



ISI SPECIAL INTEREST GROUP ON SPORTS STATISTICS



# Ambra Macis



# Big Data analytics in sports

[bdsports.unibs.it](http://bdsports.unibs.it)

UNIVERSITY OF BRESCIA

Home Team Research Analytics Teaching Dissemination BasketballAnalyzeR

**BDsp<sup>o</sup>RTS**

**Big Data analytics in sports**

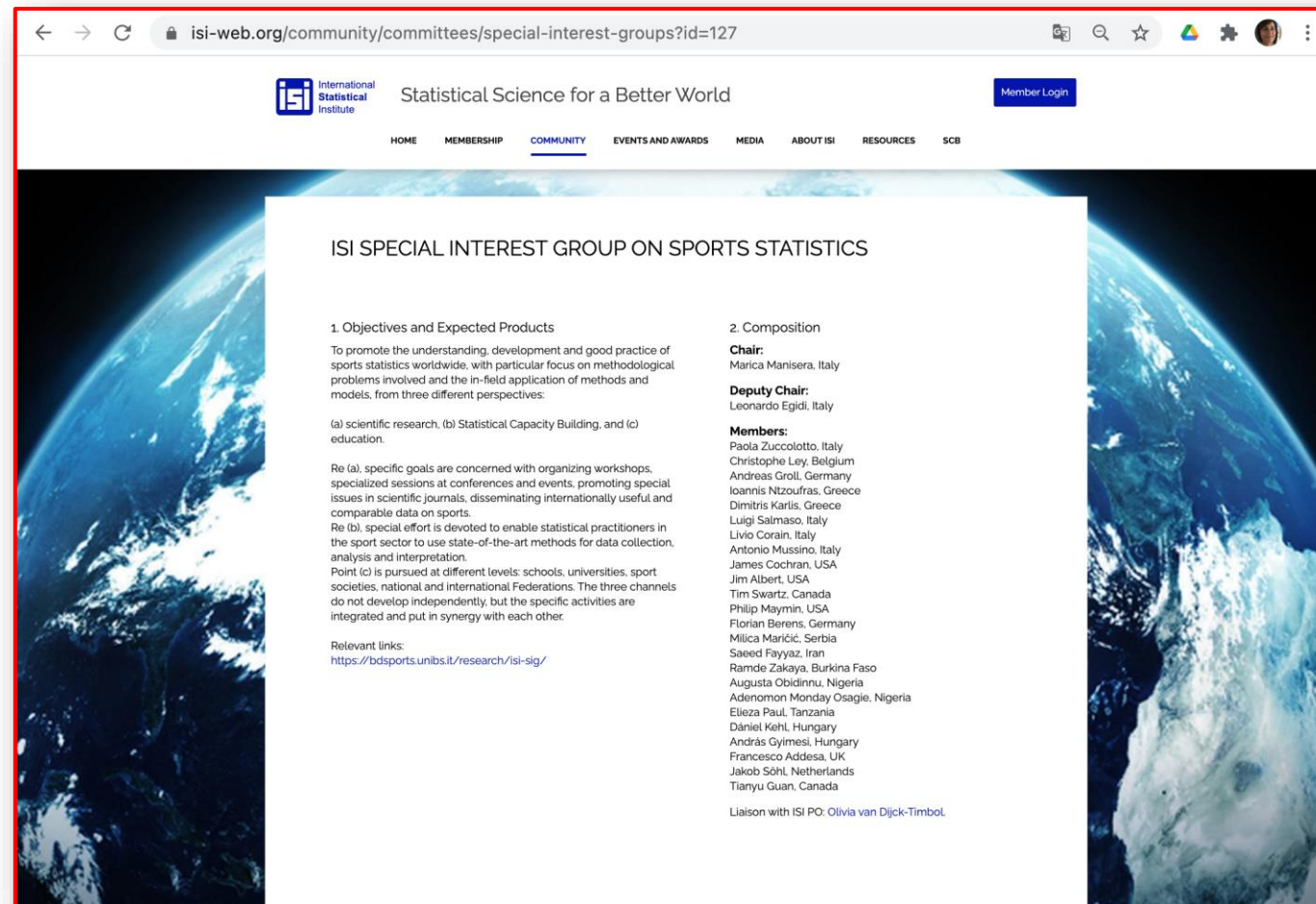
Scientific coordinators of the project: Paola Zuccolotto and Marica Manisera

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

Description of the Project Our partners

# ISI Special Interest Group on SPORTS STATISTICS

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



The screenshot shows a web browser displaying the ISI website. The URL in the address bar is <https://www.isi-web.org/community/committees/special-interest-groups?id=127>. The page header includes the ISI logo, the tagline "Statistical Science for a Better World", and a "Member Login" button. A navigation menu contains links for HOME, MEMBERSHIP, COMMUNITY (which is highlighted), EVENTS AND AWARDS, MEDIA, ABOUT ISI, RESOURCES, and SCB. The main content area features a large background image of Earth from space. The title "ISI SPECIAL INTEREST GROUP ON SPORTS STATISTICS" is centered at the top of the content. Below the title, there are two columns of text. The left column is titled "1. Objectives and Expected Products" and contains several paragraphs describing the group's goals and activities. The right column is titled "2. Composition" and lists the Chair, Deputy Chair, and Members of the group. At the bottom of the right column, it mentions the Liaison with ISI PO.

## ISI SPECIAL INTEREST GROUP ON SPORTS STATISTICS

**1. Objectives and Expected Products**

To promote the understanding, development and good practice of sports statistics worldwide, with particular focus on methodological problems involved and the in-field application of methods and models, from three different perspectives:

(a) scientific research, (b) Statistical Capacity Building, and (c) education.

Re (a), specific goals are concerned with organizing workshops, specialized sessions at conferences and events, promoting special issues in scientific journals, disseminating internationally useful and comparable data on sports.

Re (b), special effort is devoted to enable statistical practitioners in the sport sector to use state-of-the-art methods for data collection, analysis and interpretation.

Point (c) is pursued at different levels: schools, universities, sport societies, national and international Federations. The three channels do not develop independently, but the specific activities are integrated and put in synergy with each other.

