

Module 1. History and evolution of *sports analytics*

Martí Casals, Professor, and researcher in the Sport Performance Analysis Research Group (SPARG) of the Centro de Estudios en el Deporte y Actividad Física (CEEAF) at the Universitat de Vic - Universitat Central de Catalunya UVic-UCC (UVic-UCC). Sports statistician and collaborator at Barça Innovation Hub – Universitas.

1.1. Introduction to the module

This module aims at knowing the first sports analysis fundamentals for your introduction and motivation into the new world of *sports analytics*. At the same time, we want to make you think critically about the opportunities and possibilities this technology has and, above all, that you don't concede a goal. What does this mean? We will see that sports analytics has existed for a long time in the professional world, we will see its great evolution and changes occurred in the last years, we will know updated information from experts to take into account in this field and we will present the introductory ideas for those who are new to this field of study.

To begin with, we invite you to watch a video, which is titled: "This is how Felipe Flores was sold to Belarus", which reflects what has happened and is happening, for example, in the world of football and in the analysis of the purchase of a player.

We currently know many things that have been said in the world of football, but occasionally we do not have all or enough information. In the sports industry, there is usually an abuse of the term *big data* and the most important is forgotten: where this data come from and whether there really is any problem or intention to use it. The video reflects the importance of the process the scientific method entails and what can be learned from it. The scientific method is a cycle that makes us think about its process before stating any opinion or belief: from the moment a problem is formulated (with its hypothesis), a plan is carried out, thinking about the type of data collection and information that is available, an analysis is performed and, afterwards, communication and interpretation of the initial problem is performed.

1.2. The *sports analytics* phenomenon

The evolution of sport in recent years has been important and, as we know, it has generated high levels of expectations, to a significant economic, health and transfer of

industry impact. It has also generated a new area of knowledge and promoted the formation of sports knowledge and innovation centres, such as Barça Innovation Hub.

Currently, sport is controlled by means of a very wide range of records that analyse very varied environments: from exercise physiology, psychology, *marketing*, medicine to modern digital systems for the analysis of the performances during competition.

At this time, the challenge of knowledge is not in the increase of the type of records that can be generated and used (which will surely grow in parallel with the technological progress), but it will come as a result of the correct selection of the records that are more accurate and relevant to expert users. The adaptation and evolution of the creation of new multidisciplinary teams will also be important (that integrate sciences such as mathematics, statistics, physics, philosophy, computing, among others), that are familiar with the context of the world of sport, that provide greater knowledge and help solve practical problems in the short, medium and long term.

The sports community closest to the scientific field understands that the statistical and mathematical treatment of this huge amount of data currently produced is opening a new area of knowledge and business. We have gone from writing down the records we see in a football or basketball game (notational analysis) to being able to obtain a wide scope of records coming from increasingly sophisticated technologies.

To sum up, we have more technology within our grasp, which results in much more data, but we still have to make better decisions based on appropriate statistical models and analyses. What the statistician Frank Harrell calls thinking is used here.

Currently, we hear much more about statistics (or *analytics*) in sport than some years ago. Sports statistics is more popularly known as **sports analytics**. The well-known *sports analyst* profiles are most used in the United States (US) and the term *performance analyst*, in countries of the United Kingdom (UK). But what is *sports analytics* really? After completing this and the following modules we will be able to answer this question, but sports analytics currently makes us think about the interest in data analysis that members of the same staff have (for example, the coach, analyst, doctor, physiotherapist, physical trainer, etc.) "thanks to the *Moneyball* film, the constant technological innovations currently available, conferences of the sports analysis industry such as Sloan, or scientific conferences and recent works by sports data scientists" (Casals, 2018, <https://barcainnovationhub.com/es/bioestadistica-ciencias-del-deporte/>).

The phenomenon of *sports analytics* is not an individual activity. From the business point of view, we tended to look for a "superstar" player or data analyst or scientist in our team. In any case, we often forget that the most important thing is to build a multidisciplinary team in which other disciplines that encompass this task are also considered, such as sports science, behavioural science, medicine and data visualization, among others, besides statistics, which must be oriented towards a more tactical approach and which optimises sport performance.

We are facing a new data-oriented culture that requires changes to the current management model of organizations. In sport, as in other fields, this will not be immediate. For moment, sports analytics departments have started to be created and renewed in clubs [for example in the NBA], allowing different professionals from a same coaching staff (analysts, physical trainers, physiotherapists, coaches) to get closer to data scientists (Casals, 2018, <https://barcainnovationhub.com/es/bioestadistica-ciencias-del-deporte/>).

The skills and backgrounds of these data scientists can be diverse, and it will be important (as discussed in the next module) to know well what type of data scientist you want on your team. Among them, sports statisticians have an interesting profile that coincides at a time in which statistical science and its profession are at their peak. This is considered one of the sexiest jobs of the XXI century.

It is very common to wonder who may be really interested in statistics applied to sport. We would be surprised to see the amount of professionals attracted to it, from sports *managers* or *decision-makers*, players, coaches, physical trainers, doctors, rehabilitation specialists, journalists, or bookmakers, *scouts* or video analysts, to sports scientists, academics, fans, psychologists, epidemiologists, scientists from other fields and statisticians (see figure 1). In fact, there is a lot of people interested in it, but the key question will be to know what the actual objective of each one of them is. That is, why do they want to analyse data in sports? And for what purpose? Most of the students of this course, interested in a profile of sports analyst, will speak with different profiles and communication between you will be key to try to solve a problem with a common global objective.

Figure 1: Who can be interested in analysing data in sport?



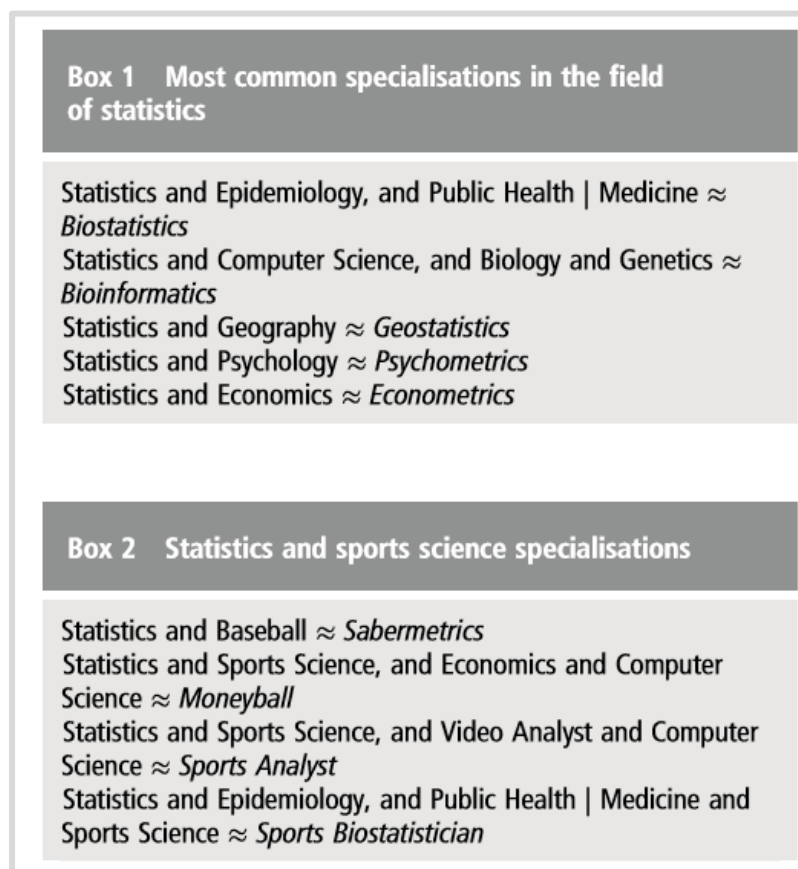
- Managers and Decision makers
- Coaches
- Trainers
- Players
- Journalists
- Bookmakers
- Betting market
- Analysts, Scouts and video analysts
- Academics
- Fans
- Sports Scientists
- Sports medicine experts
- Physicians
- Physiotherapists
- Psychologists
- Epidemiologists
- Researchers (other fields)
- Statisticians (design and develop new statistical analysis models...)

