

# Module 2. Real Problem-Based Learning (PBL) in the sports analytics phenomenon (I)

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.

## 2.1. Introduction to the module

Nowadays, learning everything that sports analysis and the application of statistics cover in introductory courses like this is not an easy task. Therefore, the objective of this module will be to know and introduce *Sports analytics* concepts, by means of a problem-based learning (PBL) exercise. PBL is a learning methodology that comes from a Canadian university, where the student learns concepts in a more autonomous way, working on a specific case or problem (Servant-Miklos, 2018). PBL is specifically a pedagogical approach that uses cases and problems as a starting point to acquire the desired learning objectives. The role of a tutor in problem-based learning is different from the role of a course instructor in a didactic or conference-based course, which is more traditional.

As previously discussed in module 1, many interested in *Sports Analytics* want to analyse data, visualize and interpret it, but one of its keys is to try, first, to know what the problem really is and, later, to try to ask good questions. For this, in this module a fictional case will be presented where, from this, *Sports Analytics* concepts will be incorporated.

## 2.2. CASE: Pedro, candidate for analyst at Barça B

This case is a text that explains the selection test to which Pedro is exposed, a boy who wants to work as a quantitative football analyst for the F.C. Barcelona B. You do not have to think that you are Pedro, you have to read, first of all, the whole case and, later, in each paragraph, write down all the words, concepts or phrases that you think you have not completely assimilated, that are related to sports and data analysis. In fact, the initial



objective will be to try to extract the best questions from concepts related to sports and statistics in each of the paragraphs in this case. From these “questions”, attempts will be made, autonomously, to provide clues in order to adequately answer them in everything related to the phenomenon of *Sports Analytics*.

## Text of the CASE: Pedro, candidate for analyst at Barça B



**Figure 1: Pedro**

My name is Pedro, I am an ex-amateur football player, ex-player of chess and professional poker, apart from having completed my career in Physical Activity and Sports Sciences. During the last year I have wanted to expand my knowledge of video analysis with an introductory online *data science* course. After updating my profile on *LinkedIn*, *Researchgate* and *Twitter*, the FCB's chief football analyst called me directly for a job interview.

It is already my third interview and right now I am on Floor 3 of the F.C. Barcelona Ciutat Esportiva, in front of 7 responsible people (head of the area, data scientist, tactical analyst, academic *sports scientist*, head of *scouting*, head of the medical area *Innovation Hub* and a data journalist) from the *Sports Analytics* department of the sports science area.

At the beginning of the interview, the person in charge of this area asks me my opinion about football clubs that are starting to have *Data Science* departments and how I think statistics can help them.

Minutes later, the data scientist talks to me about the importance of being familiar with research, and how it adapts to *Big Data* and *Data Science* phenomena. He also explains me that his group meets every Tuesday to talk about some scientific articles of interest related to football. He tells me that, at the beginning of the league, he was surprised by the number of yellow cards and, for this reason, he would like to know how they could influence football. I quickly tell him that a study that appeared in the *Journal of Quantitative Analysis in Sports* (Anders, A. and Rotthoff, K. W. (2011) a few years ago sounds familiar to me. "Yellow Cards: Do They Matter?" *Journal of Quantitative Analysis in Sports*, 7(1)).

The data scientist is in a bit of a hurry and introduces me to one FCB tactical or game analysts of the first team. This provides me with a database (validated by the IT team) of the last 15 league matches of all the teams (they have not been able to obtain the first 10

matches), where there are different characteristics of the teams (example: Stadium, Home / Away), initial squad (example: 4-3-3, 4-4-2), possession percentage, number of shots on the left, number of shots on the middle, number of shots on the right, conversion rate, number of corners, number of fouls, final result, so I can see some patterns and describe some behaviours. The analyst adds that he does not have a specific question to ask me, but for a long time he has wanted to know if there is an association between the fact of scoring before the opponent and the final result of the match.

Later they introduce me to an analyst with a more academic profile and, as test, he gives me a sheet with the following *abstract* from a scientific study (Mitrotasios, M. and Armatas, V. (2014). *Analysis of goal scoring patterns in the 2012 European Football Championship*. United States Sports Academy). Based on reading it, he invites me to chat with him over coffee.