Relevant links:  
<https://bdsports.unibs.it/research/isi-sig/>

**2. Composition**

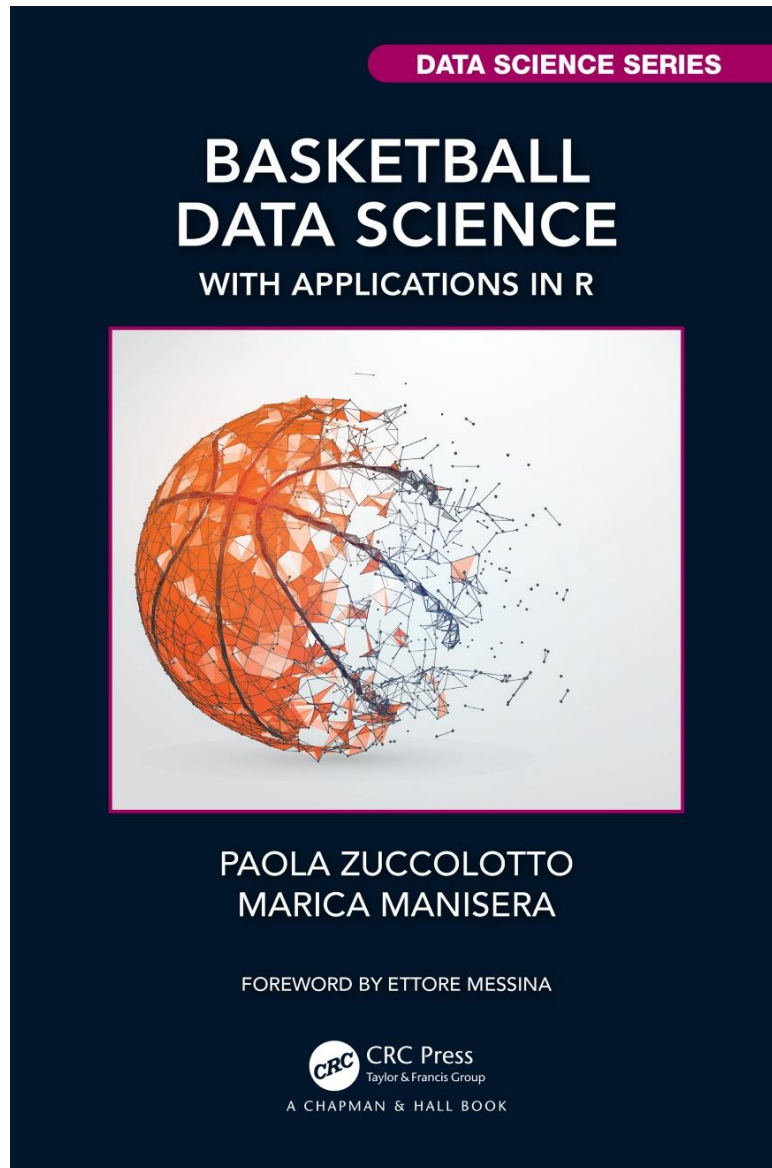
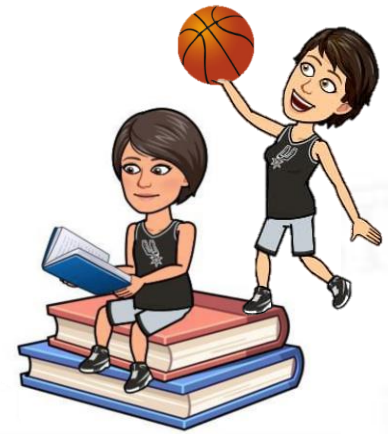
**Chair:**  
Marica Manisera, Italy

**Deputy Chair:**  
Leonardo Egidi, Italy

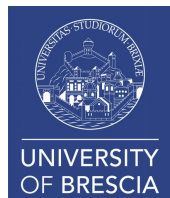
**Members:**  
Paola Zuccolotto, Italy  
Christophe Ley, Belgium  
Andreas Groll, Germany  
Ioannis Ntzoufras, Greece  
Dimitris Karlis, Greece  
Luigi Salmaso, Italy  
Livio Corain, Italy  
Antonio Mussino, Italy  
James Cochran, USA  
Jim Albert, USA  
Tim Swartz, Canada  
Philip Maymin, USA  
Florian Berens, Germany  
Milica Maričić, Serbia  
Saeed Fayyaz, Iran  
Ramde Zakaya, Burkina Faso  
Augusta Obidinnu, Nigeria  
Adenomom Monday Osagie, Nigeria  
Eleza Paul, Tanzania  
Dániel Kehl, Hungary  
András Gyimesi, Hungary  
Francesco Addesa, UK  
Jakob Söhl, Netherlands  
Tianyu Guan, Canada

