BDsports

Big Data analytics in sports

UNIVERSITÀ STUDIORUM BRIXIAE

UNIVERSITY OF BRESCIA

International Statistical Institute

ISI SPECIAL INTEREST GROUP ON SPORTS STATISTICS

UNICAEN UNIVERSITÉ CAEN NORMANDIE

cnrs

GREYC

Laboratoire de recherche en sciences du numérique

CAEN NORMANDIE

cnrs GDR Groupement de recherche

Sport et activité physique

Marica Manisera
Paola Zuccolotto

YOU CAN'T MEASURE A PLAYER'S HEART (FIRST PART)

DO NOT REDUCE THE ROLE OF STATISTICS TO A COLLECTION OF INDEXES

We are used to starting our talks on basketball data science from the wrong side of understanding sports analytics. The vast majority of people believe that Statistics in basketball can be reduced to counting the number of shots, baskets, points, assists, turnovers. The most skilled ones are aware of the existence of a huge amount of additional and more complex indexes (available for example at NBA.com), but all these statistics (and we now deliberately use lowercase to distinguish it from Statistics, which is the science we want to talk about) per se, have a more or less limited meaning. As the famed NBA coach and executive, Red Auerbach, quite rightly stated "*I don't believe in statistics. There are too many factors that can't be measured. You can't measure a ballplayer's heart.*" He meant lowercase statistics.

We do not assert that we are able to measure a player's heart, or the flicker of imagination that generates a sensational dunk or, even the stroke of luck that turns the tide of a game. All these elements, which make our hearts skip a beat, lift us to our feet to rejoice, fill us with wonder - all that sports really is in our hearts – fortunately, cannot be measured.

But, on the other hand, it is not correct to reduce the role of Statistics to a collection of indexes trying to get the game down to a set of numbers. For a statistician, these indexes are simply the data to start from. These data collected in large quantities and appropriately re-elaborated – often using very complex techniques and algorithms – can be transformed into useful information to support technical experts in their decisions. And, if Statistics is not able to measure a player's heart, it is, however, able to measure very subtle aspects, such as the way a player interprets his role, the impact of high-pressure game situations, the teamwork and the positive or negative influence of teammates on a player's performance. This is not done by counting the number of assists, points, steals. You do not need to be an expert to understand it. Statistics can rely on a large set of techniques, models and algorithms for data description and visualization, multiple and complex relationships detection, drivers identification, prediction. But beyond all of this, a well-developed research line deals with the study of latent variables, that is, all those variables that are not concretely and physically measurable. Not exactly the player's heart, but his mental toughness, sense of confidence, leadership, sense of motivation, inner desire to succeed, self-discipline, sense of belonging, these are all psychological traits that can be measured by Statistics and possibly studied in relation to the player's or the team's performance.

What we want to stress is that Statistics is a tool for decisions and not a substitute for human intelligence. Statistics alone does not provide decisions but supports decisions. Decisions remain up to the basketball experts. Algorithms can never replace the human brain when it comes to pulling together the strings of all the evidence and formulating a final judgment, which will be based on all the gathered information, bringing together the quantitative and qualitative issues into a single perception. At the same time, all the imponderable and unconscious elements, intuitions, expertise and other knowledge that cannot be formalized in a structured way or howsoever measured must be taken into account. The player's hearts must definitely be taken into account as well.

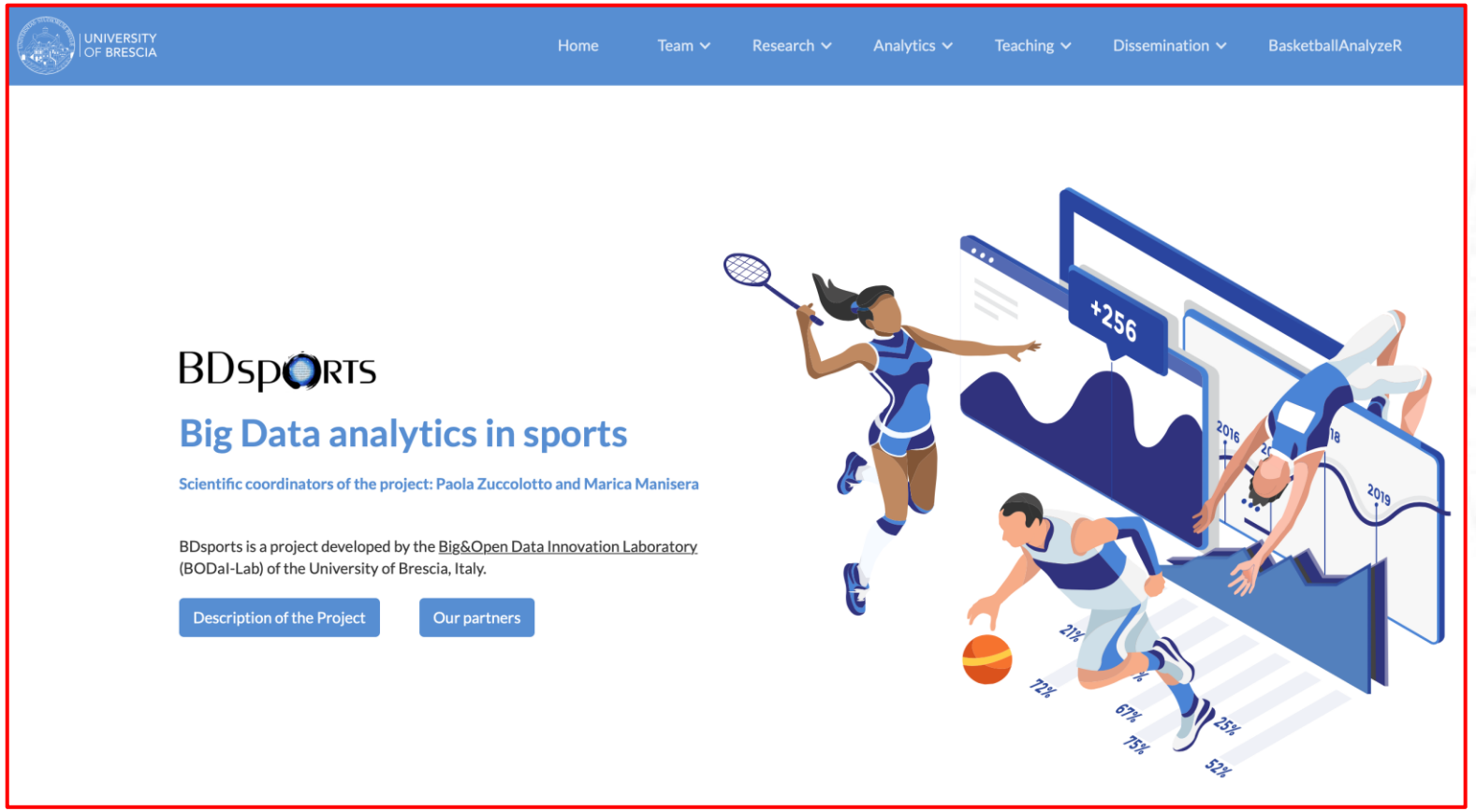
STATISTICS ALONE DO NOT PROVIDE DECISIONS

In doing that, the teamwork of the statistician and the sports expert is of paramount importance. The expert describes the matter, by posing the right research questions, which allows the statistician to understand the problem and narrow the focus. Then, expert and statistician together plan the research design: decide which data to collect, considering constraints, practical feasibility, pros and cons of each choice. Subsequently, the analyst carries out the analysis, using his statistical expertise, and then the ball goes back to both of them jointly for the interpretation of the results, never forgetting the strengths and limitations of the techniques that have been used. The information and knowledge drawn from the data thanks to Statistics is the support for the basketball expert's decisions. As mentioned above, he must summarize all the available information: the quantitative evidence provided by Statistics and the qualitative ones deriving from his intuitions which, coming from experience, are no less

