

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

PAOLA ZUCCOLOTTO – MARICA MANISERA – DAVIDE VERZELETTI  
UNIVERSITY OF BRESCIA, Italy

# AGENDA:

- Background: stressful game situations in basketball
- Datasets
- Preliminary analyses
- Scoring probability prediction with CART
- New shooting performance measures

- **Background: stressful game situations in basketball**
- Datasets

- Preliminary analyses
- Scoring probability prediction with CART
- New shooting performance measures

## Aim:

to analyze players' shooting performance taking into account factors that may generate pressure or stressful game situations.

## BEWARE:

- a big amount of play-by-play data is needed in order to obtain robust estimates of the scoring probabilities
- the analyzed phenomenon is characterized by complex relationships and interactions among the involved variables
- the reaction to pressure may be different according to the professional level of players

- **Background: stressful game situations in basketball**
- Datasets

- Preliminary analyses
- Scoring probability prediction with CART
- New shooting performance measures

## STRESSful 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 match (QUARTER)

- **Background: stressful game situations in basketball**
- Datasets

- Preliminary analyses
- Scoring probability prediction with CART
- New shooting performance measures

## Aim:

to analyze players' shooting performance taking into account factors that may generate pressure or stressful game situations.

## BEWARE:

- a big amount of play-by-play data is needed in order to obtain robust estimates of the scoring probabilities → **BiQ DATA**
- the analyzed phenomenon is characterized by complex relationships and interactions among the involved variables
- the reaction to pressure may be different according to the professional level of players

- **Background: stressful game situations in basketball**
- Datasets

- Preliminary analyses
- Scoring probability prediction with CART
- New shooting performance measures

## Aim:

to analyze players' shooting performance taking into account factors that may generate pressure or stressful game situations.

## BEWARE:

- a big amount of play-by-play data is needed in order to obtain robust estimates of the scoring probabilities → **BiQ DATA**
- the analyzed phenomenon is characterized by complex relationships and interactions among the involved variables → **DATA MINING**  
**ALGORITHMIC MODELLING**
- the reaction to pressure may be different according to the professional level of players

- **Background: stressful game situations in basketball**
- Datasets

- Preliminary analyses
- Scoring probability prediction with CART
- New shooting performance measures

## Aim:

to analyze players' shooting performance taking into account factors that may generate pressure or stressful game situations.

## BEWARE:

- a big amount of play-by-play data is needed in order to obtain robust estimates of the scoring probabilities → **Big DATA**
- the analyzed phenomenon is characterized by complex relationships and interactions among the involved variables → **DATA Mining**  
**Algorithmic Modelling**
- the reaction to pressure may be different according to the professional level of players → **TWO different DATASETS**

- Background: stressful game situations in basketball
- **Datasets**
- Preliminary analyses
- Scoring probability prediction with CART
- New shooting performance measures

Two datasets (play-by-play records of all the matches played during two competitions):

- Italian “Serie A2” Championship 2015/2016 (A2ITA)
- Olympic Basketball Tournament “Rio 2016” (RIO16)

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)



- Background: stressful game situations in basketball
- Datasets
- **Preliminary analyses**
- Scoring probability prediction with CART
- New shooting performance measures