Liaison with ISI PO: Olivia van Dijk-Timbol

# The Book



 Package  
developed by Marco Sandri



# 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



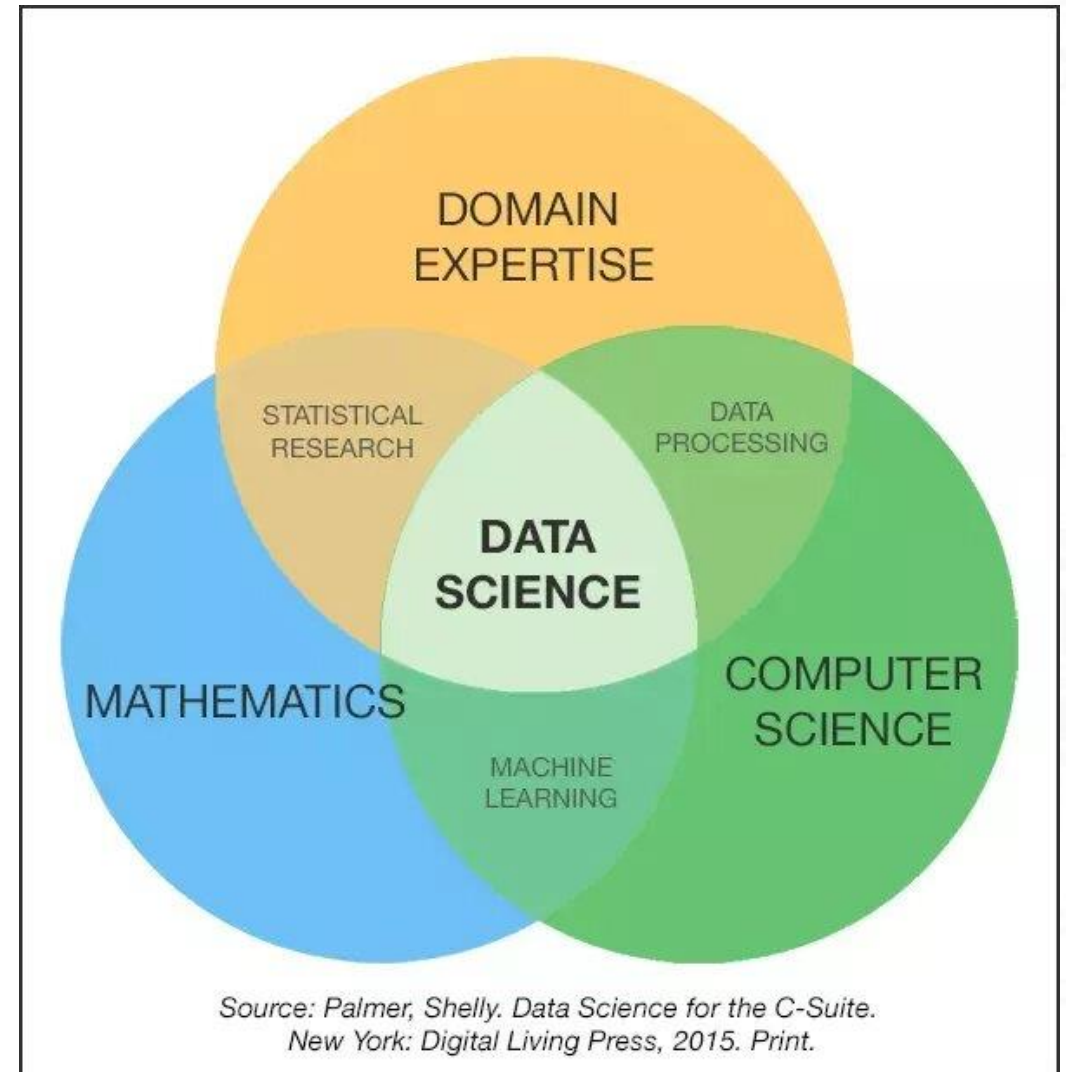


# What is Data Science?

Discipline aimed at **extracting knowledge** from data in various forms

Multidisciplinary

Applicable to a wide range of fields



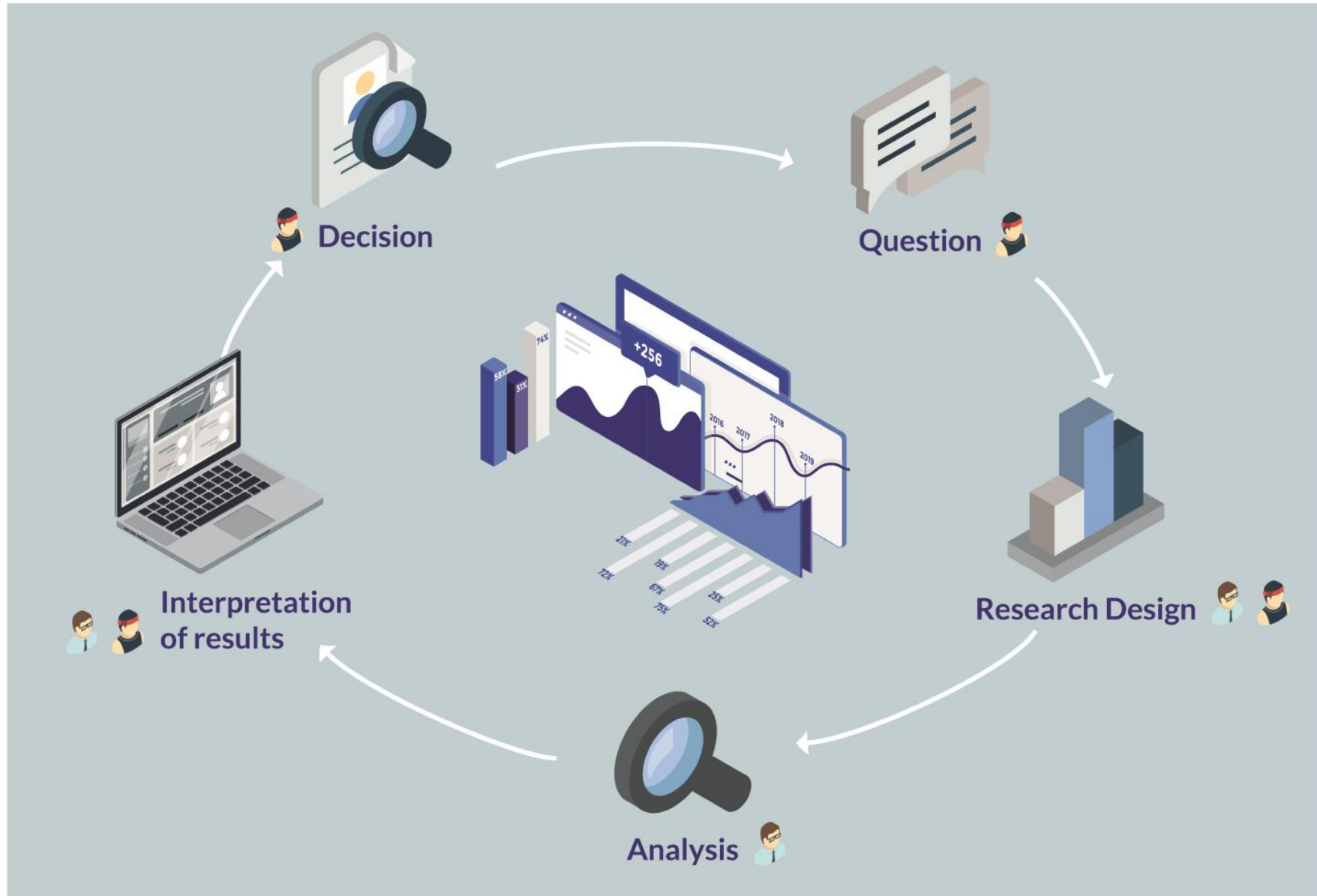
# Data Science...

- ... aims 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



# 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**



# 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

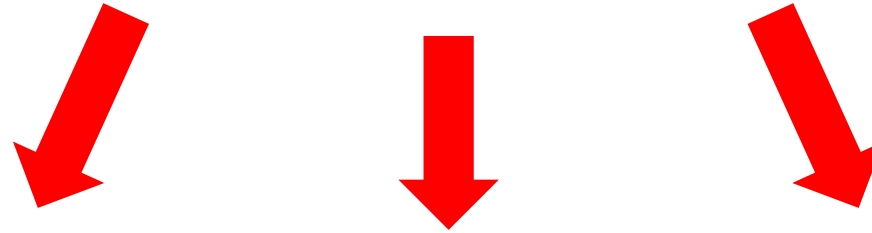
(2021)

“...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**”



# 2 – Basketball Analytics: state of the art

# Basketball Analytics



Official  
Statistics

Sport Analytics  
Services

Scientific  
Research

TEAMS STATISTICS

LEADERS GAMEWIDE

TEAM LEADERS

Points per Game Overall

#	Team	GP	W	WPC	PPG	PTS	FGM-FGA	FG%	3PM-3PA	3P%	FTM-FTA	FT%
1.	USA	8	8	100.0	100.9	807	34.0-73.9	47.1	10.4-28.1	36.9	21.0-28.0	75.0
2.	Argentina	8	6	75.0	86.5	519	28.7-71.7	41.4	11.3-32.3	35.1	15.0-23.0	65.0
3.	Spain	8	6	75.0	86.1	689	32.4-69.9	46.3	8.0-28.9	27.6	19.0-21.0	74.0
4.	Australia	8	5	62.5	85.4	602	31.0-64.3	48.4	7.4-22.3	33.1	14.3-19.4	74.0
5.	France	8	5	62.5	83.1	602	29.0-63.7	45.7	7.4-22.3	33.1	16.3-21.4	77.0
6.	Brazil	8	2	25.0	82.2	611	29.0-69.4	41.8	6.4-21.6	29.6	16.2-22.2	73.0
7.	Hong Kong	8	2	25.0	81.7	490	32.7-60.2	53.8	3.0-12.2	24.6	14.0-18.0	77.0
8.	China	8	2	25.0	80.9	493	34.0-69.0	49.3	8.0-23.0	34.8	18.0-24.0	74.0
9.	Hong Kong	8	2	25.0	79.4	390	27.4-63.2	43.7	10.0-27.4	36.7	13.0-17.0	76.0
10.	Lithuania	8	2	25.0	79.0	406	29.0-62.2	46.6	7.0-22.3	31.6	10.0-14.0	71.0
11.	China	8	2	25.0	68.4	318	23.0-52.9	43.7	3.0-11.4	26.7	18.0-22.4	79.0
12.	Venezuela	8	2	25.0	63.0	313	22.4-63.4	37.0	6.0-20.4	29.5	13.4-17.4	76.0