Casals M, Finch CF. Sports Biostatisticians – a critical member of all sports science and medicine teams for injury prevention. *Injury Prevention*. DOI: [10.1136/injuryprev-2016-042211](https://doi.org/10.1136/injuryprev-2016-042211)

Source: Casals and Finch, 2016.

In the scientific world, interest in the application of statistics has grown and, above all, in its greatest scientific rigour, in sports medicine and sports science in general. Statistics, together with other skills or professions, has created different common specializations, such as biostatistics, bioinformatics, geostatistics, econometrics and psychometry (see figure 2). The film *Moneyball*, directed by Bennett Miller and produced by Michael De Luca, also helped to awaken the interest of sports scientists and, in this way, to know different sports specialisations, such as *sabermetricians*, *Moneyball*, *sports analysts* and *sports biostatistics*, where the skill of statistics is always present.

Figure 2: Common specialisations in the field of statistics.



Source: Casals and Finch, 2016.

1.3. Specialisations in sports statistics

Sabermetrics:

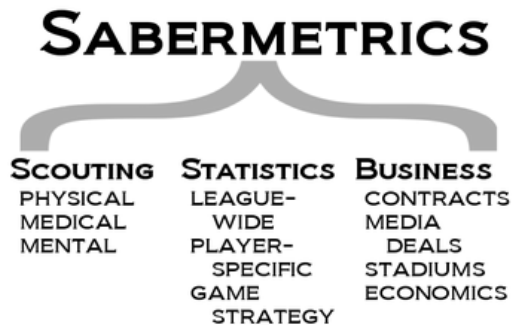
One of the first professions related to data analysis in sport are *sabermetricians*, experts in advanced statistics in baseball (see figure 3).

Figure 3: *Sabermetrics*, the study of advanced statistics in baseball.

$$f(\text{Sabermetrics}) = \text{statistics} + \text{scouting} + \text{business} + \epsilon$$

In words: Sabermetrics the study of baseball statistics, baseball scouting, baseball business, and anything yet-known or missed by myself (which is the “ ϵ ” epsilon).

More specifically:



Source: own elaboration based on [Untitled image about *Sabermetrics*], 2013.

Sabermetrics:

“It is the empiric analysis of baseball, especially based on the statistics collected during the games. Sabermetricians collect and compile the relevant data produced during a game to answer specific questions. The term comes from the acronym SABR (Society for American Baseball Research), founded in 1971. The word Sabermetrics was created by Bill James, who is one of the pioneers in this field and is considered the visible face of the movement» (Lluís Carrillo, 2017, <https://www.beisbolmlb.com/sabermetrics-que-es>).

In 1980, Bill James defines *sabermetrics* as “the search for objective knowledge about baseball” (...) often using statistical analysis to question traditional measures of baseball evaluation such as batting average and pitcher win” (Birnbaum, undated, <https://sabr.org/sabermetrics>).

One of the research guides on *sabermetrics* was created by Phil Birnbaum and can be found in <https://sabr.org/sabermetrics>. This guide provides the key points to bear in mind before going into the world of *sports analytics*, where aspects of the base of *sabermetrics* are reviewed, how to ask good questions, how to find *raw data*, and where to publish information related to baseball science and analytics. Besides, it contains information from the *By the numbers newsletters* magazine for the *Statistical Analysis*

Research Committee that offers data to understand the history of sports analysis to date, always taking into account the scientific world and also the sports industry.

Moneyball:

Moneyball is a book (Lewis, 2003) and then a film that marked a turning point in the analytical thinking of low-budget teams. It tells the story of a baseball team that uses novel statistical approaches to help *general managers* make decisions to achieve sports success. The film tells the story of a poor team: Oakland Athletic's (a relatively good team, but that at the end of the 2002 season loses 3 players). Its *general manager* Billy Beane decides not to do the same as other teams and, betting on statistical methods, he tried to sign players and build his team taking into account his low budget. This team was, consequently, the first to prioritise quantitative analysis by means of statistics and data to make decisions. Later, with the film, many low-budget clubs (Brewers, Indians, and Rays) followed the Oakland Athletic's approach (see figure 4). Brad Pitt, in the role of Billy Beane, got Oakland Athletic's to win the same number of games as the New York Yankees, although they were later eliminated in the *playoffs*. Two years later, another team (the Red Sox) won the league with the same model used by the Oakland Athletic's.

Figure 4: Oakland Athletic's.



Source: Hoffman, 2012.

The film *Moneyball*, different from what sabermetrics is, focuses more on the value and price of players. In fact, criticism and comments were made by the *sabermetricians* about how they had focused and showed their profile in the film. More information

about it can be found in <https://sabr.org/latest/commentary-critiques-new-moneyball-movie>.

The *moneyball* is a strategy for using data analysis and focuses on how to get the most value out of the investment made in the construction of a sports franchise. The concept consists in selecting and analysing data in order to identify undervalued players so that teams with lower budgets can compete at a good level.

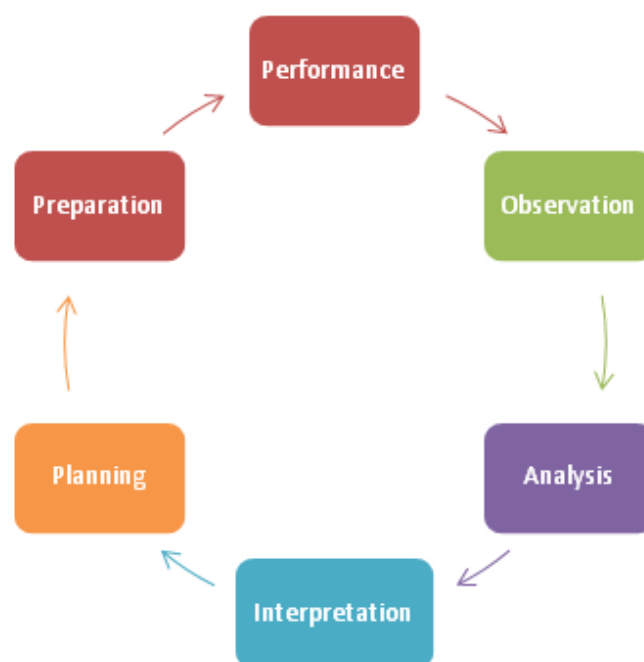
The *moneyball* can be effective, but it is not always the best strategy for franchises or teams. Sports information and analysis systems may require an important investment in tools and staff.

Performance analysis:

Performance analysis (as it is known in the United Kingdom) can be generally classified into two main categories: *notational analysis* and *motion analysis* (Carling, 2005). Notational analysis provides a register of the characteristics on the position of the ball, the players involved, the action in consideration, the time, and the result of the activity, etc. Motion analysis, for its part, focuses more on the raw characteristics of an individual's activity and movement; for instance, it identifies fatigue and workload measurement.

Performance analysis is not about just analysing games, but it is important in the training programme in order to help coaches improve the players' performance. The following figure shows a training cycle based on three stages: observation, analysis, and interpretation.

Figure 5: The training cycle and its stages.



Source: Chan, 2019, p. 1.

Anyway, the *performance analysis* concept ends up meaning more than an organization can imagine and it is not easy to cover the three stages discussed above

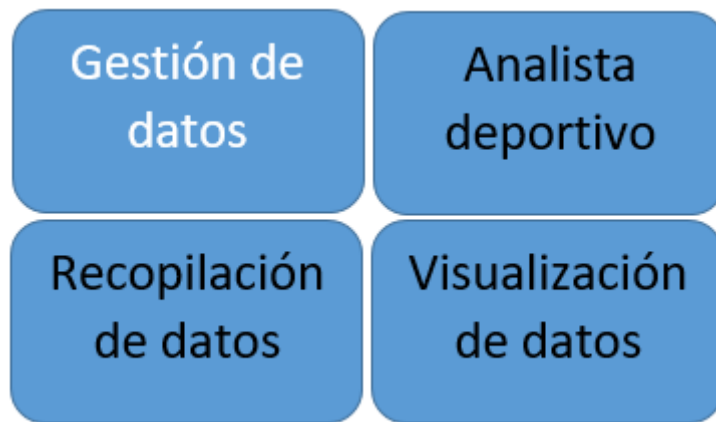
Figure 6: Concepts and activities covered by performance analysis.