## DATA MINING Tools:

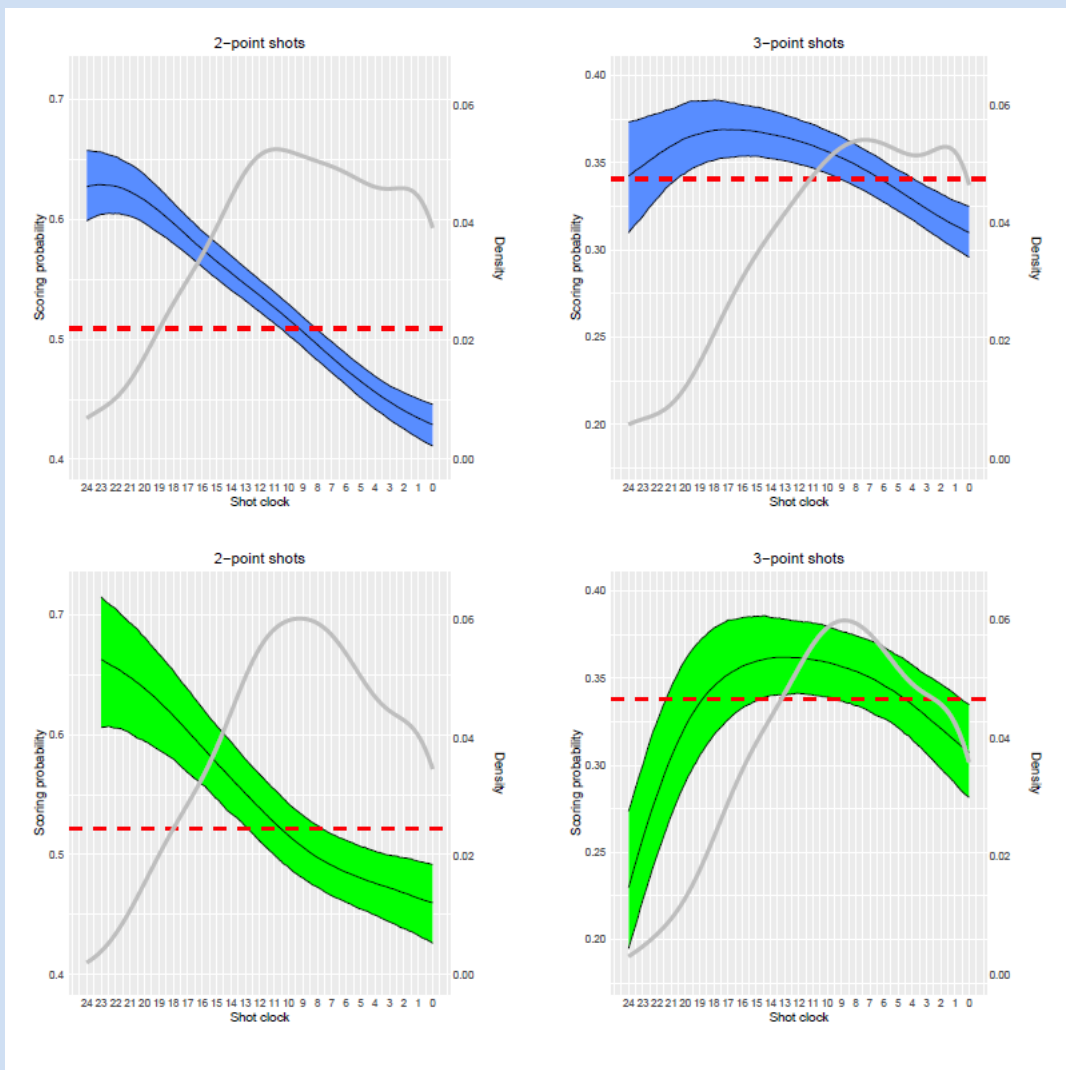
- **univariate** non parametric regressions via kernel smoothing on the dependent variable MADE (assuming values 1 and 0 according to whether, respectively, the attempted shot scored a basket or not)
- 1000 bootstrap samples of size  $nboot = 5000$  and  $nboot = 1000$  for the dataset A2ITA and RIO16,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

➡ **few univariate relationships detected** - Just [SHOT.CLOK](#) and [MISS.PL](#)