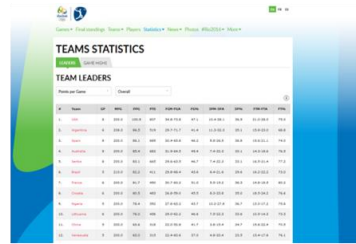




# Basketball Analytics



Official  
Statistics



Sport  
Analytics  
Services



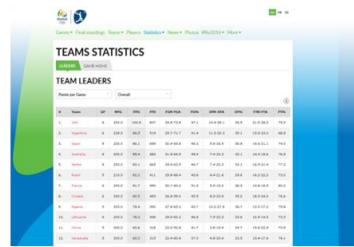
Scientific  
Research



# Basketball Analytics



Official  
Statistics



Sport  
Analytics  
Services



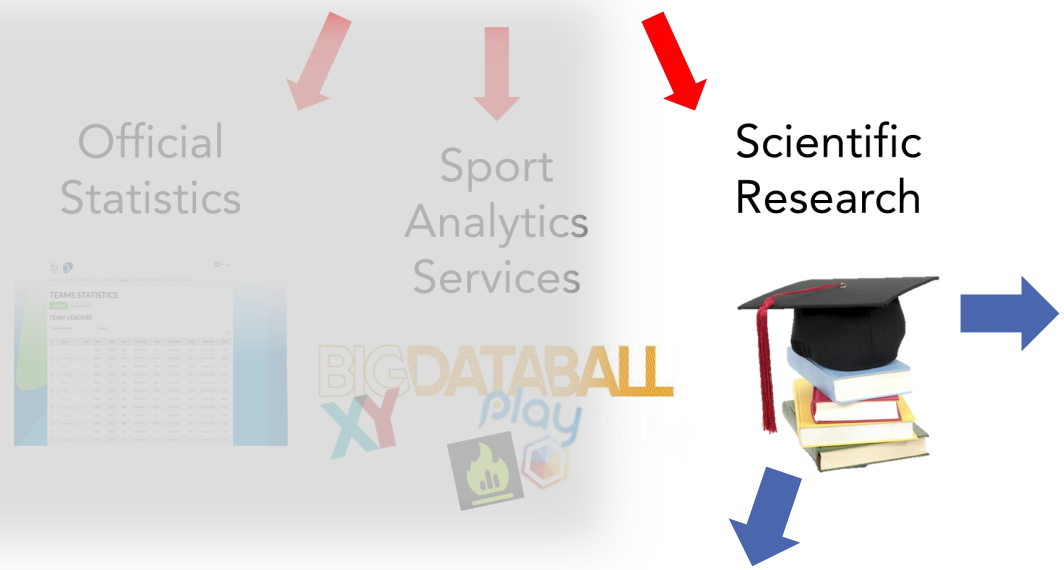
Scientific  
Research



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



# Basketball Analytics



## Scientific literature

## Scientific journals



## Special Issues



# 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 increasing 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, ...).

# 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

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

Splits

TOTAL SPLITS

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	13.0
Andre Iguodala, SF	51	10	9.5	6.2	0.7	3.2	3.9	3.4	0.93	0.39	0.7	1.2	4.9	23.7
Javale McGee, C	54	0	25.6	6.4	0.7	2.0	3.0	0.2	0.24	0.69	0.6	1.5	0.4	15.9
Zaza Pachulia, C	51	10	9.5	6.2	1.1	2.0	3.0	0.2	0.95	0.41	1.3	2.4	1.6	10.1
Shaun Livingston, PG	44	44	18.8	5.9	2.1	3.9	6.0	2.0	0.47	0.22	0.8	1.6	2.1	16.6
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	1.3	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
<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>	<b>31.0</b>	<b>9.55</b>	<b>6.57</b>	<b>14.4</b>	<b>19.2</b>	<b>2.2</b>	<b>--</b>



Stats

[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 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	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	32.1	10.2	1.4	6.8	8.8	7.3	2.09	1.51	2.3	3.0	3.2	17.2
Ian Clark, SG	51	0	14.3	0.0	0.0	1.2	1.4	1.3	0.55	0.12	0.6	1.0	1.9	13.5
Andre Iguodala, SF	54	0	25.4	0.0	0.2	3.2	3.1	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	51	1	17.3	5.2	0.4	1.6	1.9	1.7	0.47	0.20	2.9	2.2	2.2	24.0
Shaun Livingston, PG	44	44	18.8	5.9	2.1	3.9	6.0	2.0	0.95	0.20	2.9	2.2	2.2	24.0
David West, PF	51	1	17.3	5.2	0.4	1.6	1.9	1.7	0.47	0.20	2.9	2.2	2.2	24.0
Patrick McCaw, PG	46	3	12.4	3.3	0.2	0.8	1.0	1.0	0.35	0.20	2.9	2.2	2.2	24.0
Kevon Looney, SF	43	3	9.3	2.9	0.8	1.7	2.6	0.6	0.30	0.20	2.9	2.2	2.2	24.0
James Michael McAdoo, SF	29	0	9.1	2.8	0.4	1.1	1.5	0.4	0.24	0.20	2.9	2.2	2.2	24.0
Briante Weber, PG	5	0	8.2	1.6	0.0	0.6	0.6	1.0	0.40	0.20	2.9	2.2	2.2	24.0
Damian Jones, C	8	0	5.6	1.4	0.8	0.6	1.4	0.0	0.13	0.20	2.9	2.2	2.2	24.0
Anderson Varejao, C	14	1	6.6	1.3	0.9	1.1	1.9	0.7	0.21	0.20	2.9	2.2	2.2	24.0
Totals	56	--	--	118.2	8.8	35.7	44.5	31.0	9.5					

Stats

play-by-play

NBA ALL-STAR GAME

Final

	1	2	3	4	T
TNT	1	2	3	4	T
EAST	53	36	47	43	182
WEST	48	46	47	48	192

Eastern Conf All-Stars 182 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 dblTemp;  
bool again = true;  
  