Source: Coles, (2014).

Rob Carroll is the main *performance analyst* and has been, for years, consultant and main adviser for several teams and sports organisations in Ireland. Besides, he's the creator of the webpage *The VideoAnalyst*, which provides complete information of performance analysis by means of tutorials, courses, interviews, job offers, and much more. This site brings us fully to the *sports analytics* phenomenon (see figure 8).

Figure 7: Skills to highlight.



Source	Translation
Gestión de datos	Data management
Analista deportivo	Sports analyst
Recopilación de datos	Data collection
Visualización de datos	Data visualisation

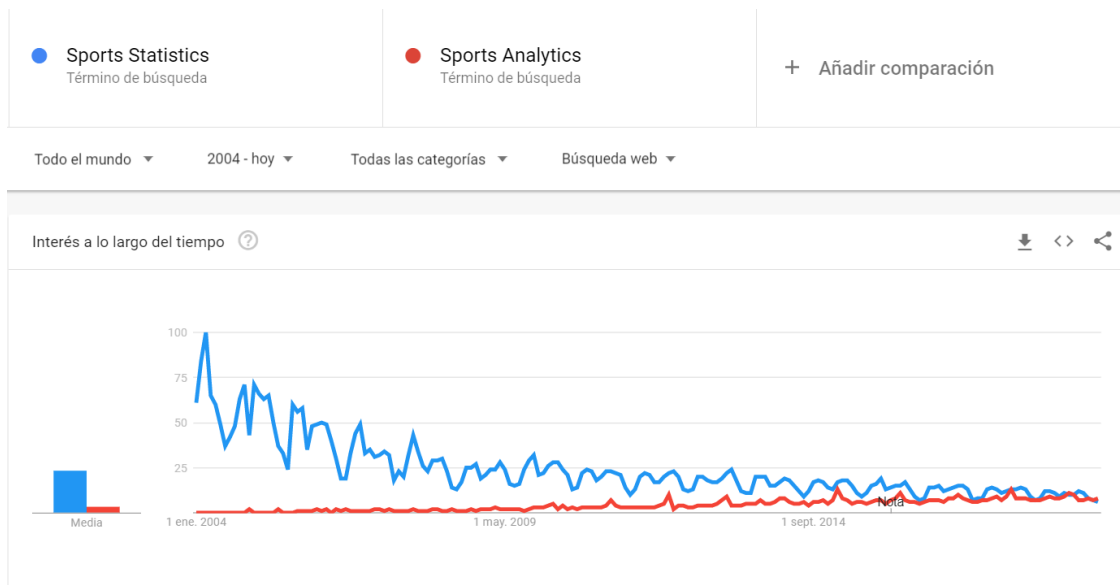
Source: own elaboration.

Finally, we should have to mention the figure of the *sports analyst*. We will explain its history and evolution in detail at the end of this module.

Sports analytics:

As previously commented, sports statistics is better known as *sports analytics* (especially in the United States and Australia).

Figure 8: Comparison of the terms “sports statistics” and “sports analytics”.



Source	Translations
Término de búsqueda	Search term
Añadir comparación	Add comparison
Todo el mundo	All the world
2004 – hoy	2004 – today
Todas las categorías	All the categories
Búsqueda web	Web search
Interés a lo largo del tiempo	Interest over time
Término de búsqueda	Average

Source: Google Trends screenshot (Author of software rights, version year) with own adaptation.

A few years ago, if we searched, for example, in Google, the terms "*sports statistics or sports statistician*" we found the figure of the *journalist* who dedicated to analysing sports statistics. Later, we could find the ASA sports statistics section, which was founded in 1992 at the *Joint statistical meeting* with the aim of promoting statistical education and, in general, making statistics accessible to people interested in sports.

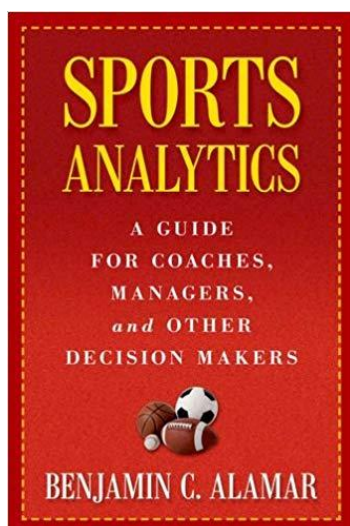
We can define the *Sports Statistician* as: sports statisticians analyse the data related to sporting events, generally for important sports such as baseball, football, or basketball. On certain occasions, data analysis, collection and recording are performed live, as they occur. Two types of sports statisticians are distinguished here: the more academic and the "*recorder*", the latter is in charge of attending sporting events and records the data in real time. These two profiles have joined together in recent years and have worked as external consultants for sports clubs and institutions and / or players.

But the sports statistician not only needs to master statistics, but much more, as some portals already informed in 2001. In fact, the term *analytics* is a trade name on the use of data to report management decisions. This term means, *sabermetrics for business*.

From the point of view of *business*, the term *analytics sports* was explained by the prestigious analyst Ben Alamar (2013), *sports management* professor at Columbia University, in New York city. Before this, he was director of *sports analytics* at ESPN and worked with NBA teams, including the Oklahoma City Thunder for 5 years. He was also the founder and editor of the Journal of Quantitative Analysis in Sports, the world's first peer-reviewed journal of *sports analytics* research.

Ben Alamar is mainly known by the book he wrote in 2013 called *Sports analytics: a guide for coaches, managers, and other decision makers* (Alamar, 2013) (see figure 9). The book describes the state of sports analysis in several fields, how many analysts' sports teams have, how they use the information, and how they perceive the use of their analysis department. The book also describes the purposes and uses of data in sport. It finishes with examples and reflections on the implementation of analytics within an organization and how to link the work of decision makers and analysts, including how to construct a successful analytics department. It is really a book for the GM to keep in mind, especially for those clubs or organizations that plan to set up *sports analytics* departments.

Figure 9: Cover of Ben Alamar's popular book on *sports analytics* published in 2013.



Source: [Untitled image of Ben Alamar's book], s.f. Taken from <https://cup.columbia.edu/book/sports-analytics/9780231162920>

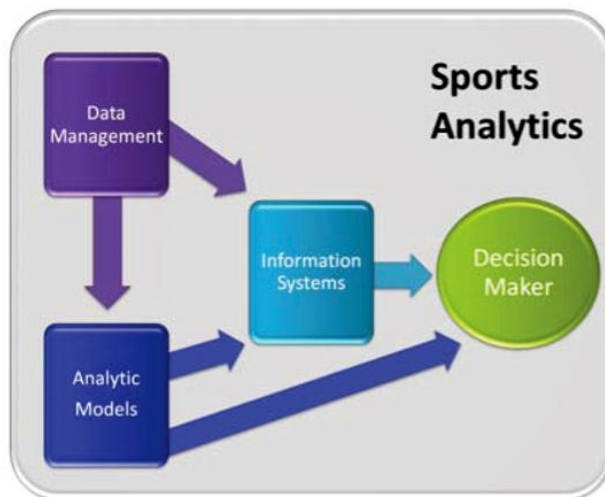
Curiously, in a 2011 editorial in the *Analytics magazine*, Ben Alamar and Vijay Mehrotra defined *sports analytics* as:

"the management of structured historical data, the application of predictive analytical models that use that data, and the use of information systems to inform decision makers and help their organizations gain a competitive advantage on the field of play." Our definition is both expansive (in the sense that it includes not only statistical models, but also the broader information value chain that surrounds these models) and restrictive (because it excludes

traditional analysis applications such as demand forecasting, revenue management and financial model, all of which are certainly relevant in the professional sports business)

This *sports analytics framework* is presented in the following figure.

Figure 10: Framework from sports analytics.



Source: Alamar and Mehrotra, 2011.

1.4. Technological evolution, data providers and visualization in sport

The sports analytics area has grown exponentially. One criticism of using technology is that it can slow down the game, but on the other hand, for many people, it is more pleasant to see that the right decisions were made.