- Background: stressful game situations in basketball
- Datasets

- **Preliminary analyses**

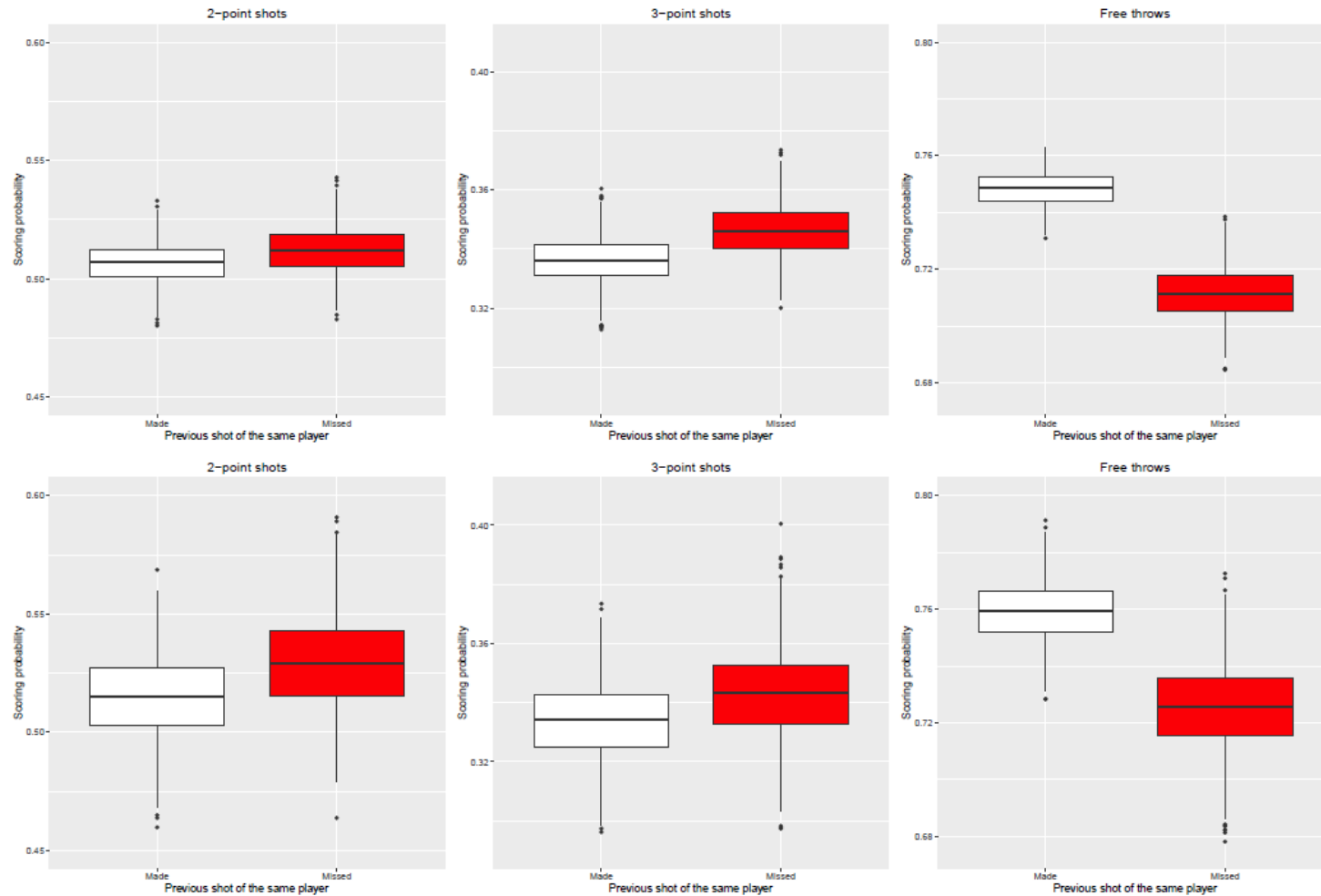
- Scoring probability prediction with CART
- New shooting performance measures