paola.zuccolotto@unibs.it
marica.manisera@unibs.it

Big Data analytics in sports

bdsports.unibs.it

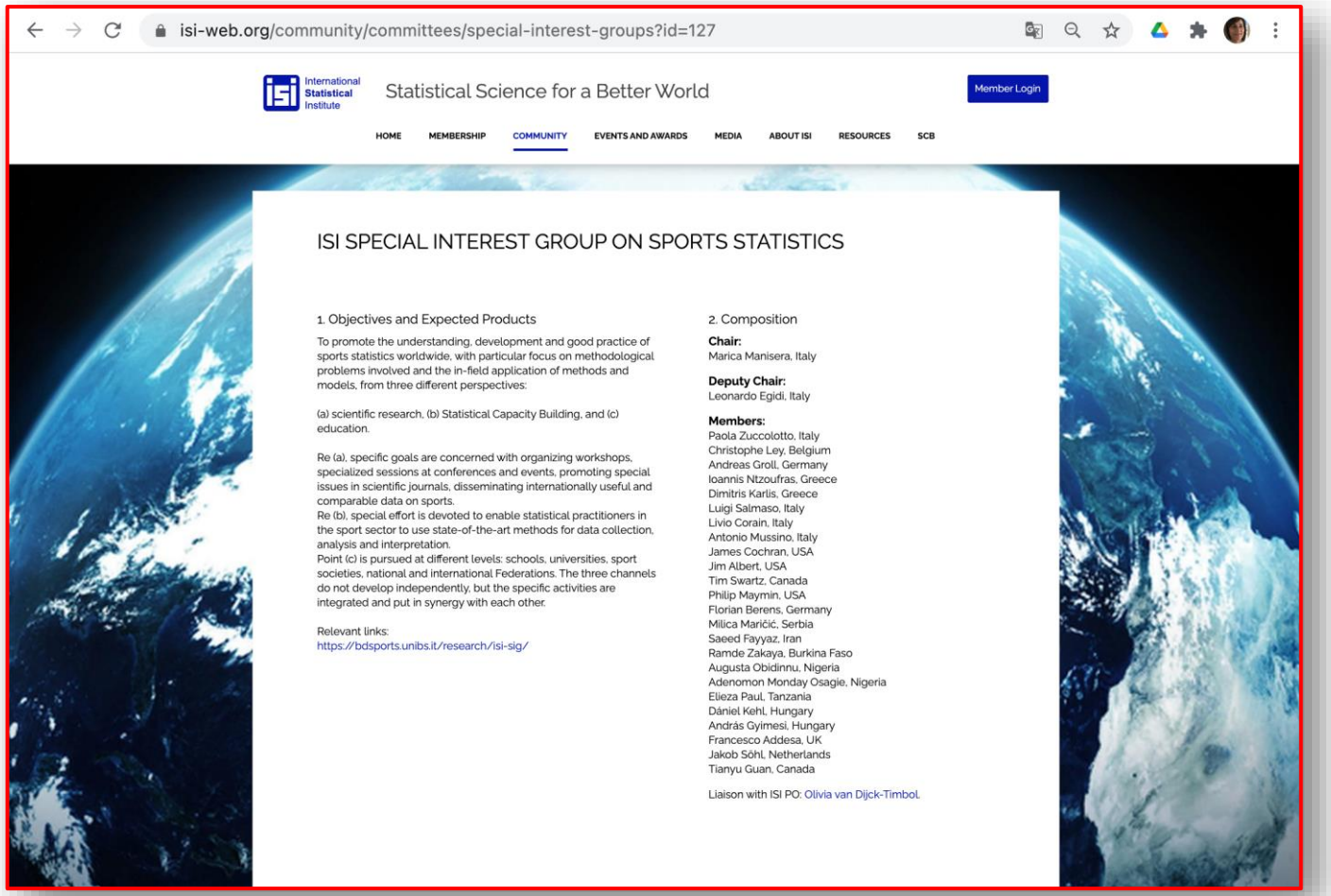


UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

ISI Special Interest Group on SPORTS STATISTICS

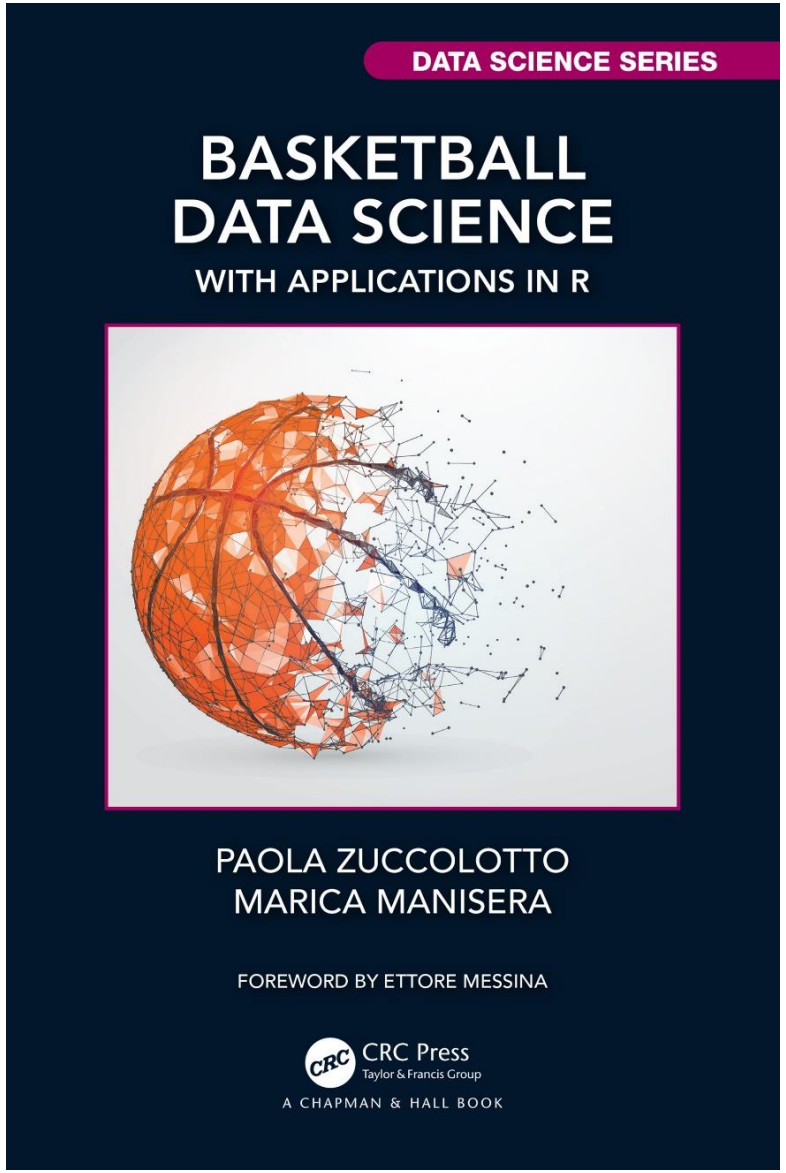
https://www.isi-web.org/community/committees/special-interest-groups?id=127

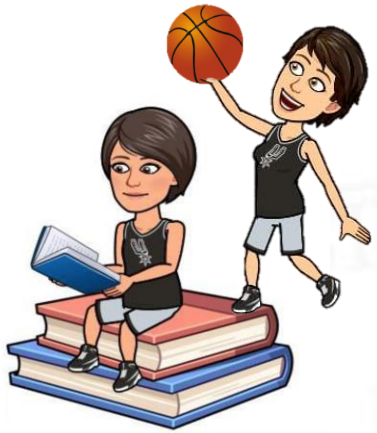


UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

Our Book





R Package
developed by Marco Sandri

UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

Agenda

1. Data Science in basketball
2. Basketball analytics: state of the art
3. Basketball data
4. Introduction to the R package `BasketballAnalyzeR`

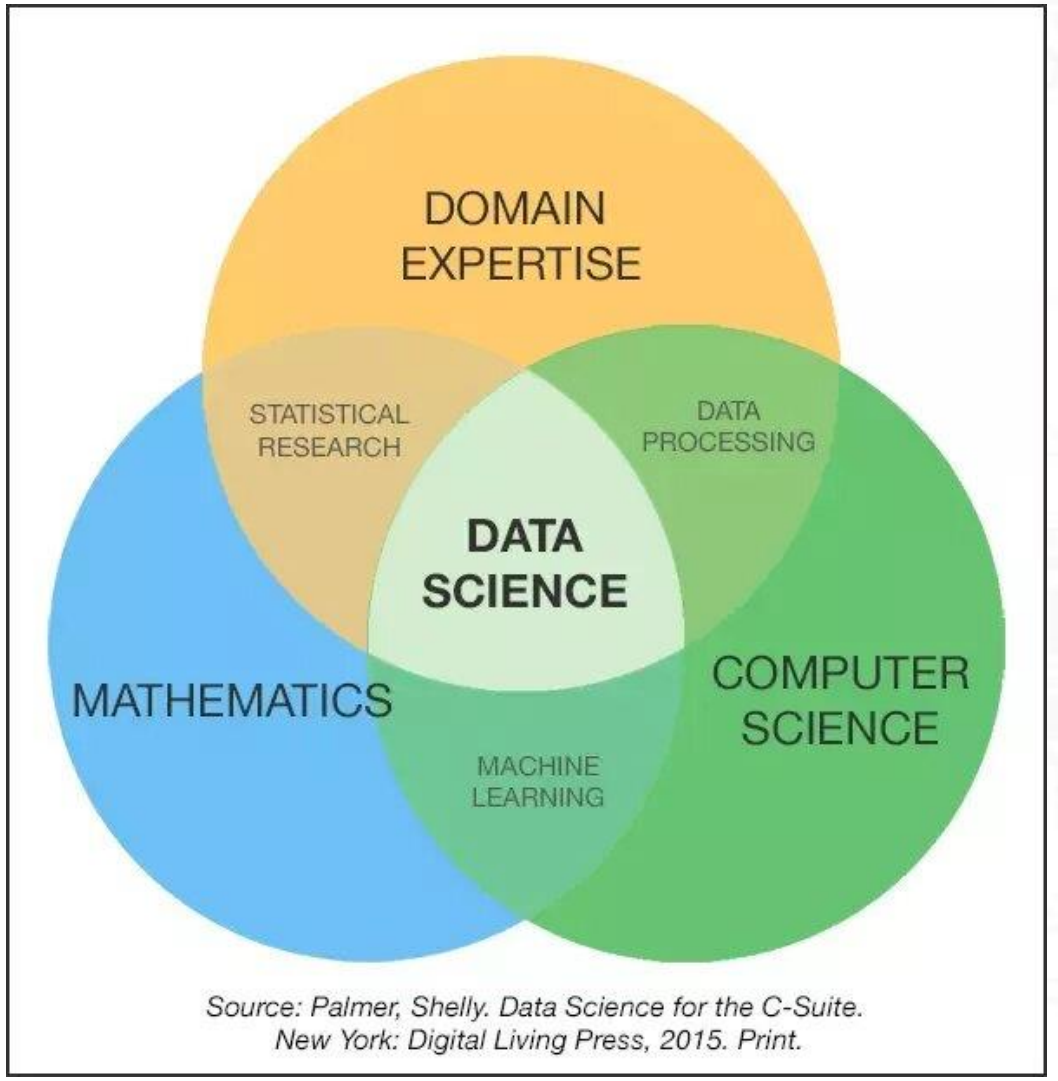
1 - Data Science in Basketball

What is Data Science?

Discipline aimed at
extracting knowledge
from data in various forms

Multidisciplinary

Applicable to
a wide range of fields



Source: Palmer, Shelly. Data Science for the C-Suite. New York: Digital Living Press, 2015. Print.