In 2006, the use of the hawk's eye (HawkEye) in tennis began to be usual, a technological device that consists of a system of high-resolution cameras installed in different parts of the stadium whose mission is to follow the trajectory of the ball. This system is also used in other sports such as basketball, cricket, Australian football, baseball, rugby, among others; until the recent VAR, the new football video refereeing (McGarry, O'Donoghue and De Eira Sampaio, 2013; Hüttermann, Noël and Memmert, 2018; Dugalić, 2018)

The NBA and STATS LLC announced in 2013 that they were expanding their agreement to install SportVU Player Tracking technology cameras in all NBA halls from the beginning of the 2013-14 season so as to quantify and analyse all the movements of the games during a whole season.

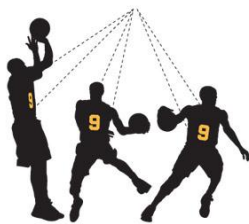
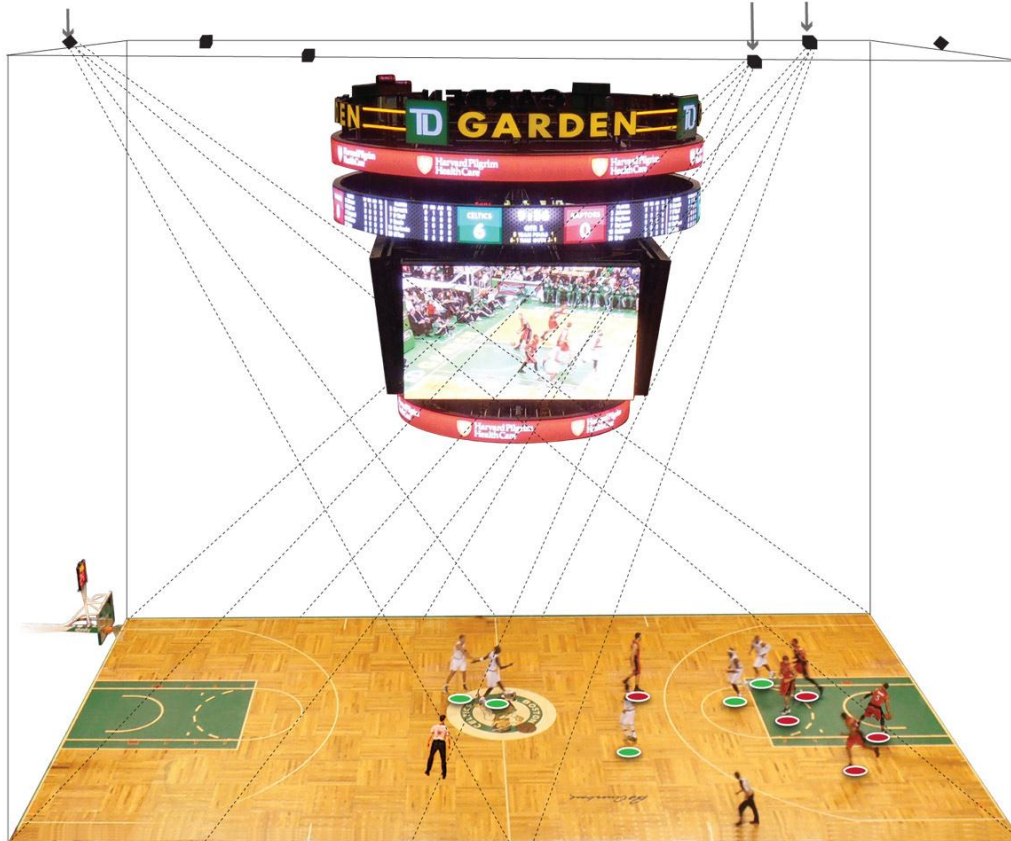
Figure 11: How SportVU technology works.

How SportVU technology works

- 1 Fixed to a metal catwalk, six small cameras are used to track the movements of all 10 players, the three referees, and the ball. The cameras use Israeli missile-tracking technology.

One camera above each basket to cover opposite half of court.

Two cameras over each sideline to help cover both halves of court.



- 2 Optical recognition software picks up jersey numbers and captures every pass, shot, and dribble. Images are collected 25 times per second.



- 3 The software produces data based on player movement in categories such as drives, speed, and distance, touches per possession, catch and shoot, and defensive impact.

Teams
Receive advanced metrics that go beyond basic stats.



Fans
Get more general information to track the game live.

- 4 The play-by-play data can be delivered in a report within 90 seconds of a play. Both team personnel and casual fans have these live statistics at their disposal.

SOURCE: SportVU

JAMES ABUNDIS, LUKE KNOX/GLOBE STAFF

Source: Goldstein, 2017.

Goldstein, (2017), states: "We are a league where data is fundamental, and our extended agreement with STATS will provide our teams and fans access to discover incredible

statistics", said Steve Hellmuth, NBA Executive Vice President of Operations and Technology. "In this era of statistical information, SportVU will be an invaluable source for basketball executives and our fans."

The NBA is one of the competitions that bet on the new era of technology and data in sports. Besides, this technological evolution has brought the world of sports science closer, with Australians being the pioneers and (initially) most requested profiles. Aaron Coutts, professor of sports and exercise science at Sydney University of Technology, holds that Australia definitely has a very strong mark in sports science

In the world of sports, the already famous **vests**, which many athletes currently wear in training and games, are made up of sensors that, incorporated into a specific joint, can extract biomedical information. Previously, it was necessary to perform a work of information and coordination crossing between different *softwares*. Now the tactical, physical, and medical aspects are integrated, and all the information can be centralized in the cloud together. In the near future, this will probably help to achieve greater multidisciplinary among the different professions, be they current ones (*managers*, trainers, doctors, physiotherapists, physical trainers, rehabilitation specialists, analysts) or the new ones that will emerge in these areas.

At FC Barcelona, for example, a technology called WIMU is used, a wireless device with sensors that monitor physical activity, providing accurate information continuously and in real time. This device is used to track the athletes' biometric data, improve their performance, and prevent injuries. It is currently used in all professional sections of the club, as it can be used in *indoor* and *outdoor* sports.

Currently different professional teams use this information to monitor or control injuries in their athletes by means of different devices. But how do athletes and staff members perceive this implementation? According to a recent study, this helps improve communication between the athlete, the technical *staff*, and the medical staff. (Barboza, Bolling, Nauta, Van Mechelen and Verhagen, 2017). Besides, from the point of view of the doctor, this collaborates to improve the knowledge and control of the athlete, to intervene when necessary (Barboza et al., 2017).

One of the main problems caused by the use of so much technology has to do with the measurement error and the validations that this type of instrument requires. In sports, many evaluations are made using tests or other instruments, but many times, the validity and reliability of this data cannot be guaranteed, in spite of its broad use. Can we ensure that the value we record after a valuation is reliable? No. Less than 15% of the technologies are validated and, therefore, we have to know what we are measuring (Peake, Kerr, and Sullivan, 2018). And, if the valuation is not reliable, does it depend on the person responsible for measuring it? We currently use this data to make decisions, but do we know if it is true or do, we know its margin of error? Do we have certain responsibility? These are questions that we must ask ourselves in a sports organization that is interested in data analysis.

A book of interest, *Advantage Play: Technologies that Changed Sporting History*, (Steve Haake, 2018), tells the story of the key technological advancements used to obtain an advantage in sports. Technological advances also go together with the evolution of other activities such as sports analysis conferences. This is the case of the popular *MIT Sloan Sports Analytics Conference*, which is a meeting where leaders, players, students, journalists and other professionals gather, performing different talks and presentations with a common topic: discussing the increase of the role of analytical statistics within sport, covering from NBA to sports such as football (*soccer*). At the Sloan conference it has been possible to enjoy the presence and talks of personalities from sports, and even politics, such as Shane Battier, Nate Silver, Billy Beane and Daryl Morey (*general manager* at Houston Rockets) and Barack Obama.