**ABSTRACT Purpose:** The aim of the present study was to investigate the factors associated with goal scoring in top class football matches. **Methods:** 76 goals analysed from 31 matches from the European football Championship in Poland-Ukraine in 2012. The Sport Scout performance analysis tool was used to time code when goal scored and the associated behaviours relating to goal scoring. **Results:** It was presented that 57.9 % of the goals were scored in the second half and most of them in the last 15-min period, while the first goal seems to play an important role in the final result of the match. Most goals were scored during open play (72.4%) but 27.6% scored after a set play, mainly following corner kick and free kick. Concerning initiation zone of the attack, it was presented that 56.6% began from the attacking half. Most goals were scored after a cross (43.7%) while regarding goal scoring area, over 90% of goals were scored inside the penalty area and specifically from the area between penalty spot and goal zone (42.1%). Finally, almost ½ of goals were scored without any pressure from defenders. **Conclusions:** The present study presented guidelines to coaches in order to design trainings similar to the actual competition and adapt the game style with regard to match status. Moreover, coaches can use this information to establish objectives for players and teams with a special reference to the offensive or defensive play.

I have been on Floor 3 for about 3 hours now and they introduce me to the analyst with a *scouting*, profile, who seems concerned because, lately, the team's results have not been as expected, and he does not know what the effect of the change of coach on professional football teams is. Besides, he is also interested in knowing the goalkeepers' performance. He tells me that he wants to know, for example, if the number of passes

goalkeepers make away from home may differ if they play at home. And from that moment on, he shows me the following tables of some studies that were previously performed for me to analyse them with them.

When talking about performance, the medical chief of F.C. Barcelona *Innovation Hub* Barcelona Innovation Hub and the club's data journalist join the interview and show me the following photo about injuries from the English league for me to do, briefly, my final in-depth analysis.

**Figure 2: Injuries from the English league.**



Source: [Untitled image of injuries from the English league]. 2016. Taken from <http://sportsdiscovery.net/journal/2016/08/05/interview-head-of-sports-science-at-swansea-city-afc-jonny-northeast/>

It is already 7 p.m., and I hope I made a good impression on them and I can join this team of analysts.

### 2.3. Analysis of the 1<sup>st</sup> paragraph of the CASE

After reading the first paragraph carefully, I am going to put in bold the words that can help me ask questions to consolidate some concepts indirectly related to *Sports Analytics*. As the first example, one of the first questions to ask about the first paragraph would be:

“My name is Pedro, I am a **former** amateur **football player**, ex-player of **chess** and professional **poker**, apart from having completed my career in **Physical Activity and**



**Sports Sciences.** During the last year I have wanted to **expand my knowledge** of **video analysis** with an introductory **online *data science* course**. After updating my profile on **LinkedIn, Researchgate and Twitter**, the FCB's chief football analyst called me for a job interview."

### **1. What skills, knowledge and training can help you become a better analyst in the future?**

Had a similar question been asked? Maybe not, but you have probably thought of similar questions that make you wonder how to answer and learn more about it. For a few minutes you can think about asking the following possible questions. After these minutes, we present you some that have been thought reading the first paragraph. Let us hope they are similar, and you are interested in answering them to be able to delve more into the phenomenon of *Sports Analytics*.

1. **What skills, knowledge and training can help you become a better analyst in the future?**
2. **What is *Data Science*?**
3. **What is the difference between *Data Science* and *Science*? And between *Data Science* and *Statistics*? Learn new means of knowledge of analysis.**
4. **Can having knowledge and making good use of social networks help me?**

We will try to answer the questions.

### **1. What skills, knowledge and training can help you become a better analyst in the future?**

There is no specific skill, knowledge, training, or recipe, but indeed there are several experiences and tips that can help. First of all, we have to ask ourselves what the general idea is of a professional dedicated to being a sports analyst. The following figure can give us some clue.

Figure 3: Some people's vision of what a performance analyst represents.



Source: [Untitled image of some people's vision of what *performance analyst* represents]. 2012. Taken from <http://thevolleyballanalyst.blogspot.com/2012/02/performance-analyst-perception.html>

The previous figure makes us think that, probably, the current skills of the analyst today are not well known. The evolution of this profile has been important, and it is necessary to take into account several of these skills. As some of the characteristics of the football analyst (also applicable to other sports) we could mention:

1. Knowing football in depth (not only tactical or technical aspects, but also physical, psychological, social).
2. Have knowledge of the squad and coaching staff to which it belongs and the competition in which the club is involved.
3. Being entrepreneurial and technological. It is essential that they know how to adapt to all the software options, platforms, and tools existing today.
4. Being a quantitative and qualitative expert, capable of making sense of the amount of statistics and numbers related to football, to correctly communicate to the squad and the coaching staff.
5. Be aligned with the coach and the club regarding the game model.