UNIVERSITY OF BRESCIA

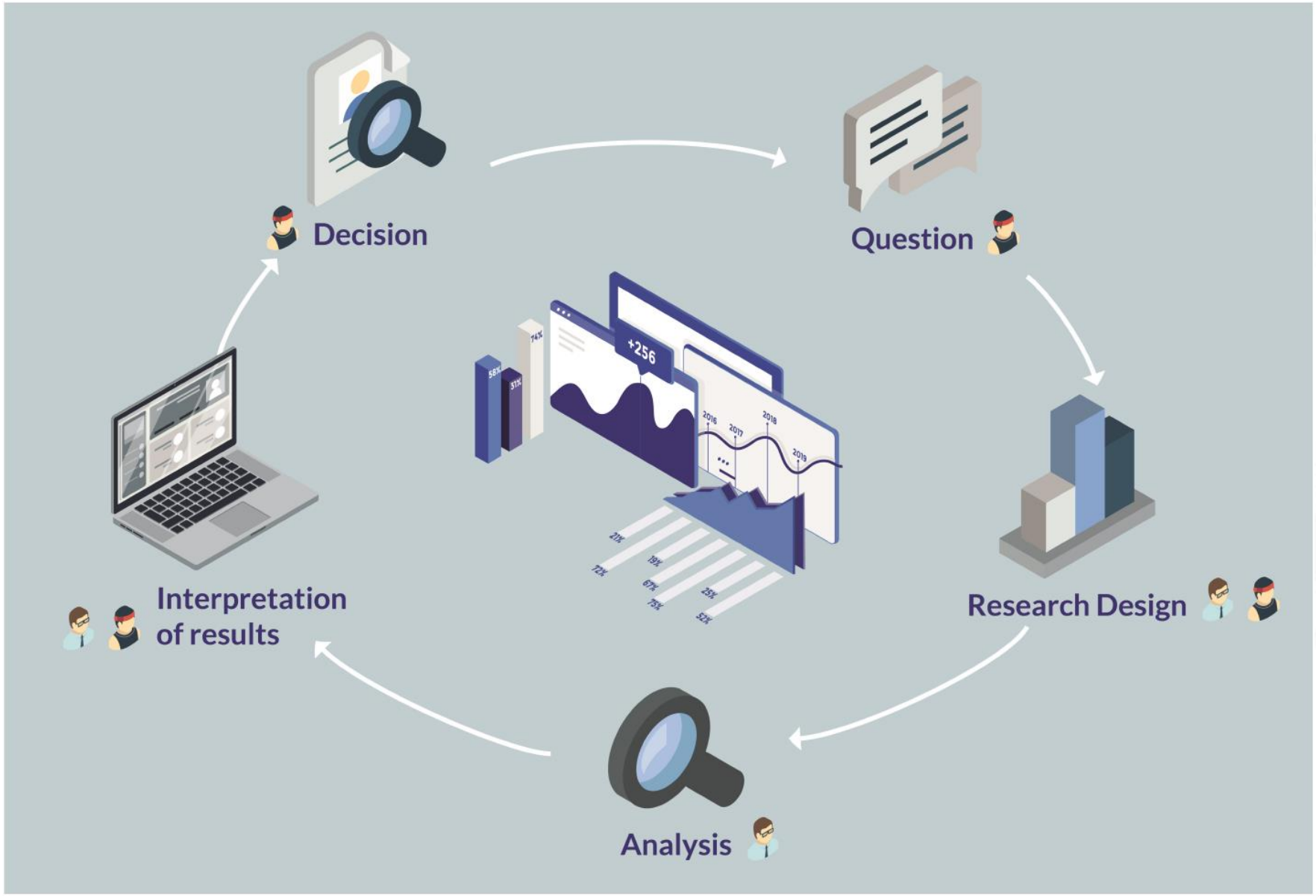
BDsports
Big Data analytics in sports

Data Science…

- … aimes at extracting knowledge from the data (interpretation of results is extremely delicate)
- … can deal with any field of human knowledge
- … can potentially answer any question, if it has the right data
- … will never be able to describe everything
- … is not a crystal ball
- … does not provide decisions, but support for decisions

Basketball data science has no ambition to replace basketball experts, but to support them in their decisions

The virtuous cycle of Sports Analytics



UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

Anatomy of a decision

Are stats killing the game of basketball?

MARC GASOL SAYS: 'STATS ARE KILLING THE GAME OF BASKETBALL' (2017)

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)

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**

UNIVERSITY OF BRESCIA

BDSPORTS
BIG DATA ANALYTICS IN SPORTS

Are stats killing the game of basketball?

Marc Gasol considers data very important and beneficial for winning

World champion and NBA basketball player Marc Gasol brought some stardust to proceedings at the Sports Tomorrow Congress (WOM+N) 2021 on Thursday

(2021)

"…the confidence of the person who showed me the data convinced me about them. We were used to seeing images of games, but they transformed those images into numbers. **Once you have the data, they help you make better decisions**"…. [they should] "educate players on the importance and benefits of data. The best thing they can do is make the most of them to squeeze as much as possible out of games, as **data is very important and beneficial for winning**"

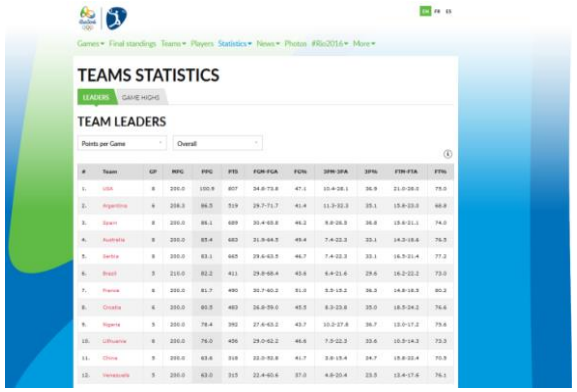
UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

2 – Basketball Analytics: state of the art

Basketball Analytics

Official Statistics



Sport Analytics Services

BIGDATABALL
XY
play

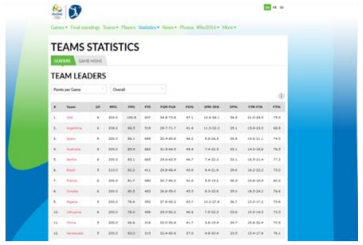
Scientific Research

UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

Basketball Analytics

Official Statistics



Sport Analytics Services

BIGDATABALL play

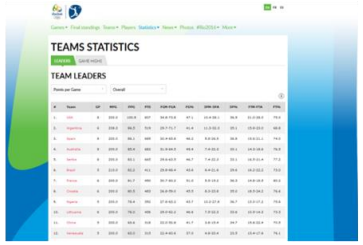
Scientific Research

UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

Basketball Analytics

Official Statistics



Sport Analytics Services

BIGDATABALL
XY
play

Scientific Research

Our analyses often integrate machine learning tools and experts' suggestions

UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

Basketball Analytics

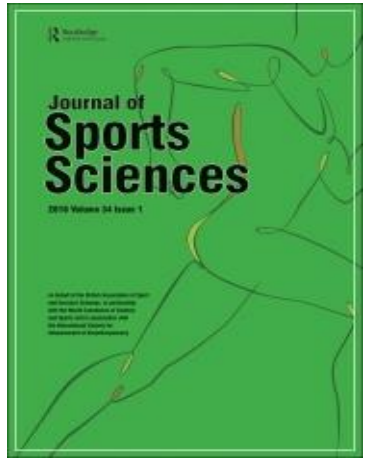
Official Statistics

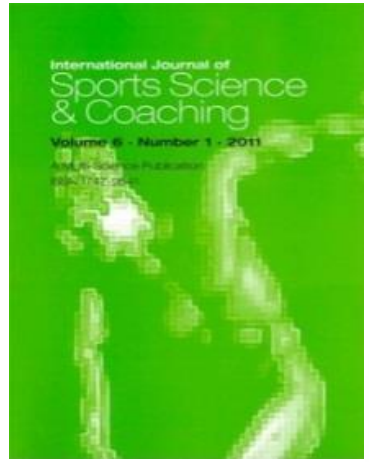
Sport Analytics Services

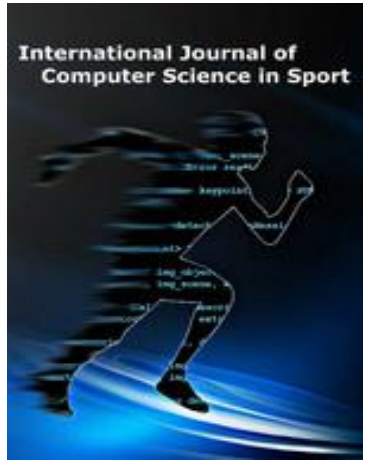
Scientific Research

Scientific literature

Scientific journals







Special Issues

UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

Basketball Analytics

Scientific literature

- **Predicting the outcomes** of a game or a tournament
- Determining **discriminating factors** between successful and unsuccessful teams
- Examining the statistical properties and **patterns of scoring** during the games
- Analysing a **player's performance** and the impact on his team's chances of winning
- Monitoring **playing patterns** with reference to roles

Basketball Analytics

Scientific literature

- Designing the **kinetics** of players' body movements with respect to shooting efficiency, timing and visual control on the field
- Depicting the **players' movements, pathways, trajectories** and the **network of passing actions**, the flow of events and the connected functional decisions
- Studying **teams' tactics** and identifying optimal **game strategies**
- Investigating possible **referee biases**

Basketball Analytics

Scientific literature

- Measuring **psychological latent variables** and their association to performance
- Epidemiology of basketball **injuries, physical, anthropometric and physiological attributes** of players, hematological parameters or other vitals
- Special **training programmes** to stimulate muscle strength, jumping ability and physical fitness in general
- **Scheduling** problems

Basketball Analytics

Scientific literature

- This list is far from being complete
- The range of possible research questions is going to grow, thanks to the availability of large data sets and the incresaing computational power
- A complete theory explaining the relationships among the variables involved in basketball analytics is still not available

→ Answering to all those questions is a very interesting challenge for Data Scientists

3 – Basketball data

Basketball Data

Data are essential to Data science and Analytics, so the procedures for obtaining and organizing data sets must be structured and validated to guarantee **Quality**:

Exhaustiveness

Accuracy

Completeness

Consistency

Accessibility

Timeliness

Basketball Data

Another important issue about data is **Context** (all the additional information necessary to correctly interpret data):

"*Data without context are just numbers*"

Several **sources** (Federations, sporting organizations, professional societies, associations, …)

Basketball Data

The **web** is a massive store of data:

- Data on payment or freely available
- Open data often require web scraping procedures
- Variety of datasets (traditional data matrices, multidimensional data cubes, unstructured text data, pixels from sensors and cameras, data from wearables, mobile phones, tablets, geocode, timestamps, …), requiring relational databases and datawarehousing tools

Basketball Data

We can distinguish four main **macro-categories**:

- Data recorded manually
- Data detected by technological devices
- Data from psychometric questionnaires
- Other data

Basketball Data

Data recorded manually, with or without technological tools for annotation. This category includes the basic statistics from box scores, notational analysis data, play-by-play (event-log) data, reports filled by technical experts and coaches during training sessions, opinions and experts' evaluations that can be combined with measurement data.