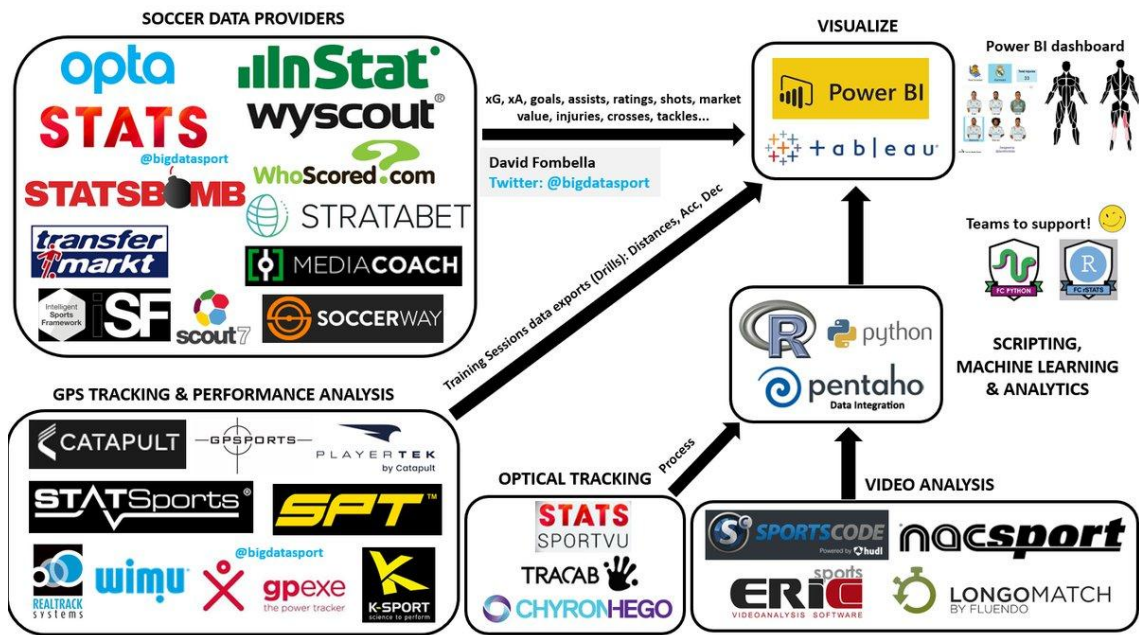
Sports analytics conferences are also growing, and it is important to consider their dates, the programmes (some with more academic, scientific, technological, sports industry or business approaches) and main topics and innovations to discuss. One of the most important academic events in sports analysis (apart from some round table organized at the American Statistical Association (ASA) is the New England Symposium on Statistics in Sports. Recently, one of the appointments that the Barça Innovation Hub has organised is the Barça Sports Analytics Summit.

Data providers:

Many times, in the corridors of a sports organization or club, the analysis department talks about data providers and it is common to hear about *OptaPro*, *Wyscout*, *Stats*, *Instat*, *StatsBomb*, among others. Every day there is more information, so the synergy and teamwork between IT and *Sports Analytics* departments is very important before starting to work on this data. It is important to become familiar with the concepts and nomenclature (for instance: *tracking data*, *metadata*, *box score*, *eventing* or *play-by-play*, *API*, *scraping*, etc.) that these departments use within the club and are multiple, comprehending both areas, although sometimes they work independently.

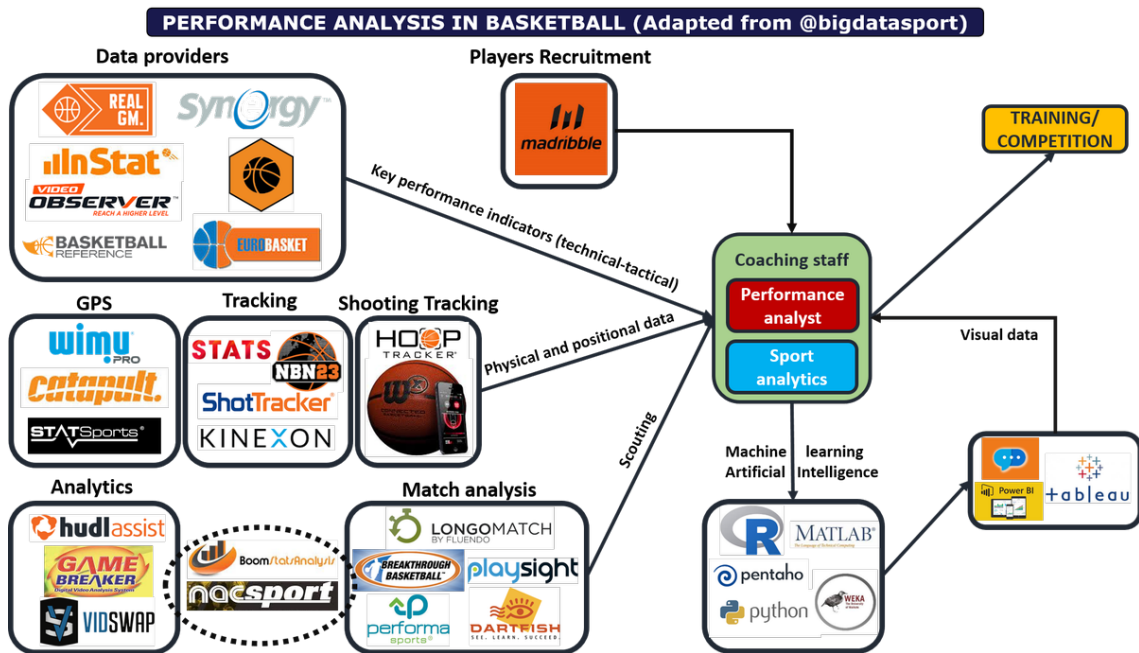
Currently, it is important to know the data providers in different sports. It will be key to be familiar with the companies that provide the data before entering the world of data analysis in sports. Figures 13 and 14 show a small sample of the existing options in sports such as football and basketball.

Figure 12: Data providers in football



Source: Fombella, D. Big Data applied to football. Sport providers analysis and visualisation. Undated.

Figure 13: Data providers in basketball.



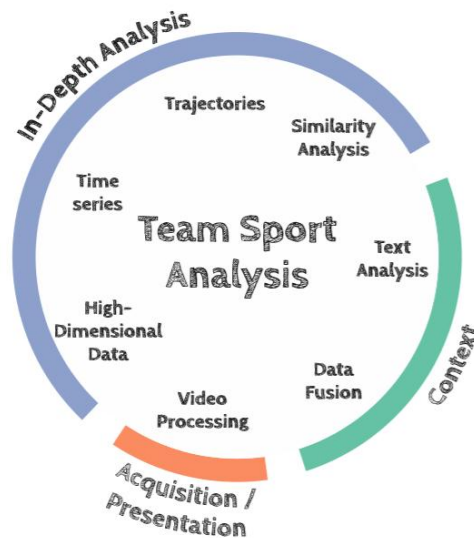
Source: Fombella, D. Big Data applied to football. Sport providers analysis and visualisation. Undated.

Stein et al, (2017) gives us a panorama of the possibilities and problems, from the simplest to the most complex, that we can find in the world of *sports analytics*, *visual analytics*, and *computer science*. It is a compulsory reading for those who want to begin



to know the possibilities of the *computer science* and visualization, it helps to understand the context and the challenges that we can find in sport (see figure 15).

Figure 14: Data acquisition and enrichment through contextual information.