And what can these professionals analyse? They can really focus their objectives on **physical performance**, technical performance (example: passes and zones where they

are made, shots on goal, any technical event), **tactical performance** (example: possessions made, zones and characteristics; set pieces actions, for, against; relationships between players), psychological performance, injury prevention and much more.

The student taking this course is probably wondering: "Do I have to know everything?" In the beginning, the analyst only took photos or **video screenshots**. This involved many hours of work but, thanks to new video analysis *software* available, this work became easier (ex: Nacsport, Longo match, ERIC Sport, Sportcode). **New technologies**, the information boom and the **professionalisation** of sport in general have created the need to ask ourselves how to train better analysts. Many do not know if I am a video analyst? Or a photo analyst? Tactical or performance analyst? Data analyst? This gives us an idea of the evolution of profiles that may there be in a sports analysis department in the same football club.

Next, we will do an activity to see if you have an idea of the different profiles of an analyst. If we look at the following photos, what coach profile would be related to, for example, *Football Analytics*?

**Figure 4: One of them is a trainer related to Sports Analytics.**

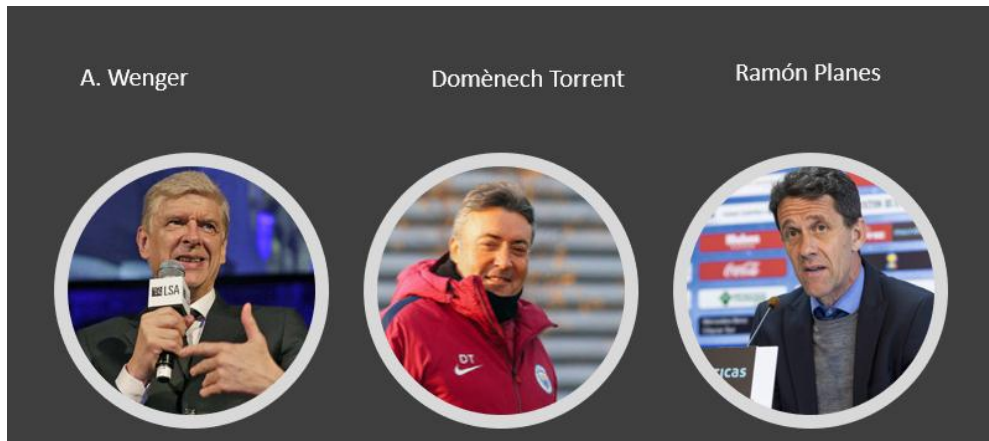


Source: own elaboration.

The right answer would be Jürgen Klopp.

And the analyst with a *scouting* profile?

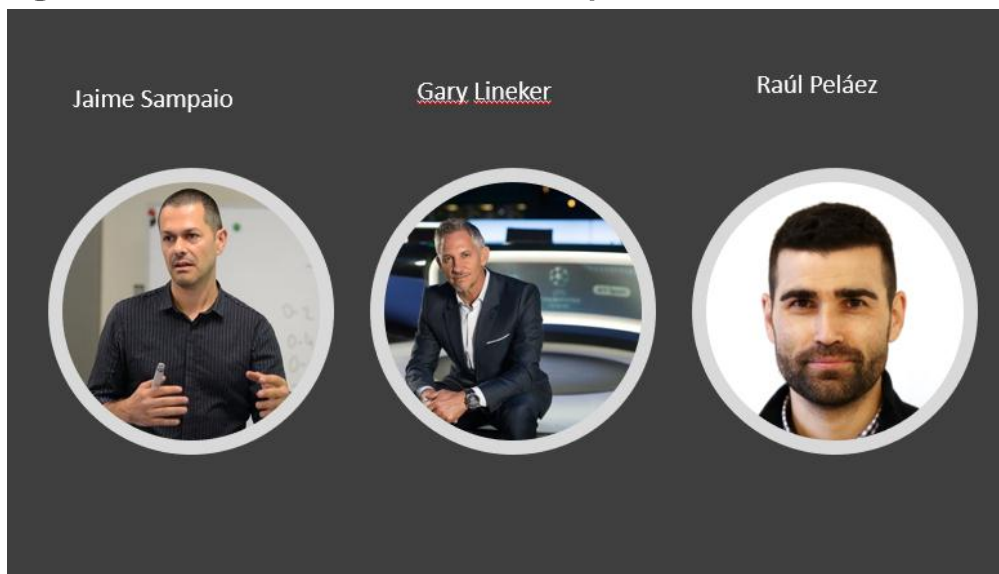
**Figure 5: One of them is an analyst with scouting profile.**



Source: own elaboration.