while (again) {  
    iN = -1;  
    again = false;  
    getline(cin, input);  
    stringstream(input) >> dblTemp;  
    length = input.length();  
    if (length < 4) {  
        again = true;  
    }  
}
```

WEB SCRAPING SOFTWARE





# Basketball Data

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

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	54	54	34.4	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	32.1	16.2	1.4	3.8	6.8	7.3	2.09	1.51	2.3	3.0	3.2	17.2
Draymond Green, PF	51	0	14.3	7.0	0.0	1.2	1.4	1.3	0.55	0.12	0.6	1.0	1.9	13.0
Ian Clark, SG	54	0	25.4	10.2	0.0	3.2	3.1	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	1.1	2.4	1.6	15.9
Javale McGee, C	44	44	18.8	5.9	2.1	3.9	6.0	2.0	0.95	1.1	1.1	1.0	1.0	0.35
Zaza Pachulia, C	51	1	17.3	5.2	0.4	1.6	1.9	1.7	0.47	0.4	2.7	2.1	0.67	0.67
Shaun Livingston, PG	42	0	11.5	4.1	0.6	2.1	2.7	2.1	0.67	1.0	1.0	1.0	1.0	0.35
David West, PF	46	3	12.4	3.3	0.2	0.8	1.0	1.0	0.35	0.6	0.6	0.6	0.30	0.30
Patrick McCaw, PG	43	3	9.3	2.9	0.8	1.7	2.6	0.6	0.24	0.6	0.6	0.6	0.24	0.24
Kevon Looney, SF	29	0	9.1	2.8	0.4	1.1	1.5	0.4	0.40	0.6	0.6	1.0	0.40	0.40
James Michael McAdoo, SF	5	0	8.2	1.6	0.0	0.6	0.6	1.0	0.13	0.6	0.6	1.0	0.13	0.13
Briante Weber, PG	8	0	5.6	1.4	0.8	0.6	1.4	0.0	0.21	1.9	1.9	0.7	0.21	0.21
Damian Jones, C	14	1	6.6	1.3	0.9	1.1	1.9	0.7	0.21	1.9	1.9	0.7	0.21	0.21
Anderson Varejao, C	14	1	6.6	1.3	0.9	1.1	1.9	0.7	0.21	1.9	1.9	0.7	0.21	0.21
<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>	<b>31.0</b>	<b>9.5</b>					

Stats



Sensor data



NBA ALL-STAR GAME

Final

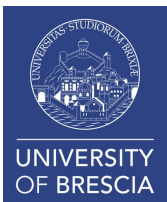
	1	2	3	4	T
TNT	53	36	47	43	182
WEST	48	46	47	48	192

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

play-by-play



# 4 - Introduction to the R package BasketballAnalyzeR

# Book and codes

## Contents

---

### PART I Basketball Analytics Fundamentals

<b>CHAPTER 1 ■ Introduction</b>	<b>3</b>