Basketball Data

Data detected by technological devices. Increasingly, technology enters both the training and the games, making available large amounts of data. Examples are the data recorded by GPS sensors or other player tracking systems, which detect the positions of the players on the court at very short time intervals (milliseconds), the video data coming from cameras, the platforms and all the wearable technologies that detect postures, body movements, vitals such as heartbeat and blood pressure.

Basketball Data

Data from psychometric questionnaires administered to athletes, aimed at the measurement of attitudes and personality traits (group dynamics, interpersonal relations, social-cognitive processes, leadership, mental toughness, personality, coping strategies, ...).

UNIVERSITY OF BRESCIA

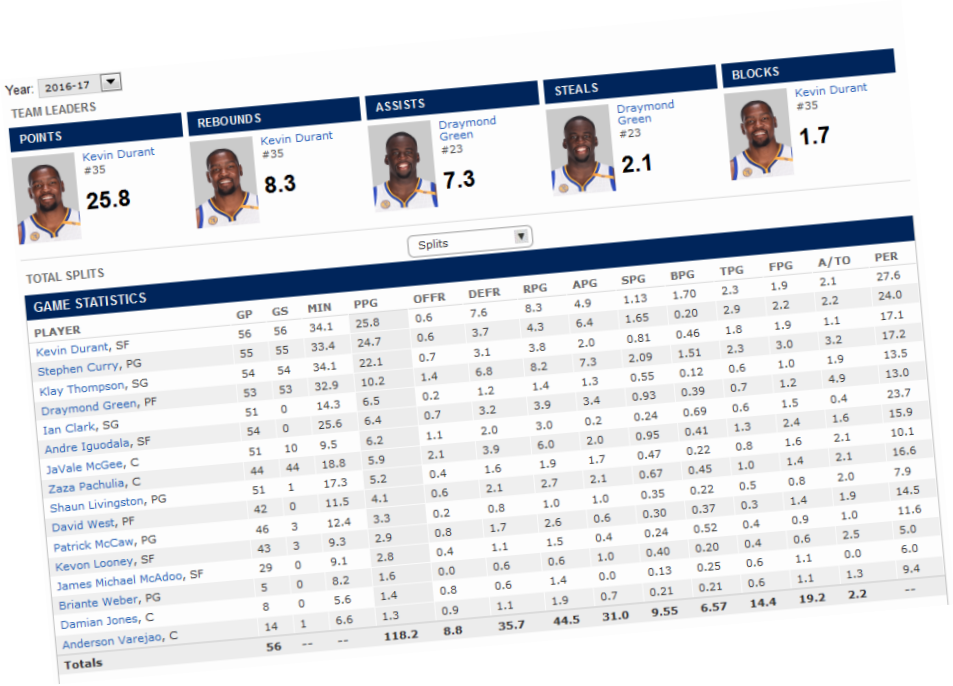
BDsports
Big Data analytics in sports

Basketball Data

Other data. In this residual category converge all the different and heterogeneous data classes that can integrate the analysis from different points of view, such as - without pretension of exhaustiveness - the market analysis data, the textual data obtained by querying the Social Networks (which can serve for example to measure the sentiment of the fans), data from Google Trends and other tools able to monitor online searches and popularity of hashtags.

Basketball 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

www.espn.com/nba

stats.nba.com

www.fiba.com

Leagues

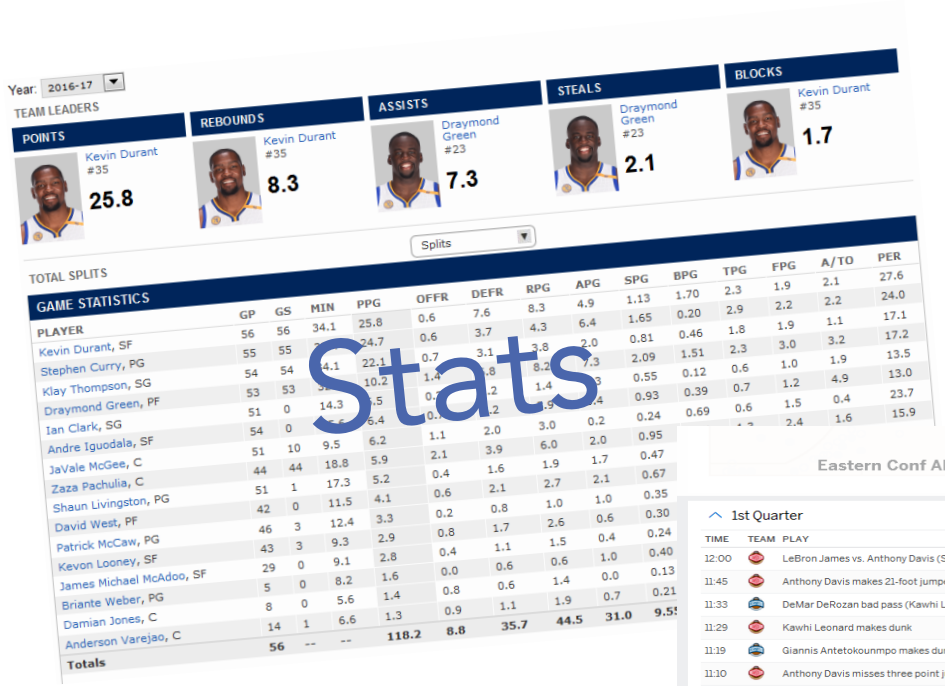
…

UNIVERSITY OF BRESCIA

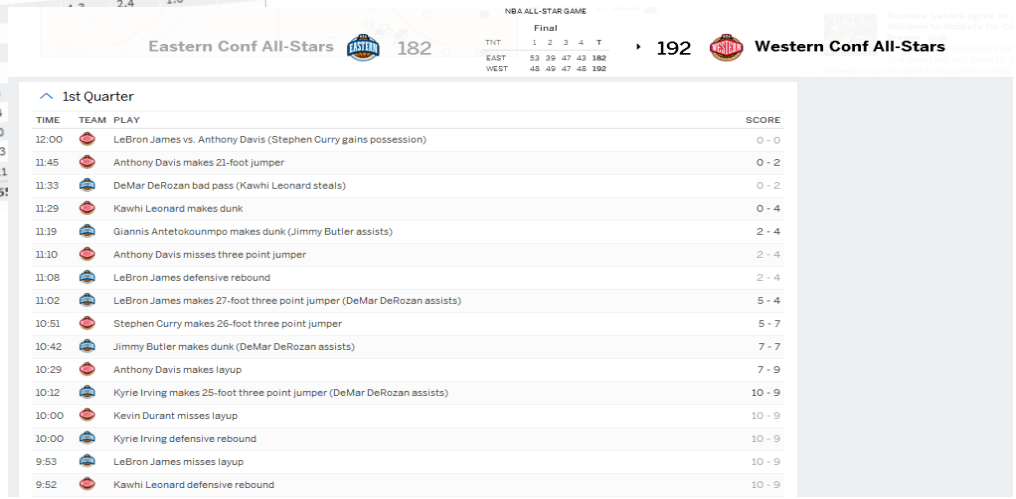
BDsports
Big Data analytics in sports

Basketball Data

Data → **Big Data**



play-by-play



UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

Basketball 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	[illegible]	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	[illegible]	10.2	1.4	[illegible]	8.2	7.3	2.09	1.51	2.3	3.0	3.2	17.2
Ian Clark, SG	51	0	14.3	[illegible]	[illegible]	[illegible]	1.4	[illegible]	0.55	0.12	0.6	1.0	1.9	13.5
Andre Iguodala, SF	54	0	[illegible]	6.4	0.7	1.2	[illegible]	[illegible]	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		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					
David West, PF	42	0	11.5	4.1	0.6	2.1	2.7	2.1	0.67					
Patrick McCaw, PG	46	3	12.4	3.3	0.2	0.8	1.0	1.0	0.35					
Kevon Looney, SF	43	3	9.3	2.9	0.8	1.7	2.6	0.6	0.30					
James Michael McAdoo, SF	29	0	9.1	2.8	0.4	1.1	1.5	0.4	0.24					
Briante Weber, PG	5	0	8.2	1.6	0.0	0.6	0.6	1.0	0.40					
Damian Jones, C	8	0	5.6	1.4	0.8	0.6	1.4	0.0	0.13					
Anderson Varejao, C	14	1	6.6	1.3	0.9	1.1	1.9	0.7	0.21					
Totals	56	--	--	118.2	8.8	35.7	44.5	31.0	9.5					

Stats

Sensor data

NBA ALL-STAR GAME

Final

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

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

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	
11:08		LeBron James defensive rebound	
11:02		LeBron James makes 27-foot three point jumper (DeMar DeRozan assists)	
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

UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

4 - Introduction to the R package

`BasketballAnalyzeR`

UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

Book and codes

Contents

UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

Book and codes

PART III Technical Appendix

UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

https://bdsports.unibs.it/basketballanalyzer/

UNIVERSITY OF BRESCIA

Home | Team | Research | Analytics | Teaching | Dissemination | BasketballAnalyzeR

BasketballAnalyzeR

`BasketballAnalyzeR` is an R package that accompanies the book:

P. Zuccolotto and M. Manisera (2020) *Basketball Data Science – With Applications in R*, Chapman and Hall/CRC. ISBN 9781138600799.