- Background: stressful game situations in basketball
- Datasets

- **Preliminary analyses**

- Scoring probability prediction with CART
- New shooting performance measures



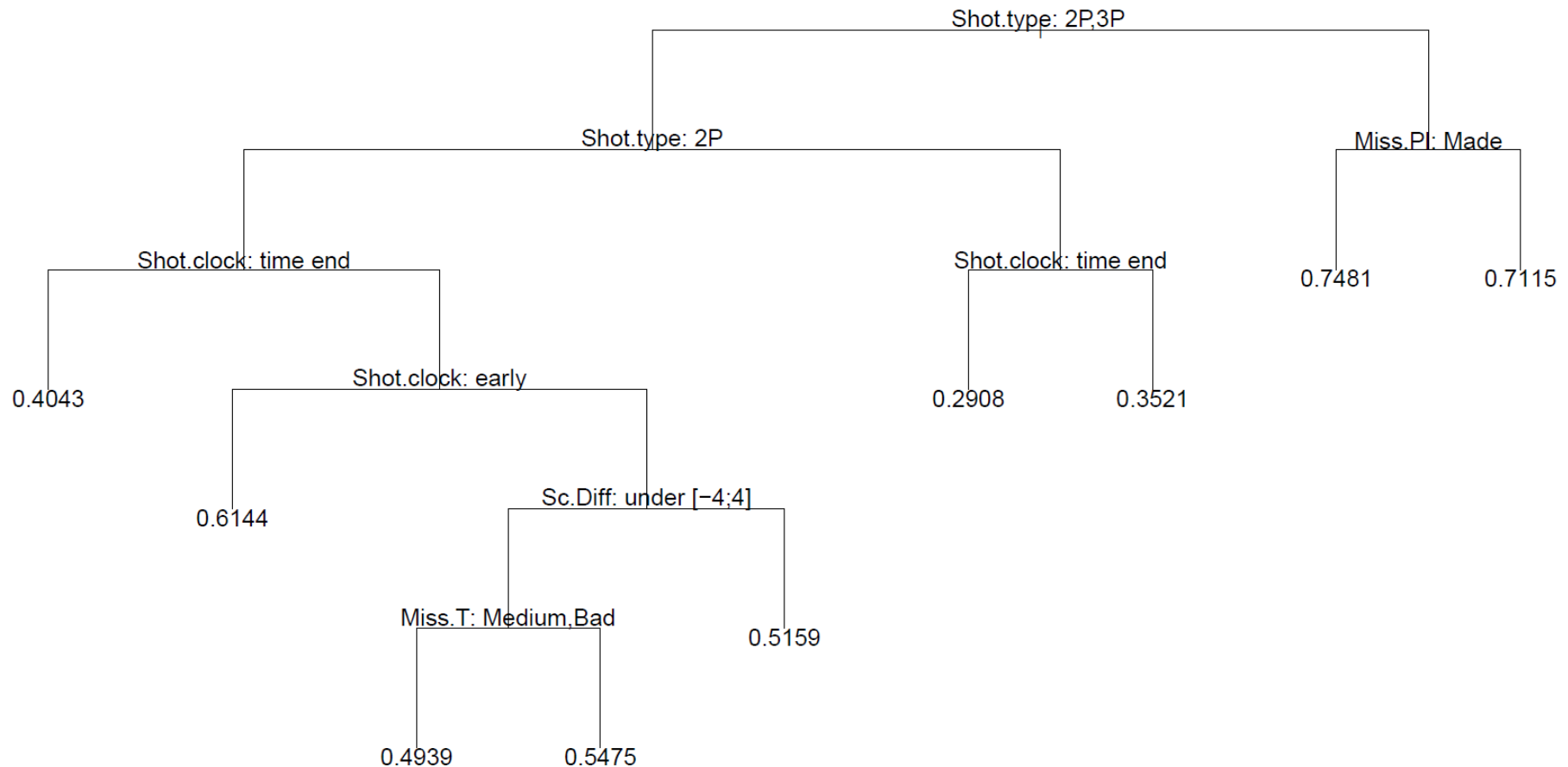
- Background: stressful game situations in basketball
- Datasets
- Preliminary analyses
- **Scoring probability prediction with CART**
- New shooting performance measures