The right answer would be Ramón Planes if we ask ourselves this question in the 2019/20 season. If we are interested in knowing a more academic analyst profile and we look at the following figure, we could say that the example profile could be Jaime Sampaio.

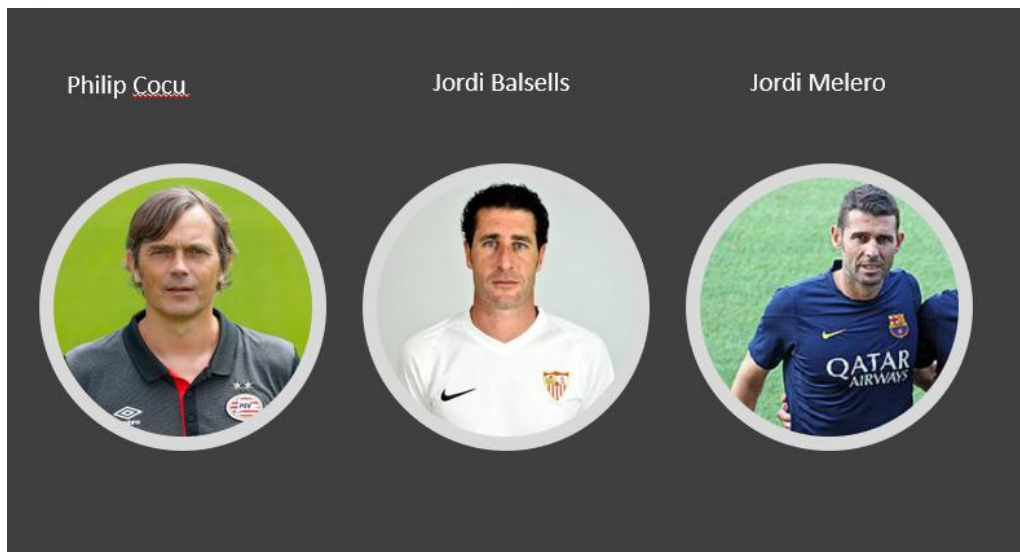
**Figure 6: One of them is an academic analyst.**



Source: own elaboration.

Another possibility would be to have a more tactical or game analyst profile in a club.

**Figure 7: One of them is a tactical or game analyst.**

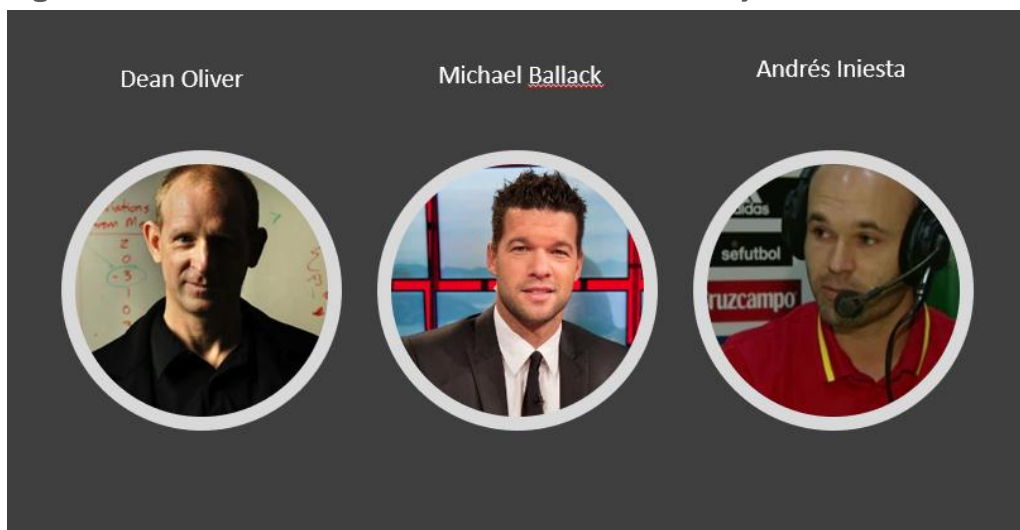


Source: own elaboration.