1.1 WHAT IS DATA SCIENCE?	4
1.1.1 Knowledge representation	5
1.1.2 A tool for decisions and not a substitute for human intelligence	6
1.2 DATA SCIENCE IN BASKETBALL	10
1.3 HOW THE BOOK IS STRUCTURED	14

<b>CHAPTER 2 ■ Data and Basic Statistical Analyses</b>	<b>19</b>
2.1 BASKETBALL DATA	20
2.2 BASIC STATISTICAL ANALYSES	26
2.2.1 Pace, Ratings, Four Factors	26
2.2.2 Bar-line plots	28
2.2.3 Radial plots	33
2.2.4 Scatter plots	35
2.2.5 Bubble plots	40
2.2.6 Variability analysis	43
2.2.7 Inequality analysis	48
2.2.8 Shot charts	53

### PART II Advanced Methods

<b>CHAPTER 3 ■ Discovering Patterns in Data</b>	<b>61</b>
3.1 DETECTING ASSOCIATIONS BETWEEN VARIABLES	62
3.1.1 Statistical dependence	63
3.1.2 Mean dependence	66
3.1.3 Correlation	68
3.2 ANALYZING PAIRWISE LINEAR CORRELATION AMONG VARIABLES	72
3.3 DISPLAYING INDIVIDUAL CASES ACCORDING TO THEIR SIMILARITY	76
3.4 ANALYZING NETWORK RELATIONSHIPS	82
3.5 ESTIMATING DENSITY OF EVENTS	95
3.5.1 Density with respect to a concurrent variable	95
3.5.2 Density in space	100
3.5.3 Joint density of two variables	102
3.6 FOCUS: THE EFFECT OF SHOOTING UNDER HIGH-PRESSURE CONDITIONS	104


<b>CHAPTER 4 ■ Finding Groups in Data</b>	<b>117</b>
4.1 CLUSTER ANALYSIS	118
4.1.1 <i>k</i> -means clustering	121
4.1.2 Agglomerative hierarchical clustering	123
4.2 CLUSTERING IN BASKETBALL	125
4.2.1 <i>k</i> -means clustering of NBA teams	126
4.2.2 <i>k</i> -means clustering of Golden State Warriors shots	134
4.2.3 Hierarchical clustering of NBA players	144
4.3 FOCUS: NEW ROLES IN BASKETBALL	150

<b>CHAPTER 5 ■ Modeling Relationships in Data</b>	<b>157</b>
5.1 LINEAR MODELS	161
5.1.1 Simple linear regression model	162
5.2 NONPARAMETRIC REGRESSION	164
5.2.1 Polynomial local regression	166
5.2.2 Gaussian kernel smoothing	170
5.2.2.1 Estimation of scoring probability	172

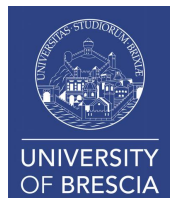