## DATA MINING Tools:

- **CART** (Classification and Regression trees), algorithm able to deal with multivariate complex relationships, also detecting interactions among predictors
- pruning

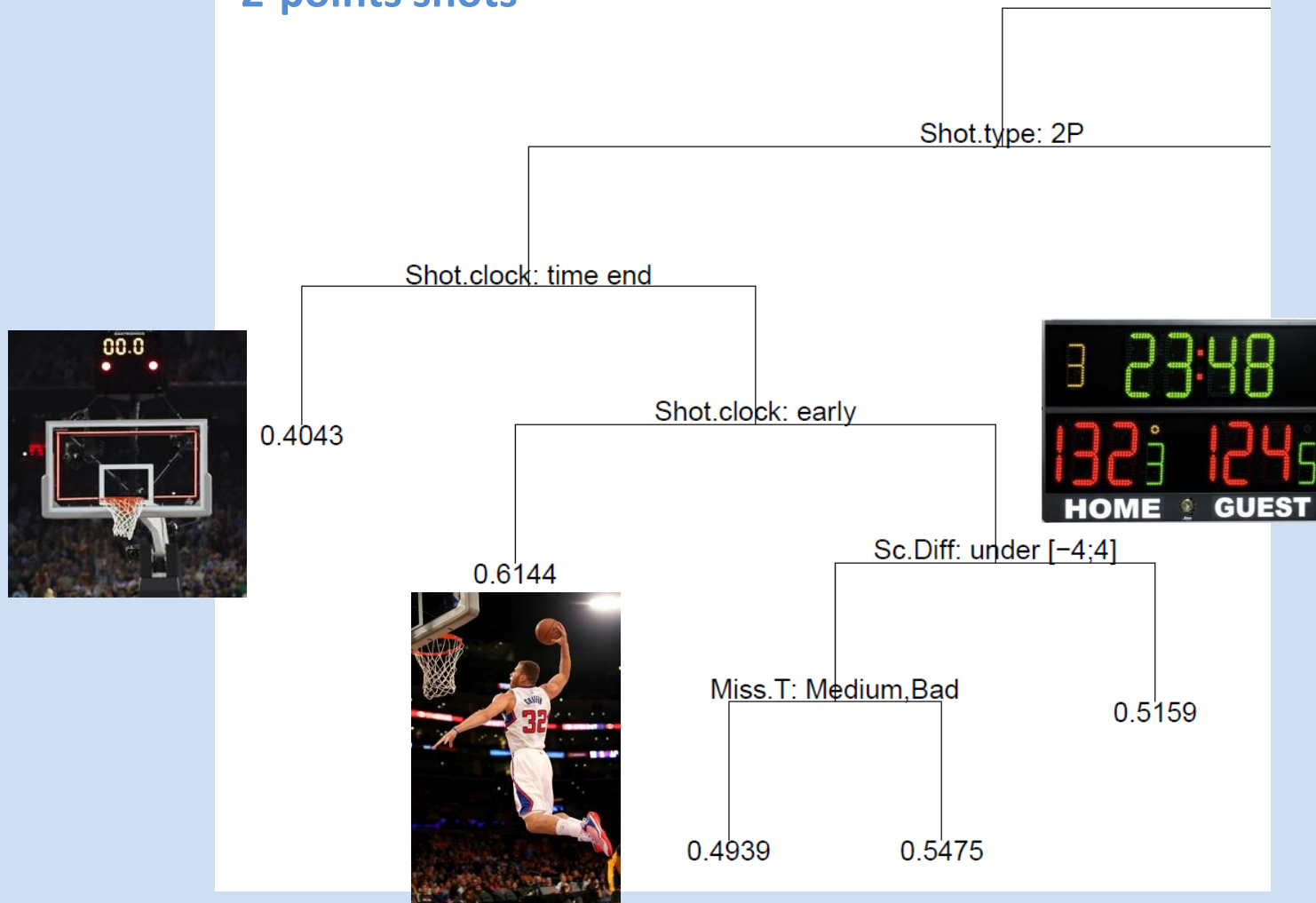
- Background: stressful game situations in basketball
- Datasets
- Preliminary analyses
- **Scoring probability prediction with CART**
- New shooting performance measures