The right answer would be Jordi Melero if we ask ourselves this question in the 2019/20 season.

And can we find commentator analysts? Yes, one of them, world famous, is an ex-football player and current specialized commentator, who made his debut on ESPN, Michael Ballack.

**Figure 8: One of them is a football commentator analyst.**



Source: own elaboration.

Last but not least, we could ask ourselves if there are data analysts or statisticians with scientific profiles. In this profile we have not placed any figure, as it is a profile with a great evolution and trajectory, and that we will surely learn from them soon. It is relevant to remember that these experts with more specific skills (example: statisticians,

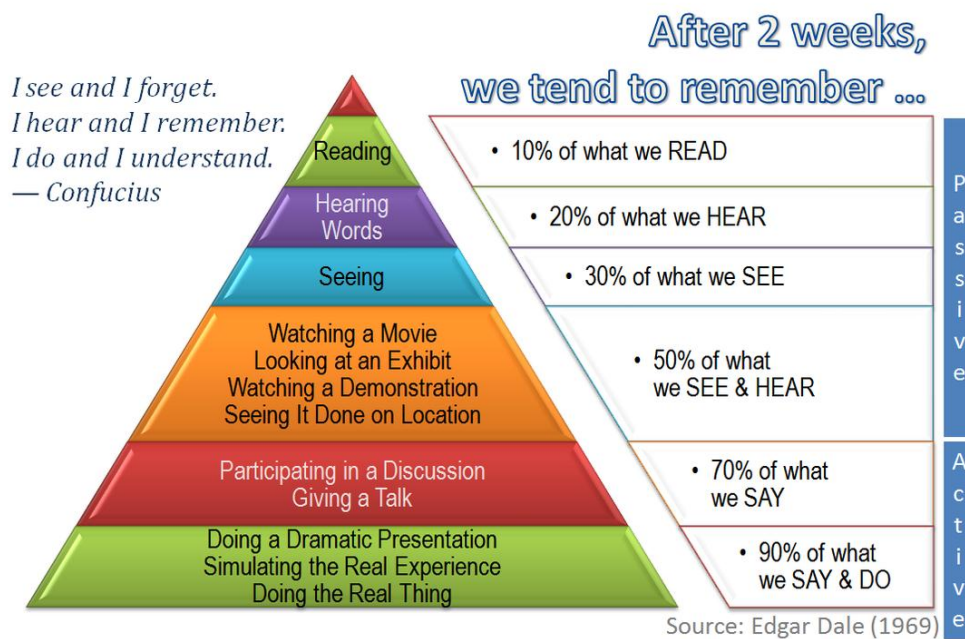
mathematicians) have to be always involved from the beginning of the study and that it is possible to resort to them as experts whenever necessary.

In this first question, we have seen what possible skills a sports analyst may have and what specialities there may be in the world of football, as an example. Another point we want to highlight to answer our initial question: **“What skills, knowledge and training can help you become a better analyst in the future?”** is about training. In this sense, many times, when we attend training courses like this, we must take into account how much information learned can be retained and used from now on. That is why it is interesting to think of Edgar Dale's pyramid from 1969, which basically informs how we remember, a little passively and actively.

**Figure 9: Learning pyramid.**

## The Cone of Learning

sparkinsight.com



Source: [Untitled image about learning pyramid] undated Taken from <https://eingleses.com/la-piramide-del-aprendizaje/>

This is very interesting because, basically, it makes you think, for example, what idea you get about the *Moneyball* phenomenon after seeing the film or, on the contrary, reading its book. We also see the importance of profiles such as speakers, teachers or experts who have to explain concepts of interest in their presentations, since their learning possibilities can be high.

In summary, and answering the first question, if you want to dedicate yourself to sports analysis you have to:

1. Know very well the **sport**.
2. Have a **mentor**.
3. **Train constantly**.
4. Have great skills in **statistics, analytics, and computer programming**.
5. Adequately know some **statistical analysis** (R, SAS, Stata, Python, Matlab), **data management** (Python, SQL, GitHub), **visualization** (Tableau, Power BI, Qlikview, Plotly) **software**.
6. **Have communication skills according to Bucheit (2017)**.

Afterwards we will cover the following questions that are closely related.