### PART III Technical Appendix

<b>CHAPTER 6 ■ The R package BasketballAnalyzeR</b>	<b>191</b>
6.1 INTRODUCTION	191
6.2 PREPARING DATA	193
6.3 CUSTOMIZING PLOTS	196
6.4 BUILDING INTERACTIVE GRAPHICS	200
<b>Index</b>	<b>223</b>

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



The screenshot shows the website for BasketballAnalyzer. At the top is a blue navigation bar with the University of Brescia logo and name on the left, and menu items: Home, Team, Research, Analytics, Teaching, Dissemination, and BasketballAnalyzer. The main content area has a large blue heading "BasketballAnalyzer". To the left is a hexagonal logo with a basketball and the text "BasketballAnalyzer" and "bdsports.unibs.it/BasketballAnalyzer". To the right, text describes the R package and its association with the book "Basketball Data Science - With Applications in R" by P. Zuccolotto and M. Manisera. Below this, it mentions the developers and the BODaI-Lab. A section for "Supplementary material for the book" includes a "pdf format" button. On the right, there is a book cover for "Basketball Data Science" with a network visualization of a basketball.



# 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("BasketballAnalyzeR")` 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*

# 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](http://www.bigdataball.com))

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



# 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](http://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/>

# Data

data (package="BasketballAnalyzeR")

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

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 ( $\geq 5$ : 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;	coordinates of the shooting player
converted_x ; converted_y	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