## Very similar results for A2ITA and Rio16



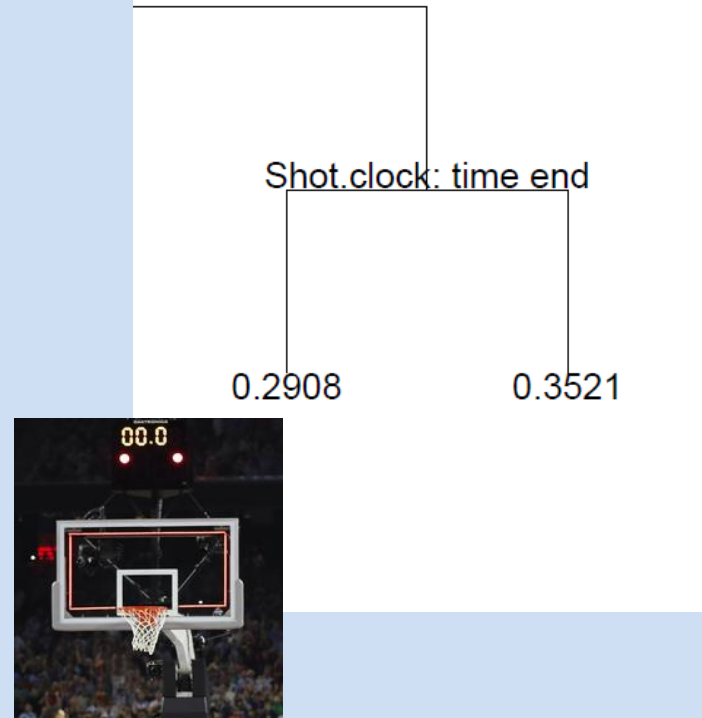
- Background: stressful game situations in basketball
- Datasets
- Preliminary analyses
- **Scoring probability prediction with CART**
- New shooting performance measures

## 2-points shots



- Background: stressful game situations in basketball
- Datasets
- Preliminary analyses
- **Scoring probability prediction with CART**
- New shooting performance measures

### 3-points shots



- Background: stressful game situations in basketball
- Datasets
- Preliminary analyses
- **Scoring probability prediction with CART**
- New shooting performance measures





- Background: stressful game situations in basketball
- Datasets
- Preliminary analyses
- Scoring probability prediction with CART
- **New shooting performance measures**