It has been developed by Marco Sandri, Paola Zuccolotto, Marica Manisera (Big&Open Data Innovation Laboratory BODaI-Lab, University of Brescia) and provides functions for analysis and visualization of Basketball Data.

This web page gives details and information about the package.

Supplementary material for the book

Codes for reproducing the case studies presented throughout the book with `BasketballAnalyzeR` – R code checked under R-3.5.3 for Windows (64 bit) and for Mac OS X 10.11

pdf format

UNIVERSITY OF BRESCIA

BDsports Big Data analytics in sports

Install R and BasketballAnalyzeR

How to install BasketballAnalyzeR

`BasketballAnalyzeR` is on CRAN and github.

Step 1 – Installing R

- Download the latest version of R from CRAN
- Install R following the instructions of the installer (you can safely use the default settings and just keep clicking Next until R starts installing)
- If you have experienced problems with the installation, read for example here or here

Step 2 – Installing BasketballAnalyzeR

There are three alternative procedures, that can be optionally chosen by the user.

PROCEDURE 1 – STANDARD INSTALLATION FROM CRAN

- Open R
- Write `install.packages("BasketballAnalizeR")` and then press *Enter*
- Wait until the package is installed
- To load the `BasketballAnalyzeR` package, write `library(BasketballAnalyzeR)` and then press *Enter*
- To test the package, write `example(shotchart)` and then press *Enter*

PROCEDURE 2 – INSTALLATION FROM LOCAL ZIP FILE

- Download the package's zip file
- Install it from local zip file (from the R menu: Packages -> Install package(s) from local files -> BasketballAnalyzeR_0.5.0.tar.gz)
- Wait until the package is installed (it can take several minutes)
- To load the `BasketballAnalyzeR` package, write `library(BasketballAnalyzeR)` and then press *Enter*
- To test the package, write `example(shotchart)` and then press *Enter*

PROCEDURE 3 – INSTALLATION OF DEVELOPMENT VERSION

With this procedure the user can install the latest version of the package, with the most recent updates in development version, not yet implemented in the CRAN version of the package.

- Open R
- Write `install.packages("devtools")` and then press *Enter*
- To download and install the BasketballAnalyzeR package, write `devtools::install_github("sndmrc/BasketballAnalyzeR")` and then press *Enter*
- Wait until the package is installed (it can take several minutes)
- To load the `BasketballAnalyzeR` package, write `library(BasketballAnalyzeR)` and then press *Enter*
- To test the package, write `example(shotchart)` and then press *Enter*

UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

Data

```
data(package="BasketballAnalyzeR")
```

```
Data sets in package 'BasketballAnalyzeR':

Obox                                    Opponents box scores dataset - NBA 2017-2018
PbP.BDB                                 Play-by-play dataset - NBA 2017-2018
Pbox                                    Players box scores dataset - NBA 2017-2018
Tadd                                    Tadd dataset - NBA 2017-2018
Tbox                                    Teams box scores dataset - NBA 2017-2018
```

- Tbox – Teams' box scores
- Obox – Opponents' box scores
- Pbox – Players' box scores
- PbP.BDB – Play-by-play data
- Tadd – Additional information

NBA Regular Season 17/18
82 games

Play-by-play: 82 games played by the Champions, Golden State Warriors (made available by BigDataBall www.bigdataball.com)

```
> PbP <- PbPmanipulation(PbP.BDB)
```

UNIVERSITY OF BRESCIA

BIG DATA ANALYTICS IN SPORTS

Data

```
data(package="BasketballAnalyzeR")
```

1. **Teams box scores**. In this data frame, called `Tbox`, the cases (rows) are the analyzed teams and the variables (columns) are referred to the team achievements in the considered games.

2. **Opponents box scores**. In this data frame, called `Obox`, the cases (rows) are the analyzed teams and the variables (columns) are referred to the achievements of the opponents of the in the considered games.

3. **Players box scores**. In this data frame, called `Pbox`, the cases (rows) are the analyzed players and the variables (columns) are referred to the individual achievements in the considered games.

Data

```
data(package="BasketballAnalyzeR")
```

4. **Play-by-play data**. In this data frame, called `PbP.BDB`, the cases (rows) are the events occurred during the analyzed games and the variables (columns) are descriptions of the events in terms of type, time, players involved, score, area of the court.

5. **Additional information**. In this data frame, called `Tadd`, the cases (rows) are the analyzed teams and the variables (columns) are qualitative information such as Conference, Division, final rank, qualification to Playoffs.

Data

Boxscores (1., 2., 3.) and Additional information (5) are about all the teams and players of the **82 games** in the regular season of the **NBA championship 2017/2018**

Play-by-play data are relative to the 82 games played by **Golden State Warriors** (the champions) during the regular season (data made available by **BigDataBall**, www.bigdataball.com)

18/19 NBA boxscores and play-by-play data of **Cleveland Cavaliers (17/18)** are available at

https://bdsports.unibs.it/basketballanalyzer/

UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

Data

```
data(package="BasketballAnalyzeR")
```

Variable	Description	Tbox	Obox	Pbox	Tadd
Team	Analyzed team (long name)	×	×	×	×
team	Analyzed team (short name)				×
Conference	Conference				×
Division	Division				×
Rank	Rank (end season)				×
Playoff	Playoff qualification (Yes or No)				×
Player	Analyzed player			×	
GP	Games Played	×	×	×	
MIN	Minutes Played	×	×	×	
PTS	Points Made	×	×	×	
W	Games won	×	×		
L	Games lost	×	×		
P2M	2-Point Field Goals (Made)	×	×	×	
P2A	2-Point Field Goals (Attempted)	×	×	×	
P2p	2-Point Field Goals (Percentage)	×	×	×	
P3M	3-Point Field Goals (Made)	×	×	×	
P3A	3-Point Field Goals (Attempted)	×	×	×	
P3p	3-Point Field Goals (Percentage)	×	×	×	
FTM	Free Throws (Made)	×	×	×	
FTA	Free Throws (Attempted)	×	×	×	
FTp	Free Throws (Percentage)	×	×	×	
OREB	Offensive Rebounds	×	×	×	
DREB	Defensive Rebounds	×	×	×	
AST	Assists	×	×	×	
TOV	Turnovers	×	×	×	
STL	Steals	×	×	×	
BLK	Blocks	×	×	×	
PF	Personal Fouls	×	×	×	
PM	Plus/Minus	×	×	×	

Variable	Description
game_id	Identification code for the game
data_set	Season: years and type (Regular or Playoffs)
date	Date of the game
a1 ... a5; h1 ... h5	Five players on the court (away team; home team)
period	Quarter (≥ 5: over-time)
away_score; home_score	Score of the away/home team
remaining_time	Time left in the quarter (h:mm:ss)
elapsed	Time played in the quarter (h:mm:ss)
play_length	Time since the immediately preceding event (h:mm:ss)
play_id	Identification code for the play
team	Team responsible for the event
event_type	Type of event
assist	Player who made the assist
away; home	Players for the jump ball
block	Player who blocked the shot
entered; left	Player who entered/left the court
num	Sequence number of the free throw
opponent	Player who made the foul
outof	Number of free throws accorded
player	Player responsible for the event
points	Scored points
possession	Player who the jump ball is tipped to
reason	Reason of the turnover
result	Result of the shot (made or missed)
steal	Player who stole the ball
type	Type of play
shot_distance	Field shots: distance from the basket
original_x ; original_y; converted_x ; converted_y	coordinates of the shooting player original: tracking coordinate system half court, (0,0) center of the basket converted: coordinates in feet full court, (0,0) bottom-left corner
description	Textual description of the event

Variable	Description
periodTime	Time played in the quarter (in seconds)
totalTime	Time played in the match (in seconds)
playlength	Time since the immediately preceding event (in seconds)
ShotType	Type of shot (FT, 2P, 3P)
oppTeam	Name of the opponent team

UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

R script

bdsports.unibs.it/basketballanalyzer/

```
##########################################################################
# July 2019
#
# The following R code allows to replicate all the analyses and examples
# in the book "Basketball Data Science" (by P. Zuccolotto and M. Manisera),
# forthcoming as a CRC Press publication.
# It is based on the "BasketballAnalyzeR" package developed with M. Sandri.
# See
# https://bdsports.unibs.it/basketballanalyzer/
# for further explanations and updates
##########################################################################

rm(list=ls())
# install.packages("devtools", repos="https://cran.stat.unipd.it/")
# devtools::install_github("sndmrc/BasketballAnalyzeR",force=TRUE)
library(BasketballAnalyzeR)

###########################################
###########################################
# CHAPTER 2                               #
# Data and Basic Statistical Analyses     #
###########################################
###########################################
#data(package="BasketballAnalyzeR")
#PbP <- PbPmanipulation(PbP.BDB)

#####################
# 2.2 BASIC STATISTICAL ANALYSES
#####################

#####################
# 2.2.1 Pace, Ratings, Four Factors
#####################
rm(list=ls())

tm <- c("BOS","CLE","GSW","HOU")
selTeams <- which(Tadd$team %in% tm)
FF.sel <- fourfactors(Tbox[selTeams,], Obox[selTeams,])

plot(FF.sel)
```