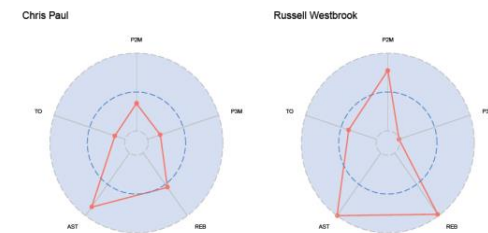
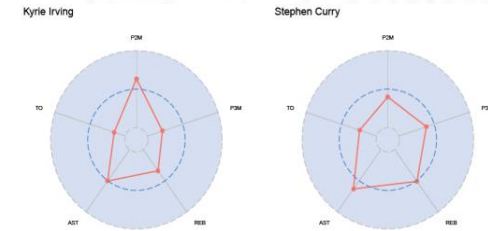
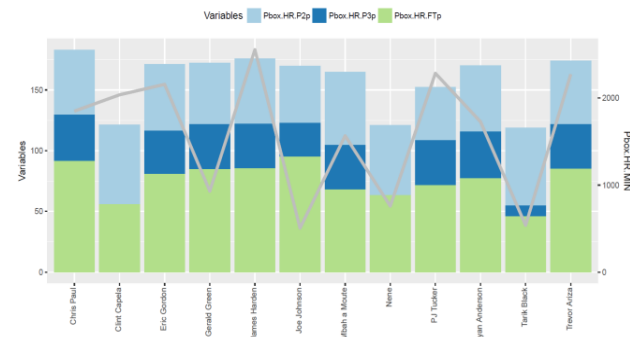
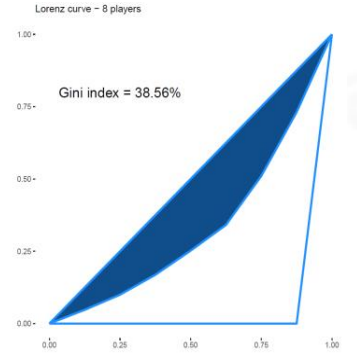
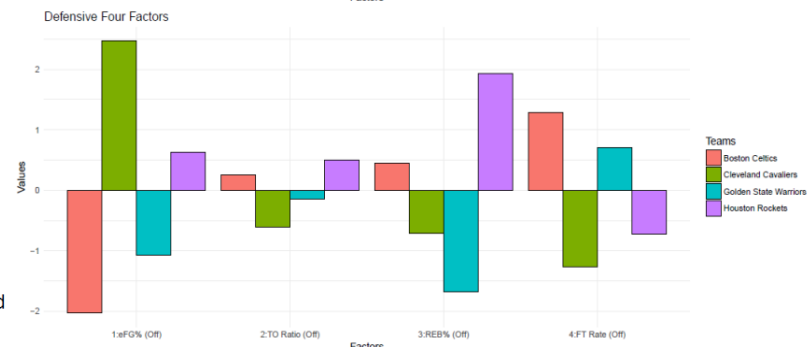
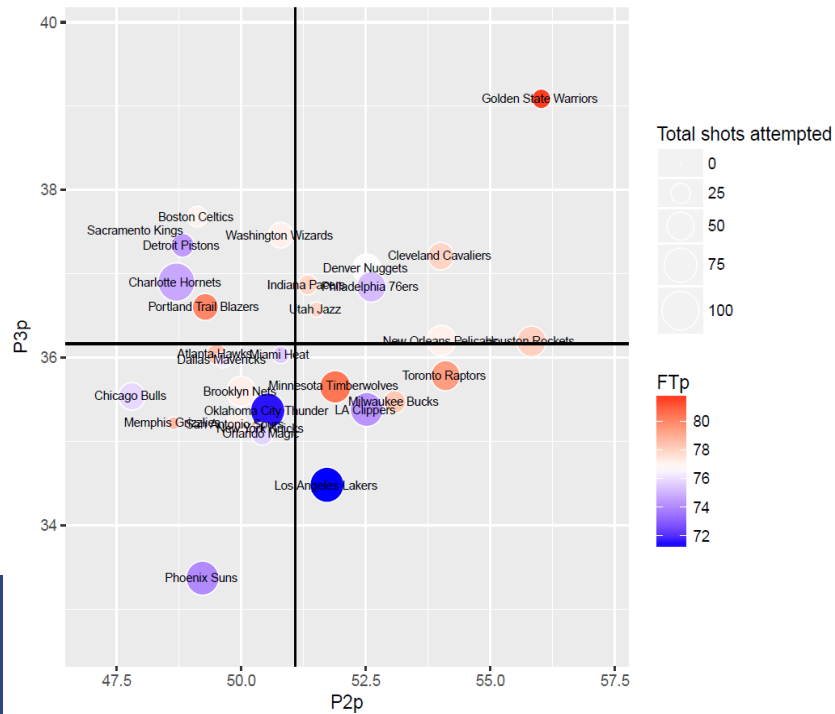
# R script

`bdsports.unibs.it/basketballanalyzer/`



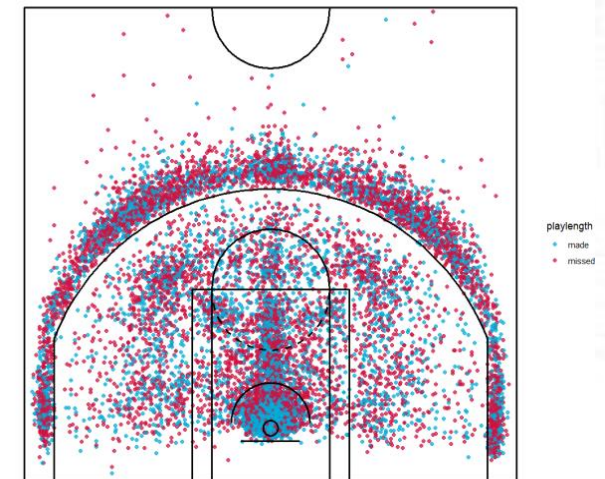
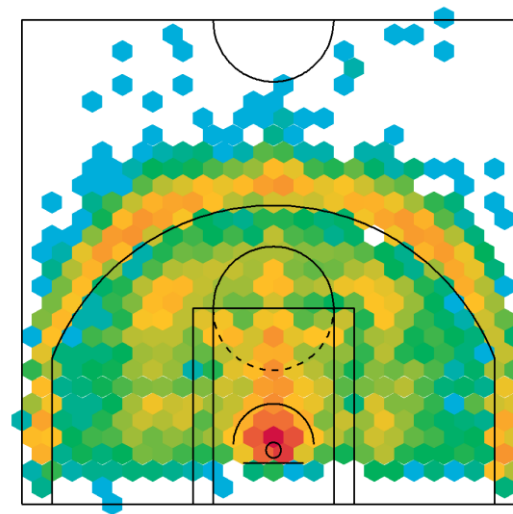
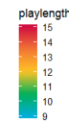
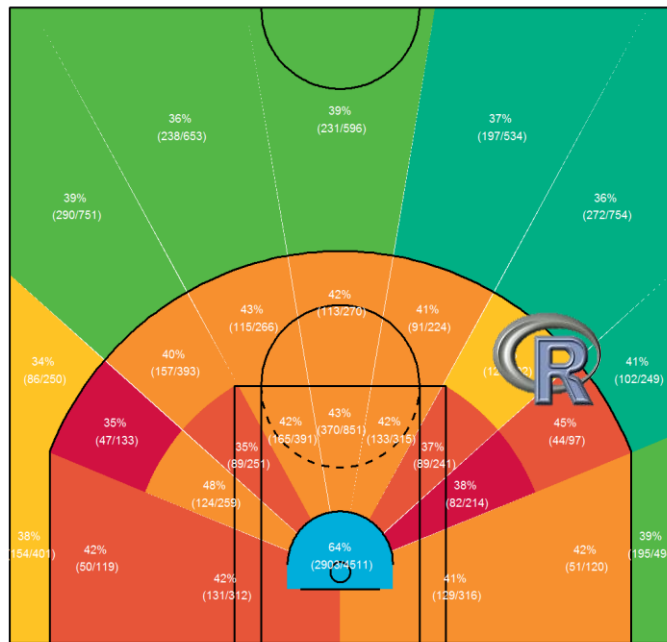
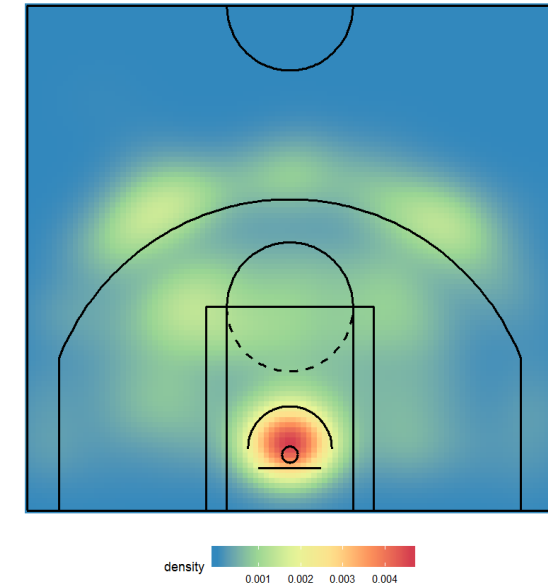
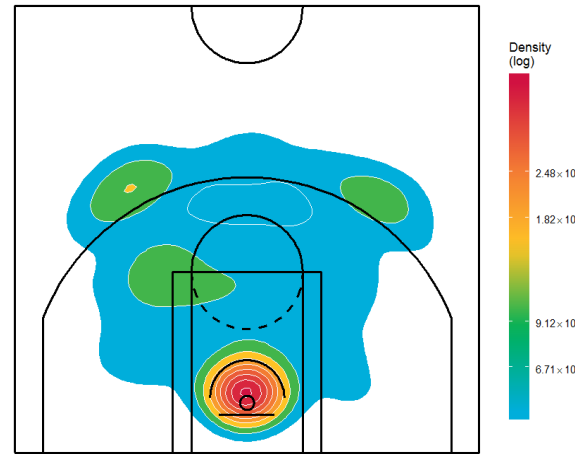
```
#####  
# July 2019  
#  
# The following R code allows to replicate all the analyses and examples  
# in the book "Basketball Data Science" (by P. Zuccolotto and M. Manisera),  
# forthcoming as a CRC Press publication.  
# It is based on the "BasketballAnalyzer" package developed with M. Sandri.  
# 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)
```

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

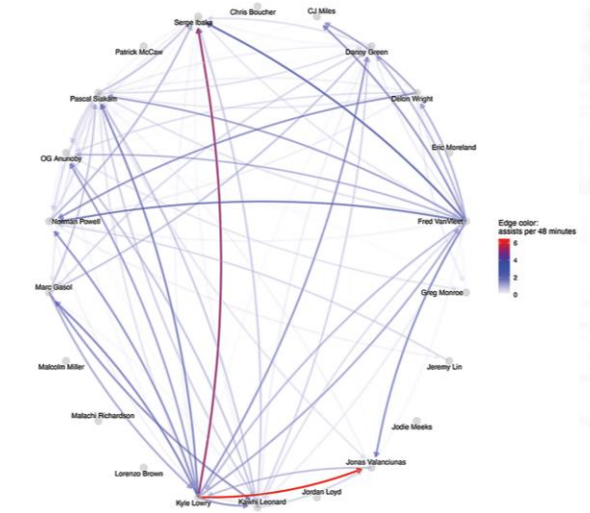
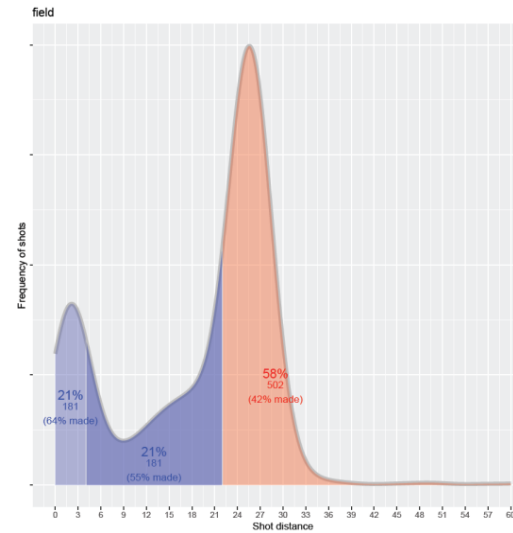
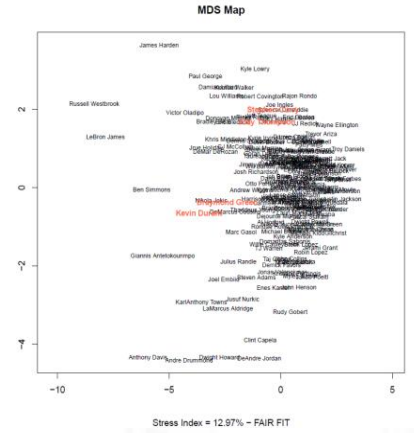
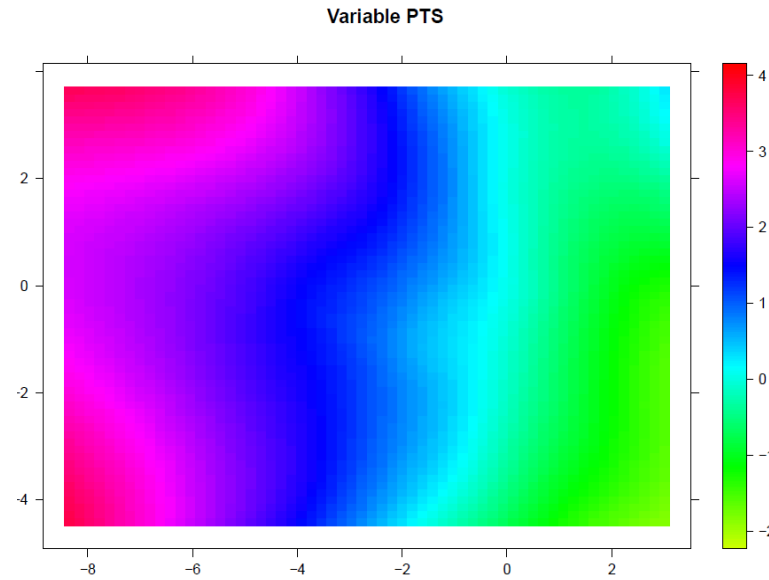
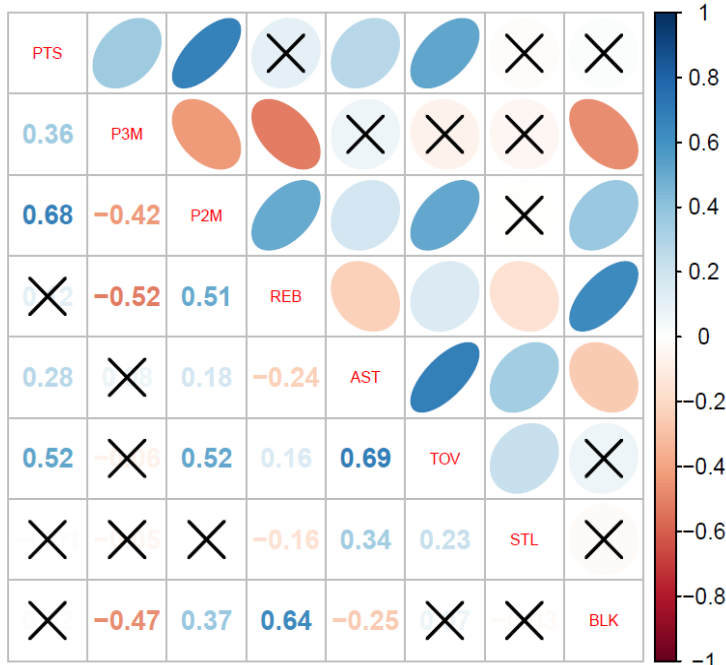




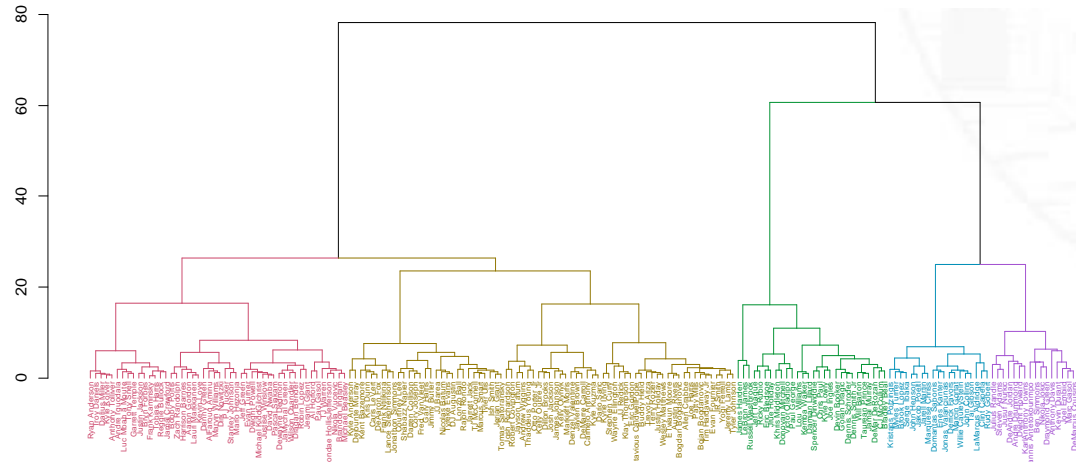
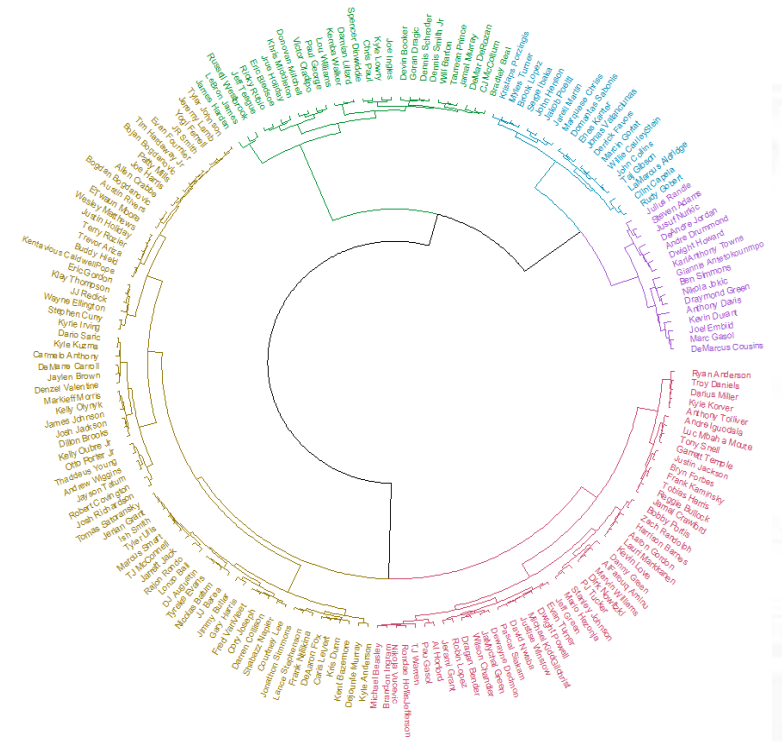
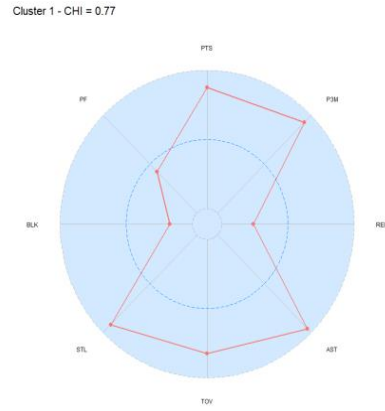
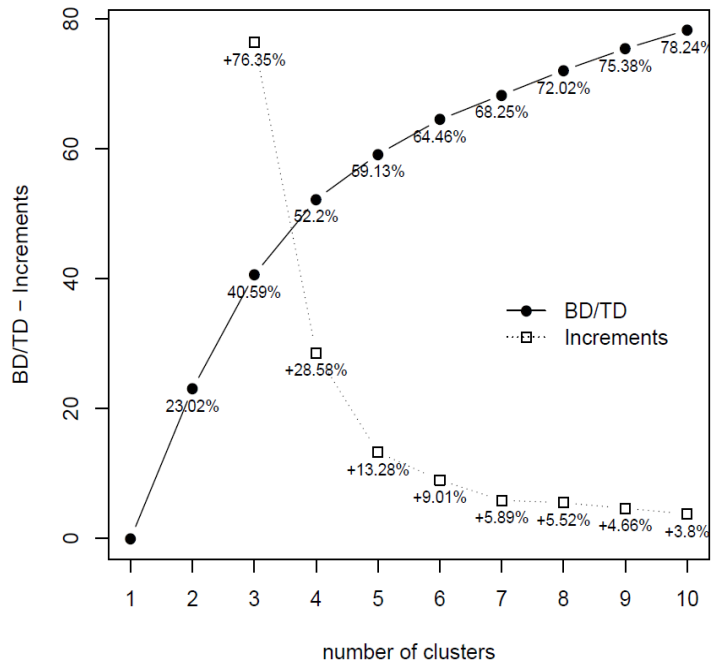
- **Basic Statistical Analyses**
- Discovering patterns in data
- Finding groups in data
- Modelling relationships in data



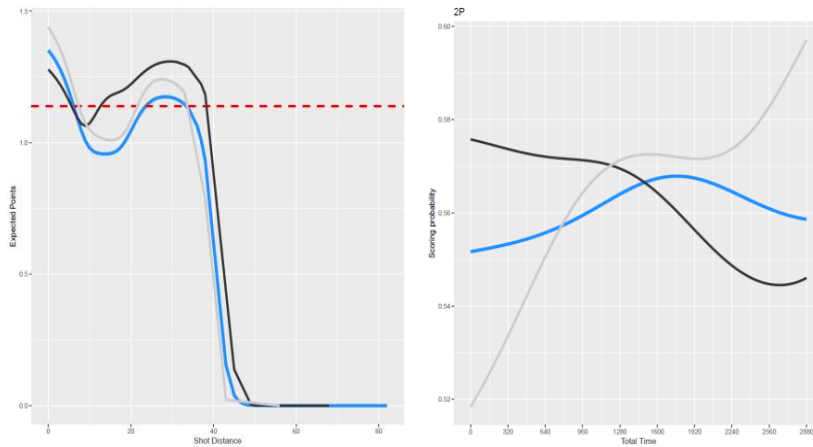
- Basic Statistical Analyses
- **Discovering patterns in data**
- Finding groups in data
- Modelling relationships in data



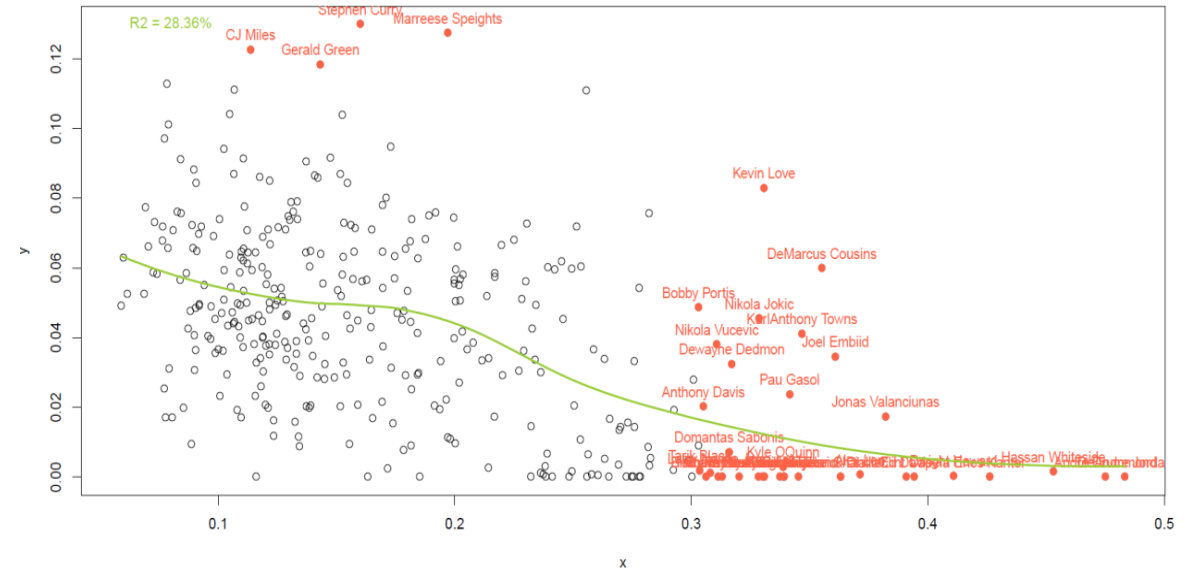
- Basic Statistical Analyses
- Discovering patterns in data
- **Finding groups in data**
- Modelling relationships in data



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



Simple regression



Simple regression