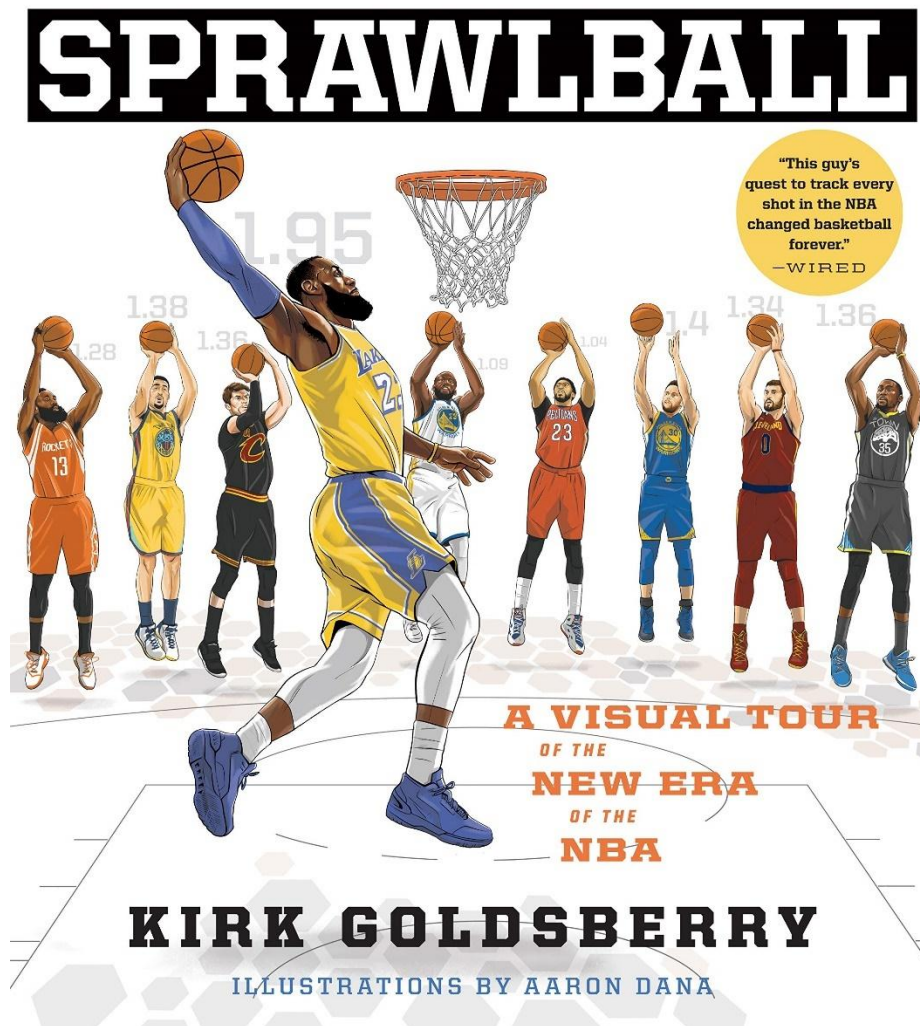


Source: Stein, et al., 2017. p. 2.

When talking about data visualization, although it is not a subject to be covered in detail in this course, it is important to take into account the figure of Alberto Cairo in order to become a good *data viz*. Alberto Cairo is a journalist and designer, director of the visualization program at the University of Miami Centre for Computational Science and famous for his books, which you can consult on his website. Speaking about the evolution of technology in sports leads us to mention the evolution of data visualization (also called ***data viz***) in the world of sports.

Although some portals on the *FiveThirtyEight* website have already shown us some possibilities for *data viz*, the author Kirk Goldsberry is an expert in geography and cartography, and in map visualization. Also, he is a basketball writer, he was vice president of strategic research for the San Antonio Spurs (NBA), he is the principal analyst for the United States basketball team, and he is a visiting researcher at the Harvard Institute for Quantitative Social Science. All this has allowed Kirk to revolutionise data visualization in *basketball analytics* thanks to the books that you can also find on his website.

Figure 15: Cover Kirk GoldsBerry *SprawlBall: a visual tour of the new era of the NBA*.



Source: [Untitled image of Kirk Goldsberry's book] (undated). Taken from <https://www.amazon.com/SprawlBall-Visual-Tour-New-Era/dp/1328767515>

From a more academic point of view, we can find Perin et al. (2018) reflecting on recent advances and challenges in the field of sports data visualisation. It is a good sports data visualisation research guide covering the breadth of visualisation tasks and goals.

1.5. History of *sports analytics*. Is it really a fad or has it existed for a while?

Many of us cannot forget that the application of mathematical models to professional sports goes back more than 50 years (Lindsey, 1959). It is important to remember that in 2005 the first issue of the *Journal of quantitative analysis in sports* was created. At that moment, only two or three NBA teams thought about the use of advanced statistics in relation to players and strategy. The book by Michael Lewis (2003) had just been published and no one had seriously thought in the application of technology and motion capture in the context of professional sports. Six years later, more than half of NBA

teams used analysis tools alongside their operations team. Currently, most sports teams in major leagues consider data analysis normal in their teams, and companies like STATS LLC and many others installed cameras in NBA stadiums and in NFL stadiums to capture more and more data.

Many times, in the world of sports we have seen professionals remember their mentors. For example, we can remember in some Pep Guardiola press conference as a football coach, after winning a final, remembering the coaches and mentors he had had (like Johan Cruyff) to reach that top.

There are different examples of recognitions in history in sport to consider. In *sports analytics* we would not be fair if we did not do it. Therefore, many people wonder... Is *sports analytics* currently a fad? No, *sports analytics* is a process that has been happening for many years (see figures 17 and 18), as shown in the *post* written by Rob Carroll, undated(<https://thevideoanalyst.com/history-sports-analysis/>) This work, which is still under development, informs us of the great events in the history of *sports analytics* from 1861 to 2015. All that was also possible thanks to contributions from Keith Lyons'blog, to the article *A history of sports notational analysis: a journey into the nineteenth century* by Simon Eaves (2015) and to the book *The numbers game* written by Alan Schwarz (2004).

Figure 16: Some references reported by Keith Lyons.

19th Century Chronology

Some examples

- 1861 Henry Chadwick's Baseball notation (Eaves, 2015:1173)
- 1883 Renshaw v Lawford Lawn Tennis in J M Heathcote (1890) (Eaves, 2015:1161)
- 1884 C G Heathcote's assessment of real tennis in J M Heathcote (1890) (Eaves, 2015:1162)
- 1890 Huntington v Hovey Lawn Tennis in V G Hall (ed) (1891:55) (Eaves, 2015:1162)
- 1890 Hobart & Hall v Carver & Ryerson in V G Hall (ed) (1891:68) (Eaves, 2015:1163)
- 1891 Hall v Wren Lawn Tennis (New York Times) (Eaves, 2015:1164)
- 1892 Corbett v Sullivan boxing match (Eaves, 2015:1170)
- 1894 Goodbody v Hobart Lawn Tennis (Eaves, 2015:1163)
- 1897 Corbett v Fitzsimmons boxing match (Eaves, 2015:1163)
- 1900 American university football, Yale v Princeton, Yale v Harvard (Eaves, 2015:1168-1169)

Source: Eaves, 2015.

Simon Eaves (2015) explains the evolution of notational analysis and how, for example, football coaches used to record themselves their own data and now find it easier to receive it from other sources.

Figure 17: Sports analytics changed the game for good.



Source: Hunter, 2014, <https://www.fastcompany.com/3033324/how-sports-data-analytics-is-upsetting-the-game-all-over-again>

We can see how the history of sport analysis begins in 1861 with **Henry Chadwick**, considered the father of baseball, who began to wonder how he could quantify the data of this sport (for example: how often or at what average in training his players could hit better or worse) and he was named **the inventor of baseball's notational system**.

Later, in 1981, **tennis** began to collect information (mainly frequencies) more focused on players' performance. The type of event for each ball played is recorded, be it a foul, a served ball, a returned ball, along with the number of exchanges of balls played.

A few years ago, thanks to new technology, in sports analysis the use of **heatmaps**, where different ranges of colours are used to identify the player's transit zones, became fashionable. Although technological resources currently allow us to see heatmaps visualisations that are sufficiently graphic, in 1987 heatmaps applied to sport already started being made.

In 1910, David F. Barrett, a commentator, began to use some measures in football, although he was not the forefather of quantitative analysis in this sport as we will see later.

It was in 1947 when a statistician was hired *full time* for the first time (Mr. Roth) in the *Major League Baseball* (MLB).

And although we all know the work (book and film) of *Moneyball* (previously described), in 1960 some similar work had already been introduced in the *National Football League* (NFL).

To know the introduction of *sports analytics* in the world of football (better known as *soccer* in the US) we go back to 1930 and **Charles Reep** cannot be left out, the creator of the first data analysis system to improve football players' performance (see figure 19).

Figure 18: Keith Lyons.



Source: [Untitled image of Keith Lyons] (2013). Taken from <https://keithlyons.me/blog/2013/11/20/a-charles-reep-project/>

Charles Reep was considered the creator of the first data analysis system to improve football players' performance. This means that he was the first person who systematically recorded data related to different aspects of the play during games, when he realized that doing so was useful to identify recurring situations and teach players how to take advantage of them to generate scoring chances (Azcona, 2015, <http://sportics.es/charles-reep-precursor-del-analisis-de-datos-en-el-futbol/>).