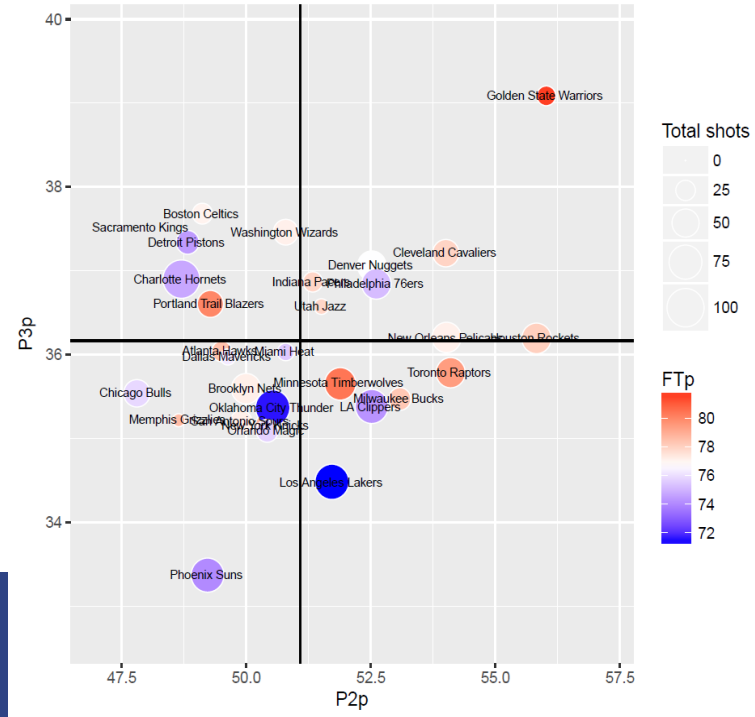
UNIVERSITY OF BRESCIA

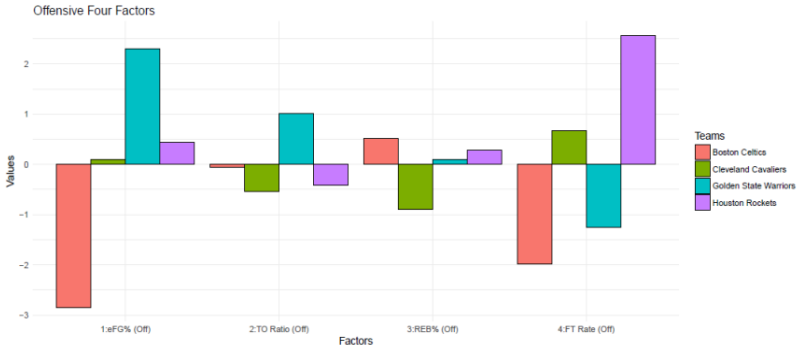
BDsports Big Data analytics in sports

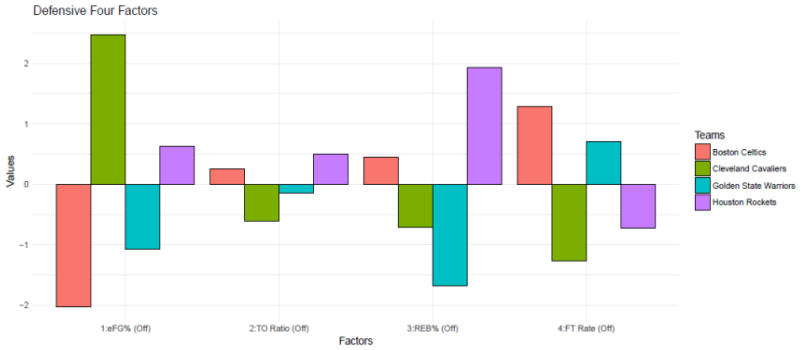
CRC Press
Taylor & Francis Group
CHAPMAN & HALL

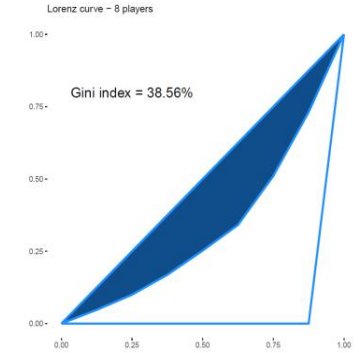
ASA
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

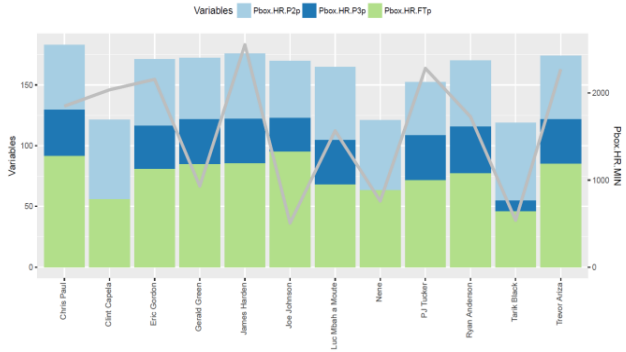
- **Basic Statistical Analyses**
- Discovering patterns in data
- Finding groups in data
- Modelling relationships in data

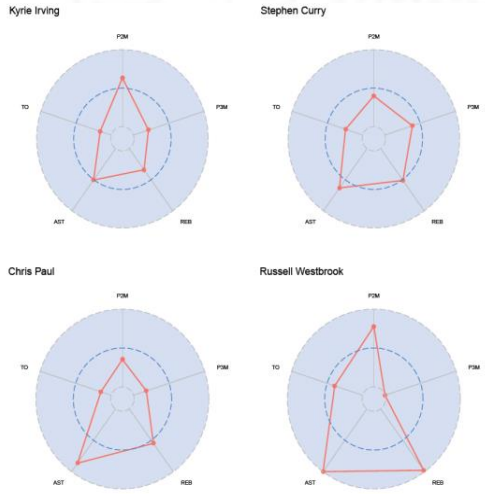












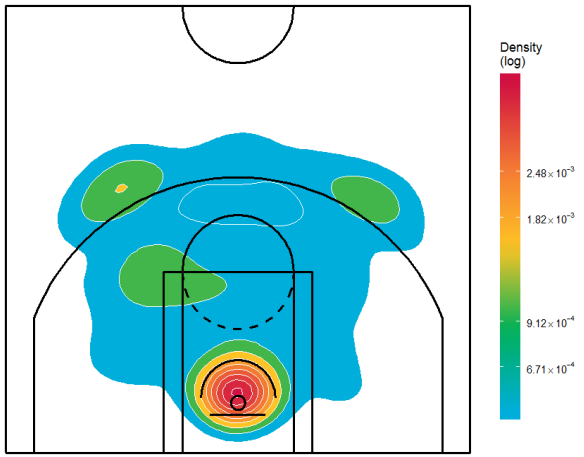
UNIVERSITY OF BRESCIA

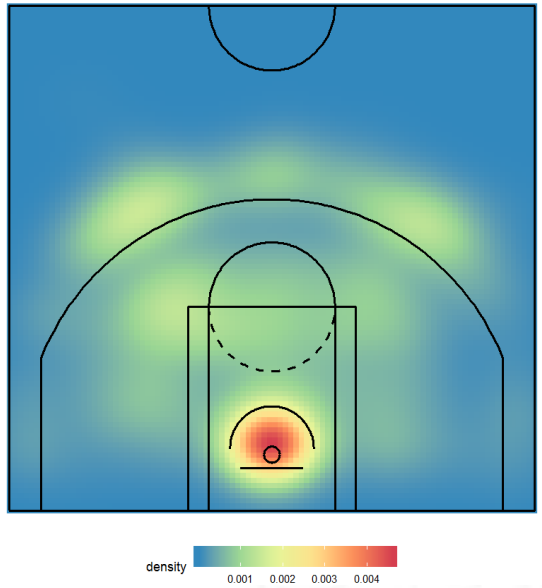
BDsports
Big Data analytics in sports

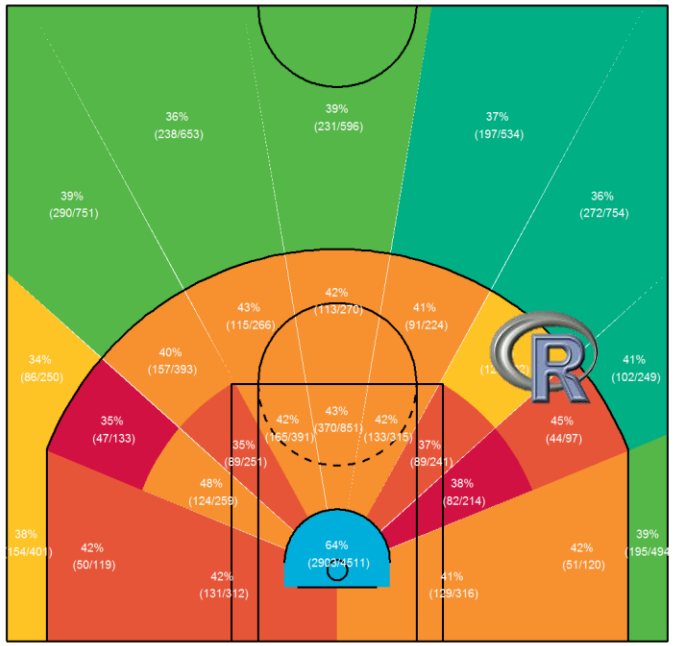
CRC Press
Taylor & Francis Group
CHAPMAN & HALL