## NEW SHOOTING PERFORMANCE MEASURE:

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

Performance of Player  $i$

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

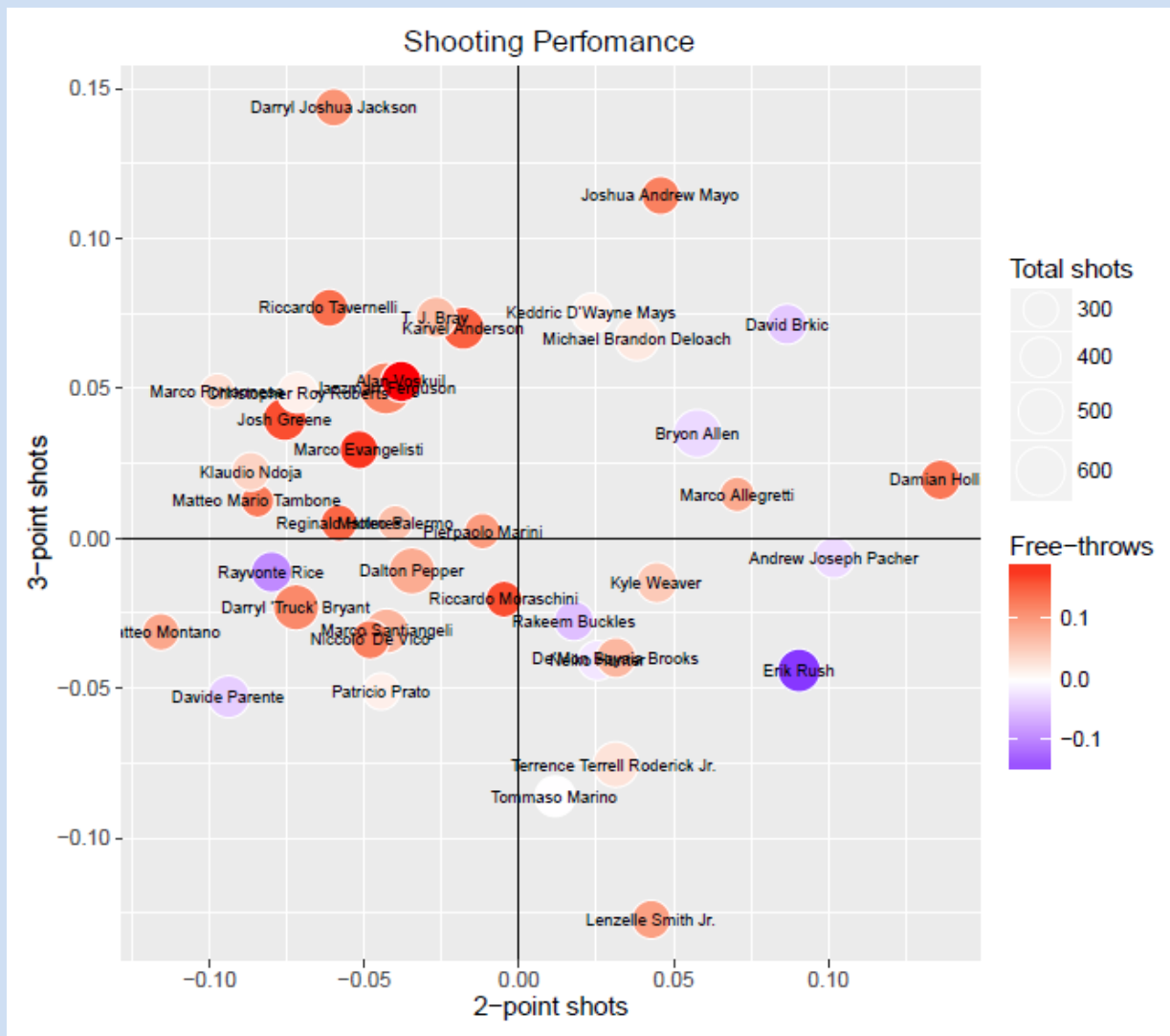
for shot type  $T$   
(2P, 3P, FT)

