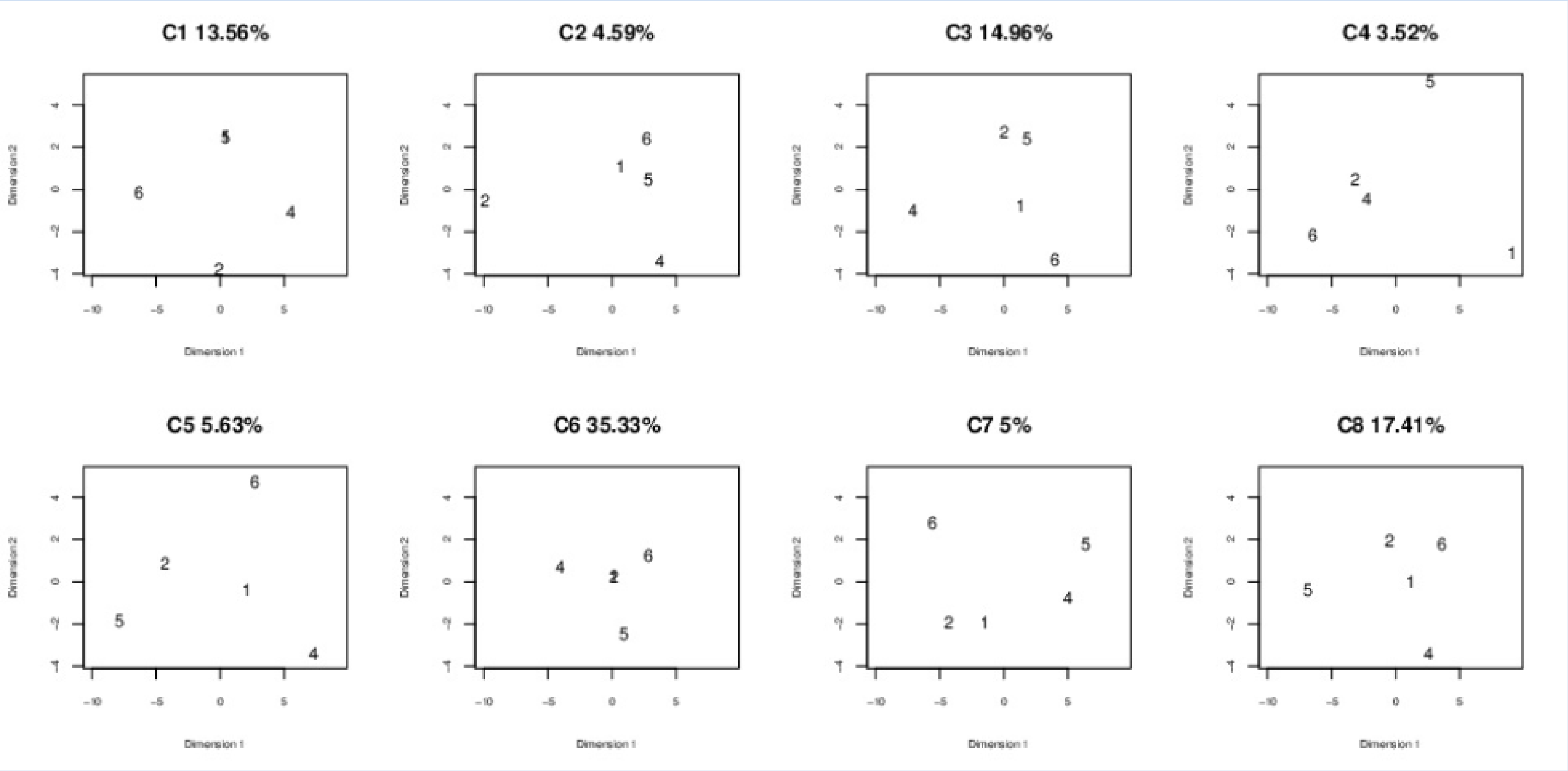
Convex Hulls Analysis

	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

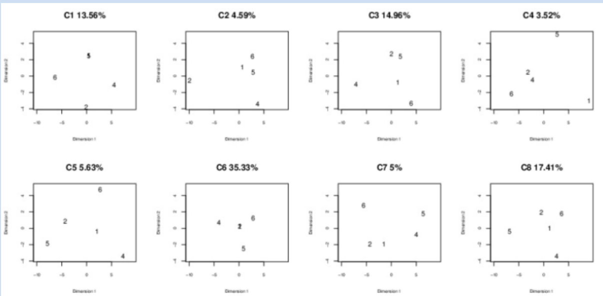
Cluster Analysis



Cluster Analysis + MultiDimensional Scaling



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

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

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**

BDSpORTS
Big Data Analytics in Sports

Marica Manisera
Paola Zuccolotto
– University of Brescia, Italy

References

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

Thank You