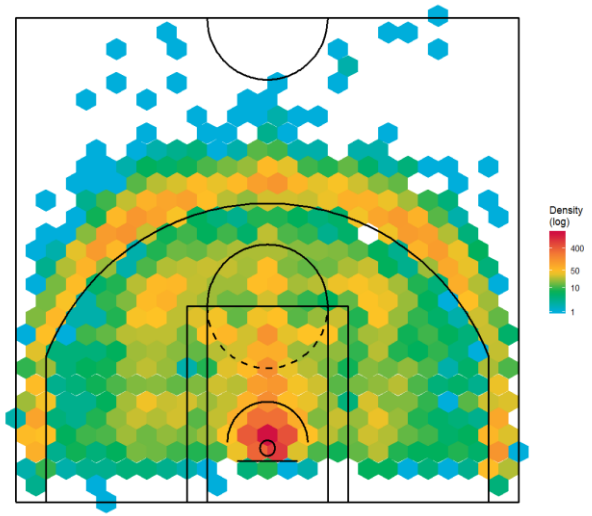
ASA
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

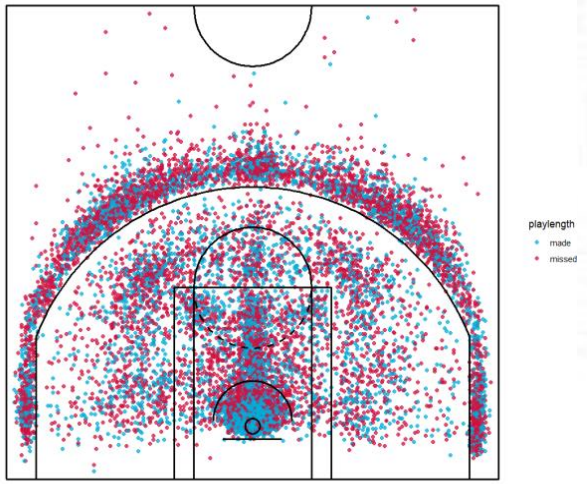
- **Basic Statistical Analyses**
- Discovering patterns in data
- Finding groups in data
- Modelling relationships in data











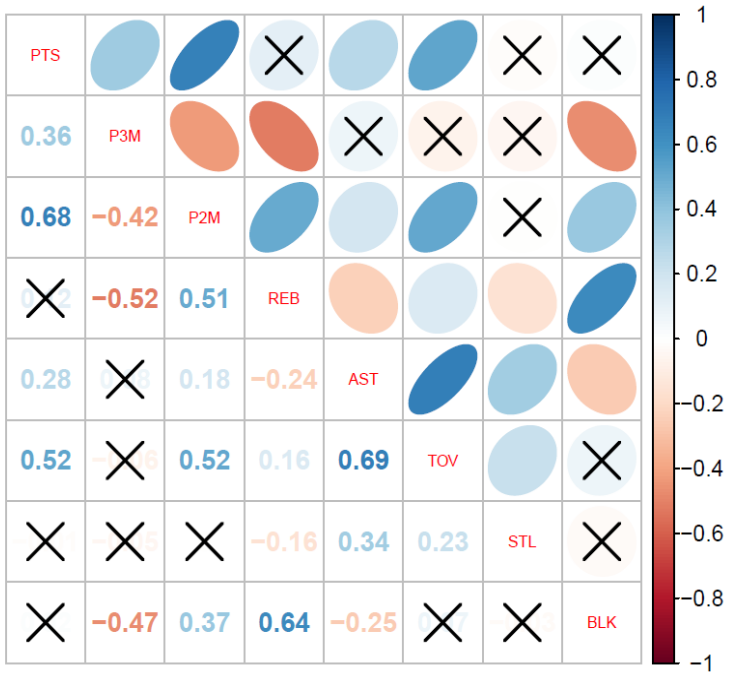
UNIVERSITY OF BRESCIA

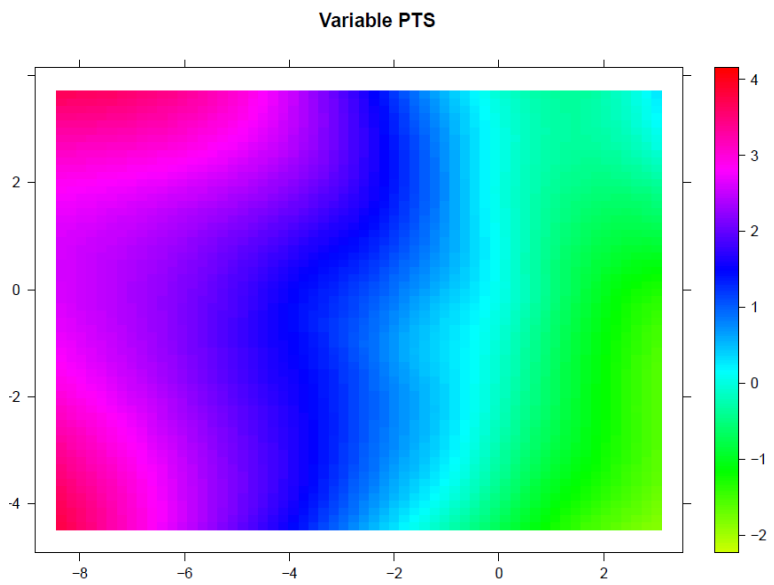
ORTS
BIG DATA ANALYTICS IN SPORTS

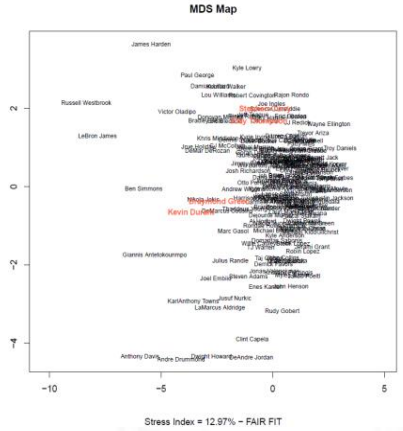
CRC Press
Taylor & Francis Group
CHAPMAN & HALL

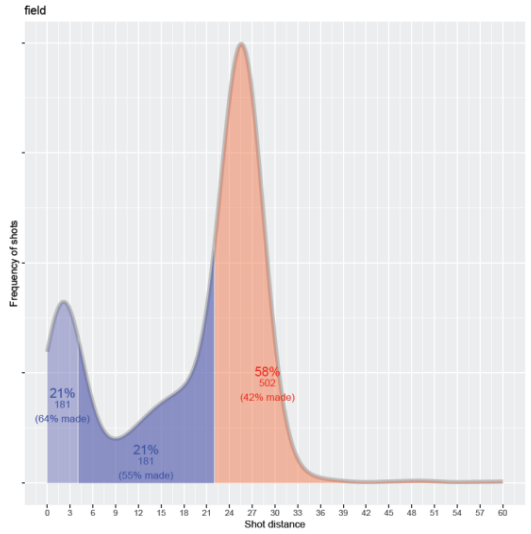
ASA
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

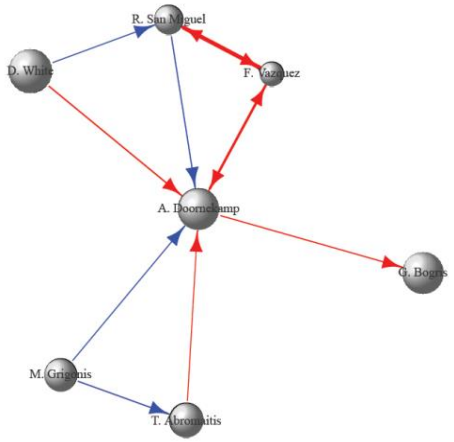
- Basic Statistical Analyses
- **Discovering patterns in data**
- Finding groups in data
- Modelling relationships in data