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

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

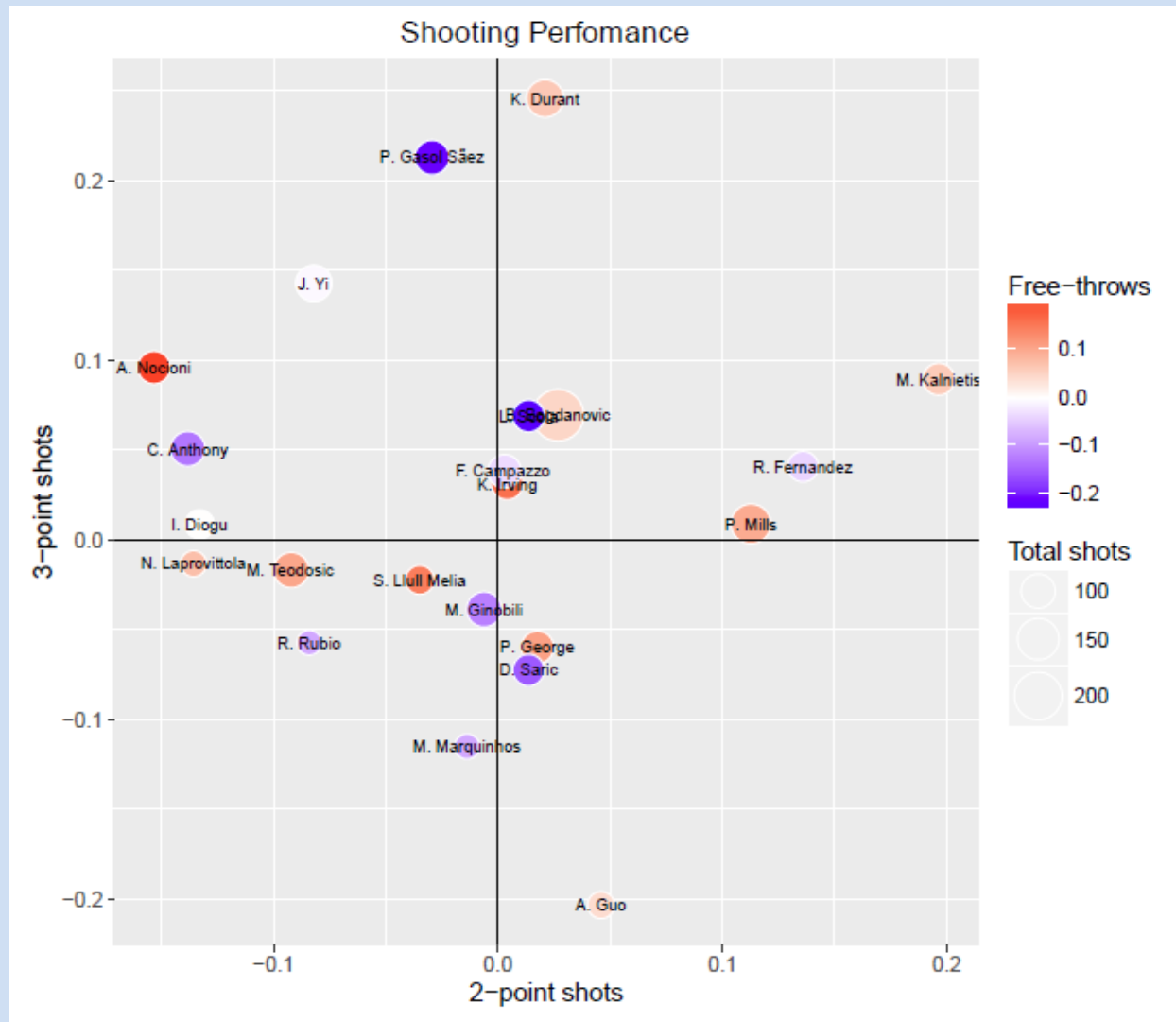
- Background: stressful game situations in basketball
- Datasets
- Preliminary analyses
- Scoring probability prediction with CART
- **New shooting performance measures**

A2ITA



- Background: stressful game situations in basketball
- Datasets
- Preliminary analyses
- Scoring probability prediction with CART
- **New shooting performance measures**

## Rio16



- Background: stressful game situations in basketball
- Datasets
- Preliminary analyses
- Scoring probability prediction with CART
- New shooting performance measures

## 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)**

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

Paola Zuccolotto, Marica Manisera and Marco Sandri

**Summary.** In this paper we analyze the basketball players' shooting performance taking account of factors that may generate pressure or stress. We consider two play-by-play datasets, related to different professional level of

- Background: stressful game situations in basketball
- Datasets
- Preliminary analyses
- Scoring probability prediction with CART
- New shooting performance measures

# THANK YOU



[bodai.unibs.it/BDSports](http://bodai.unibs.it/BDSports)