Reep was a curious person, he studied accounting and he was also a football player. One of his known data analyses was when the GM of the Arsenal team contacted him to inquire about his analysis work. From here a first collaboration between the two began and helped his "conviction that goal opportunities could be generated by making the wings perform specific movements" (Azcona, 2015, <http://sportics.es/charles-reep-precursor-del-analisis-de-datos-en-el-futbol/>).

According to the *post* of Sportics (Azcona, 2015)

it was during a game that took place on April 15, 1950 in the pitch of the small local football club of Swindown that he realized that recording the movements of a team in real time was not complicated but it was

impossible to do it for both contenders at the same time. His experience led him to develop a system that would allow complete tracking of the movements of both teams during a game. At the end of the 49-50 season he had already designed it.

In the 1950-51 season Charles committed to work with a club for the first time, Brentford. In the moment he joined as an analyst, the team was fourth at the bottom of the championship standings. From that moment on, and relying on Reep's analyses, they modified some playing patterns, winning thirteen of the fourteen following games. Their only loss came when the coach decided to bet on being more defensive in the game against the Preston North End team, coached by Tom Finney (Azcona, 2015, <http://sportics.es/charles-reep-precursor-del-analisis-de-datos-en-el-futbol/>).

It was in 1970 when baseball started talking about the "Measuring Clutch".

In basketball, for example, on many occasions it is tactics, concentration, sacrifice or knowing how to lead the game towards victory. Occasionally, it has situations where only the players with the most confidence and experience can perform the most exciting moments of this sport: the 'clutch shots', that is, the shots at decisive moments of the games. There are two points, three points, suspended, with a defender on top or even two. Some serve to force the extra time, others to take the victory. But one thing they have in common is the adrenaline they generate in oneself and in rivals, in fans, coaches and anyone who is passionate about the game of ball and basket.

In 1971, the *Society for American Baseball Research* became an entity, thanks to Robert Davids, where the term "statistorians" was used to refer to those who were interested in the history of baseball and statistical research.

In 1970 the new era begins (1970 to 2004), called **Jamesian era** among sports statisticians. At this time, society and sports professionals begin to see that statistics can be applied in sports. In 1977 Bill James published the book *The Bill James baseball abstract* with detailed statistics, that went on sale through a small advertisement in *The Sporting News*. During the next three years, James' work gained respect, including highly favourable criticism of Daniel Okrent (1981) in *Sports Illustrated*. In 1982 sales have multiplied by ten and different media agreed to publish and distribute future editions. In this same era, in 1981, the Oakland A's team was the first client team of *Sports Team Analysis & Tracking System* (STATS). In 1992, the first Analysis Performance Centre was created and in 1996 Optais the first to partner with the Premier League through the launch of a ranking system, The Opta Index.

In 2003 before the beginning of a new era or stage, Michael Lewis's book "*Moneyball: The Art of Winning an Unfair Game*" appears.



In the modern stage or **Modern era** (from 2004 to the present) we see how the executives of NBA teams start using the SportVU technology, previously mentioned, from the 2010/11 season (Dallas Mavericks, Houston Rockets, Oklahoma City Thunder and San Antonio Spurs). In 2014 for football fans, the first *OptaPro Analytics forum* was created in London. In this Modern era, sports professionals are seduced by sport analytics and try to obtain more and more data.

We can say that we are on the verge of a **"future era"** (from the present to the next few years) in which sports organizations are beginning to worry that they have too much data and must think about what to do and how to act with it to use it well and make good decisions.

1.6. **Sports analytics history and its academics**

Statistics applied to sport started in baseball and, later, in basketball thanks to the famous book *Basketball on paper* by Dean Oliver (2004). (We will talk more about this author in the next modules).

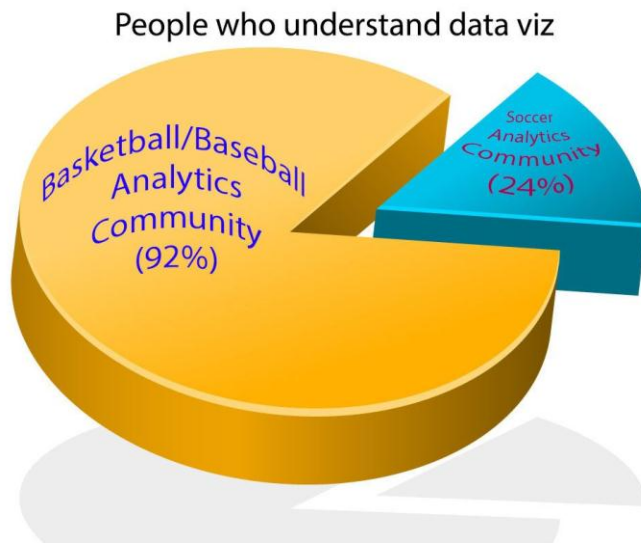
As previously discussed, the sports statistician or *sports analyst* may have a more academic role. In this sense, Tim B. Swartz (2018) wrote an article in which he mentions the great opportunities to perform sports analysis research. In this work, possible fields and specialized magazines are identified and some summary statistics are presented, and experiences and opinions are also commented. In this sense, he reminds us that, from the field of statistical departments, the application of theoretical statistics to sports data has not been considered correct. This work could not have been done without other contributions in academic terms that began several years ago (see figure 21).

Some of the academic *sports analysts* in effect in 2000 that activated this world: Ruud Koning, Jim Albert, David Berri, Jose, A. Martínez, Jaime Sampaio, Steven Caudill, and Jeremy Arkens.

Dutchman Ruud Koning, is one of the previously highlighted, who has focused on the world of *sports statistics & economics* in football. Another prominent academic is Jim Albert, one of the most influential personalities and statistician in the world of sports analytics who has been able to apply and teach statistics in baseball. Jim Albert has collaborated and published in different books and applications of statistics and sports. Other academics from the US more focused on basketball were Steven Caudill, Jeremy Arkes, or for example David Berri, who was the promoter of the winscore metric. Besides, two Europeans also focused on basketball analytics such as Jaime Sampaio and José A. Martínez helped to bring this discipline closer to Europe.

Data analysis has not only been applied in the sports industry, mainly in baseball or basketball, it can also be observed in the academic field. In fact, as Luke Born, another influential sports statistician in the world of *sports analytics*, informs us, the application of data analysis in football has taken more time to consolidate (see figure 22).

Figure 19: Sports Analytics: People who understand data viz.



Source: Luke Born (Sport Analyst of Sacramento Knings and exanalyst (Roma; soccer))

Probably, in the academic world one of the most important points has not only been research, but statistical education applied to sport. In this sense, it is worth mentioning the current work of Michael Lopez, current director of *data and analytics* in the National Football League (NFL) and also an associate professor of statistics at Skidmore College. His research covers causal inference and the application of statistics to sports and is accompanying young promises in this field, linking academia with the sports industry. Of the sports statisticians, he is probably the most important in the academic field, along with the aforementioned, Luke Born.

Curiously, when mentioning statistical education applied to sports, the great work of Keith Lyons cannot be left out, from another perspective, but key for those who want to keep up to date with *sports analytics*. Lyons promoted some years ago the *online* course called "Sport Informatics and Analytics" at Canberra University.

Michael Lopez revolutionised sports statistics and Keith Lyons revolutionised sports analytics education.

Other academics, already more familiar in *sports sciences* and *performance analysis* (who are not part of the *sports statistics and/or sports analytics* community according to Swartz, 2018) and interested in data analysis, have helped explain the applicability of this discipline on sport. In this way, they managed to get different members of the staff of a club or organization to ask themselves good questions regarding this topic.

For instance, European academics such as Jaime Sampaio, Carlos Lago, Miguel Ángel Gómez, Julen Castellano or Daniel Memmert, interested in positional data, have applied a bit what they learned from their predecessors such as Mike Huges or Ian Franks a little.