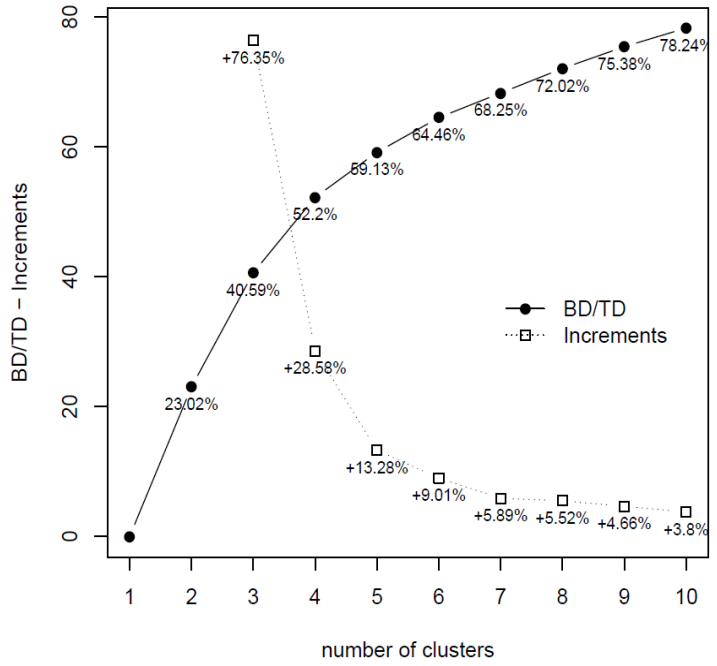
UNIVERSITY OF BRESCIA

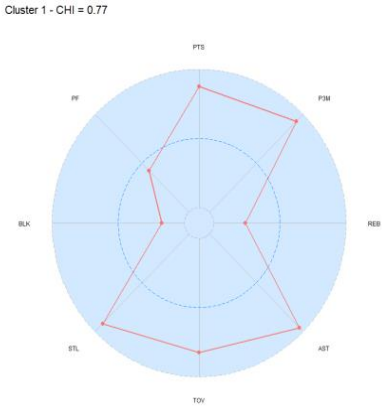
BDsports
Big Data analytics in sports

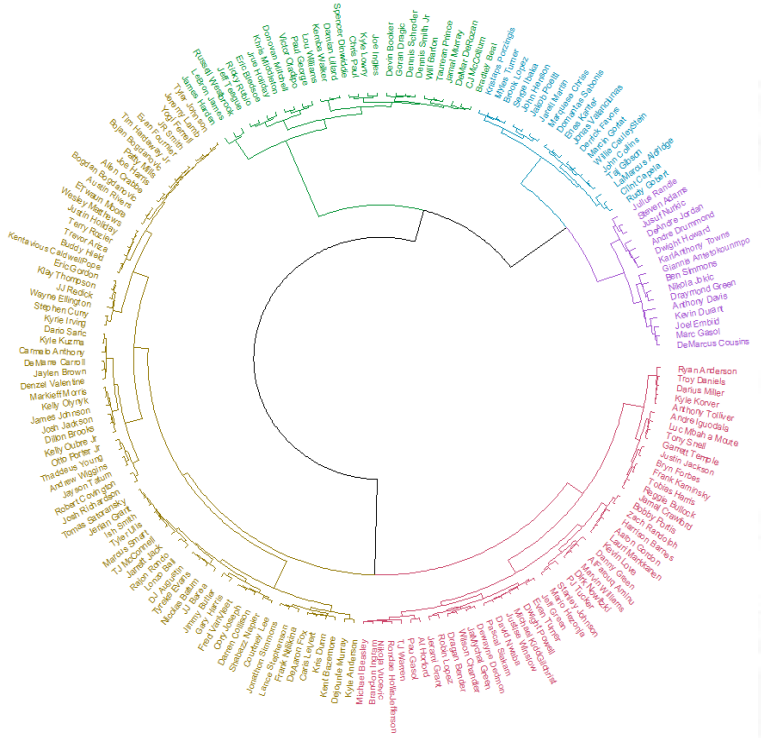
CRC Press
Taylor & Francis Group
CHAPMAN & HALL

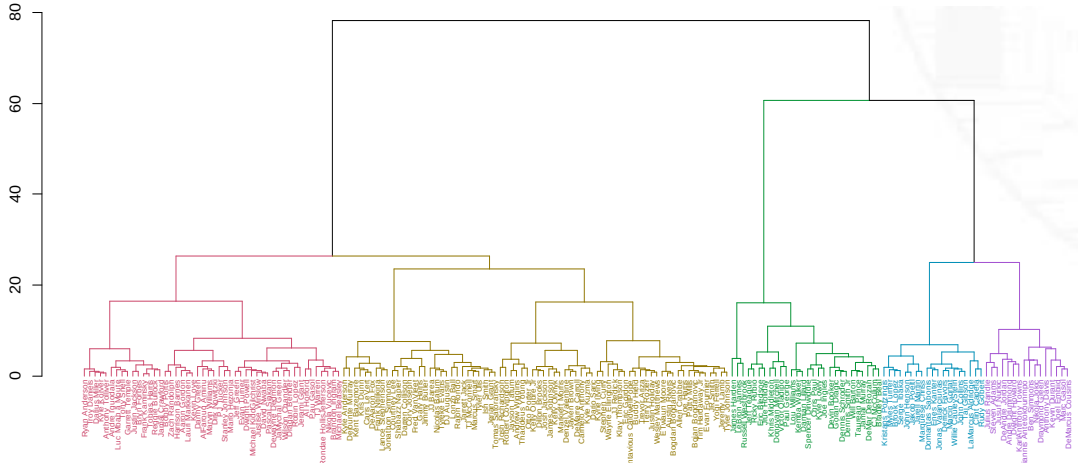
ASA
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

- Basic Statistical Analyses
- Discovering patterns in data
- **Finding groups in data**
- Modelling relationships in data





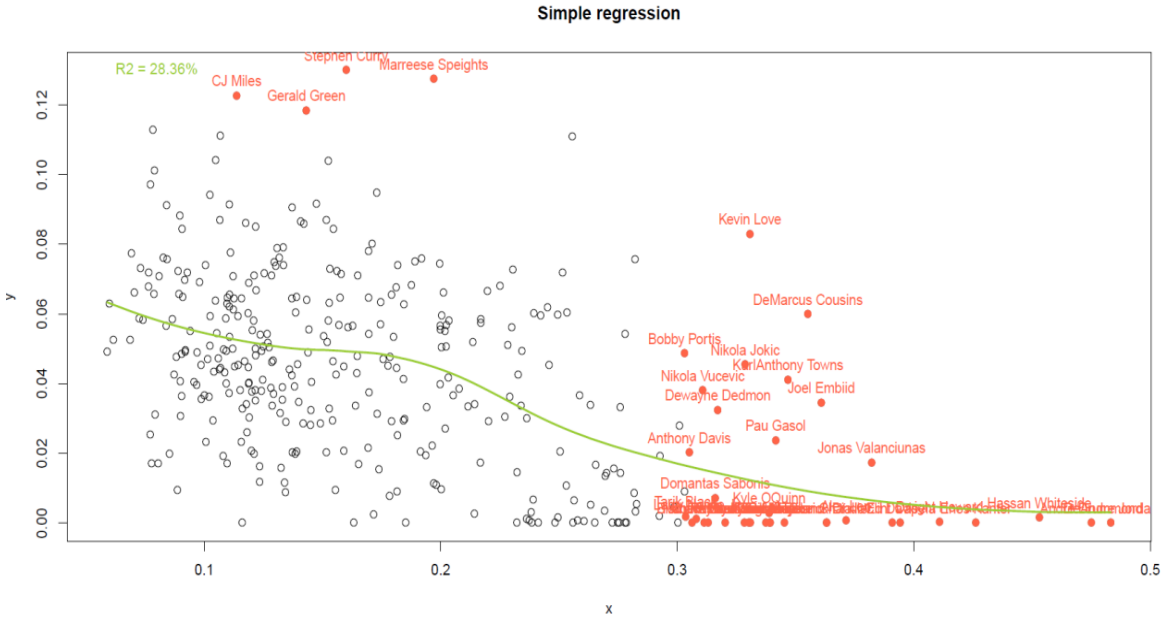


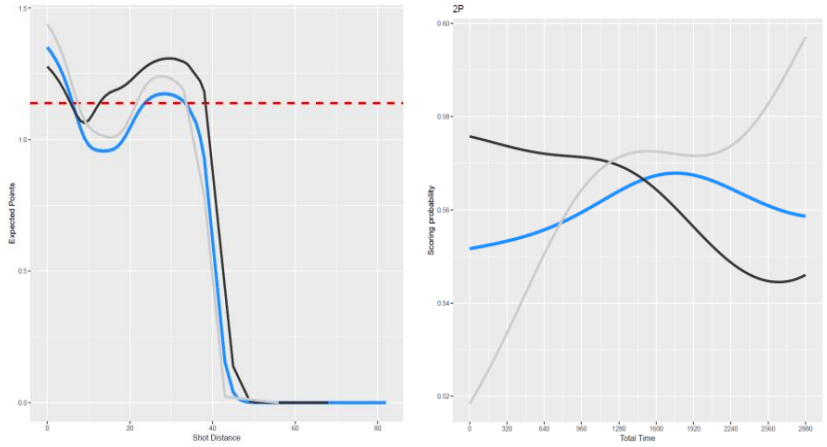


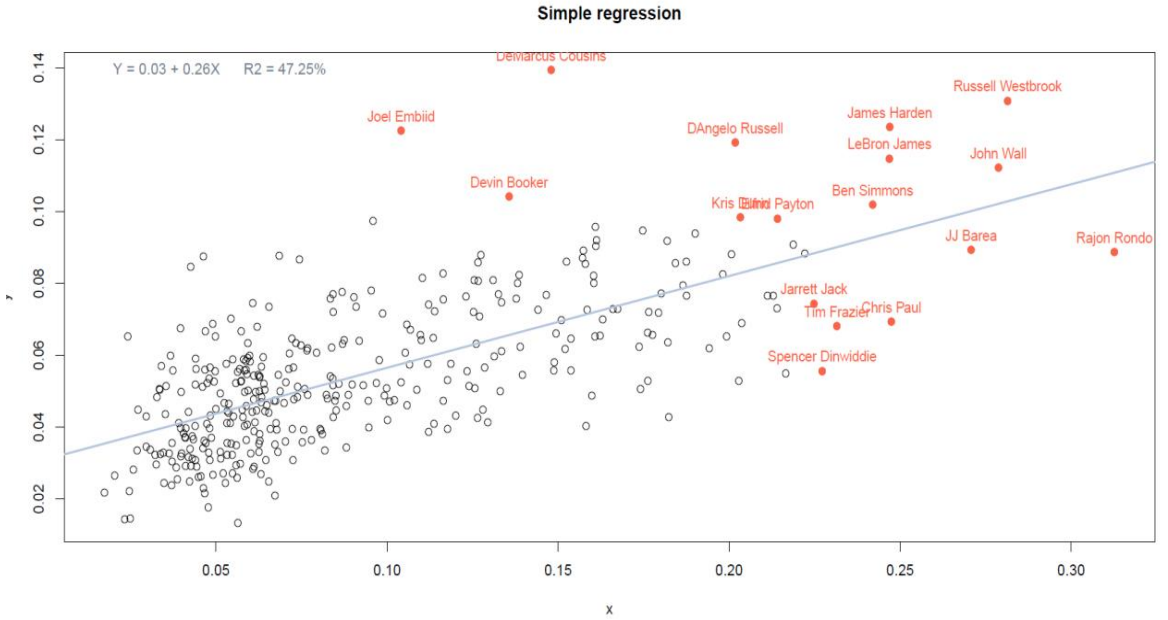
UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports

- Basic Statistical Analyses
- Discovering patterns in data
- Finding groups in data
- **Modelling relationships in data**







UNIVERSITY OF BRESCIA

BDsports
Big Data analytics in sports