In this sense, if we already mentioned how Rob Carroll had influenced the evolution of *performance analysis* (listen to the interview via podcast), from sports science and complex systems there has been an important approach towards sports analytics. One of the people who has promoted this approach is the Australian Sam Robertson, a currently very active sports scientist and expert in machine learning (ML) applications for practical problems in high performance sports. Sam also works to improve decision-making for athletes and organizations through tracking data analytics and is an associate professor of sports science at Victoria University (VU). In the following link we can see him in an interview about the future of *sports sciences* and he mentions the following: “*We’re in an interesting time in Australian sport. We’re at a stage where we’re trying to find a happy medium on how much to spend on analytics and sports science*” (Sky News Australia [@skynewsaustralia], 2019, <https://twitter.com/SkyNewsAust/status/1154938074985459712?s=19>).

When talking about sports performance, from another perspective (athlete health and injury prevention), there has recently been more concern in the NBA, for example. There are also different academics such as Rasmus Nielsen, Caroline Finch, Martí Casals, among others who have focused more on sports epidemiology and biostatistics. In this sense, Rasmus Nielsen focuses on the use of applied biostatistics to improve athletes' performance and health.

When referring to the academic world and the sports industry, we cannot leave out the distance between research (academics) and practice (*practitioners* or *staff members*). This *gap* in recent years has become narrower and it is normal to see collaborations between sports and research institutions.

Finally, it is worth mentioning that the *data science* team of the sports science department of Fútbol Club Barcelona, betting on innovation and the transmission of knowledge from the Barça Innovation Hub platform and taking advantage of the celebration of the 2nd edition of Barça Sports Analytics Summit, will launch a football

analytics guide with the aim of giving a current vision about how the use of data is revolutionizing the analysis of sports performance. For this, the work of different researchers at international level in sports science and analytics in collective sports was gathered together in order to offer a learning space regarding the possibilities offered by the analysis of sports data. The guide has a series of chapters that describe the latest technological advances, current and available statistical models, algorithms, or metrics, and how the visualisation of this information also plays an important role. Besides, all this scientific vision will be accompanied by interviews with professionals and experts in the sports industry. It concludes with the mention of future lines of research, as well as the closest horizon in *football analytics*.

References

[Untitled image of How SportVU Technology works], 2017. Taken from <http://www.mbadmb.com/2017/04/20/23316/>

[Untitled image of described abilities in the web]. (2019). Taken from <https://thevideoanalyst.com/>

[Untitled image of Keith Lyons] (2013). Taken from <https://keithlyons.me/blog/2013/11/20/a-charles-reep-project/>

[Untitled image of Ben Alamar's book] (s.f.). Taken from <https://cup.columbia.edu/book/sports-analytics/9780231162920>

[Untitled image of Kirk Goldsberry's book], (s.f.). Taken from <https://www.amazon.com/SprawlBall-Visual-Tour-New-Era/dp/1328767515>

[Untitled image of *sabermetrics, the research of advanced statistics in baseball*] (2013). Taken from <https://georgetownssportsanalysis.wordpress.com/2013/09/22/a-simple-guide-to-basic-sabermetrics/>

Alamar, B. C. (2013). *Sports analytics: A guide for coaches, managers, and other decision makers*. Printing site: Columbia University Press.

Alamar, B. y Mehrotra, V. (2011). *Beyond 'Moneyball': Rapidly evolving world of sports analytics, Part I*. Taken from <http://analytics-magazine.org/beyond-moneyball-the-rapidly-evolving-world-of-sports-analytics-part-i/>

Azcona, N. (December 1st, 2015). Charles Reep pioneer of data analysis in football [blog entry]. Taken from <http://sportics.es/charles-reep-precursor-del-analisis-de-datos-en-el-futbol/>

Barboza, S. D., Bolling, C. S., Nauta, J., Van Mechelen, M., y Verhagen, E. (2017). Acceptability and perceptions of end-users towards an online sports-health surveillance system. *BMJ Open Sport & Exercise Medicine*, 3(1).

Birnbaum, P. (undated). *Guide to Sabermetric Research*. Taken from <https://sabr.org/sabermetrics>

Carling, C. et al. (2005). *Handbook of Soccer Match Analysis*. Oxford: Routledge.

Casals, M. (20 de agosto de 2018). *The importance of biostatistics in sports science and sports medicine*. Taken from <https://barcainnovationhub.com/es/bioestadistica-ciencias-del-deporte/>

Casals, M., y Finch, C. F. (2016). Sports Biostatisticians – a critical member of all sports science and medicine teams for injury prevention. *Injury Prevention*, 23(6), 423-427.

Coles, J. (2014). Comment on what is Performance Analysis or Analytics? Accessed on 27/12/14. Available online at: <http://www.visualperformanceanalysis.com/sports-blogs/performance-analysisanalytics/>.

Chan, L. (2019). *How coaching can benefit from performance analysis*. Taken from <https://footballperformanceanalysis.com/2014/02/19/how-coaching-can-benefit-from-performance-analysis/>

Dugalić, S. (2018). Sport, media and digitalization. *Sport–Science & Practice*, 8(1), 56-69.

Eaves, J. S. (2015). A history of sports notational analysis: a journey into the nineteenth century. *International Journal of Performance Analysis in Sport*, 15(3), 1160-1176.

Fullagar H. H, McCall A., Impellizzeri, F. M, et al. (2019). The Translation of Sport Science Research to the Field: A Current Opinion and Overview on the Perceptions of Practitioners, Researchers and Coaches. *Sports Medicine*. 1-8

Goldstein, G. (20 de abril de 2017). *La Big Data Sixième homme des parquets?* Taken from <http://www.mbadmb.com/2017/04/20/23316/>

Hoffman, B. (October 2nd, 2012). Trying to Build for the Future, the Athletics Win in a Hurry. *New York Times* [digital issue]. Taken from <https://www.nytimes.com/2012/10/03/sports/baseball/athletics-trying-to-build-for-future-land-in-the-playoffs.html>

Hunter, L. (21 de julio de 2014). *How Sports Data Analytics Is Upsetting The Game All Over Again*. Taken from <https://www.fastcompany.com/3033324/how-sports-data-analytics-is-upsetting-the-game-all-over-again>

Hüttermann, S., Noël, B. y, & Memmert, D. (2018). Eye tracking in high-performance sports: Evaluation of its application in expert athletes. *International Journal of Performance Analysis in Sport*, 17(2), 182-203.

James, B. (2001). *The Bill James Historical Baseball Abstract*. Edit. Print (Book). United States.

Lewis, M. (2003). *Moneyball: The Art of Winning an Unfair Game*. W. W. Norton & Company.



Lindsey, G. R. (1959). Statistical Data Useful for the Operation of a Baseball Team. *Operations Research*, 7 (2), 197-207.

McGarry, T., O'Donoghue, P., & y de Eira Sampaio, A. J. (Eds.). (2013). *Routledge handbook of sports performance analysis*. Routledge.

Nassis, G. P. (2017) Leadership in science and medicine: can you see the gap? *Sci Med Football*. 1(3), 195-196.

Oliver, D. (2004) Basketball On Paper: Rules and Tools for Performance Analysis. Ptomamac Books, inc.

Peake, J., Kerr, G. K. y, & Sullivan, J. P. (2018). A critical review of consumer wearables, mobile applications and equipment for providing biofeedback, monitoring stress and sleep in physically active populations. *Frontiers in Physiology*.

Perin, C., Vuillemot, R., Stolper, C. D., Stasko, J. T., Wood, J. y, & Carpendale, S. (2018). State of the art of sports data visualization. *In Computer Graphics Forum*, 37(3), 663-686.

Schwarz, A. (2004). *The numbers game: Baseball's lifelong fascination with statistics*. Macmillan.

Sky News Australia [@skynewsaustralia] (2019). We're in an interesting time in Australian sport. We're at a stage where we're trying to find a happy medium on how much to spend on analytics and sports science [Publication on Twitter]. Taken from <https://twitter.com/SkyNewsAust/status/1154938074985459712?s=19>

Stein, M., Janetzko, H., Seebacher, D., Jäger, A., Nagel, M., Hölsch, J. y Grossniklaus, M. (2017). How to make sense of team sport data: From acquisition to data modeling and research aspects. *Data*, 2(1) 2.

Steve Haake (2018). *Advantage play: technologies that changed sporting history*. Editor: Birlinn General.

Swartz, T. B. (2018). Where Should I Publish My Sports Paper? *The American Statistician*, 1-6.