## 2. What is *Data Science*?

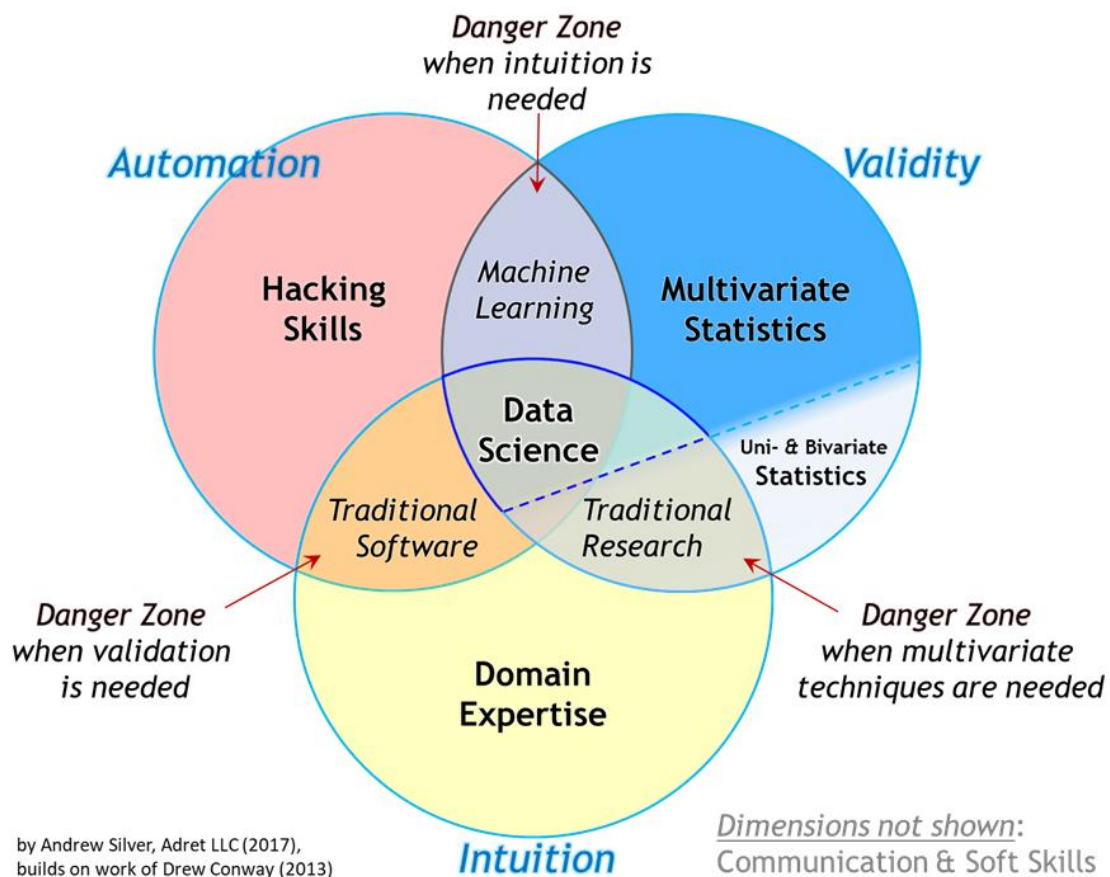
### 3. What is the difference between *Data Science* and *Science*? And between *Data Science* and *Statistics*? Learn new means of knowledge of analysis.

In 2012, the *Harvard Business Review* magazine already commented on what would be the sexiest job of the 21st century: *Data Scientist*. From the beginning, it has been thought that *Data Science* (Varian, 2009; Davenport y Patil, 2012; Paarlberg, 2015) could be easily defined as:

- Data science is a **sexy job. The salaries are high**, the work is interesting and there's significant prestige that comes with the title. As a result, many people want to be data scientists.
- A data scientist is someone who is **better at statistics than any software engineer** and **better at software engineering than any statistician**.
- The term data science has caused **excitement, confusion, and controversy**. Some of the confusion is from the lack of a consistent definition. There is an **ecosystem of related terms (e.g. analytics, business intelligence, big data, and data mining)**.
- The problem is that the **field has grown too fast**: there are now far too **many data scientists with very little experience** heading into a job market that has very few experts. Times have changed. When the term was still relatively new, companies were accepting candidates with only a basic knowledge of data and getting them to learn on the job.

The *Data Science* approaches described above are not correct and it is precisely the following Venn diagram described by Drew Conway (2013) and Andrew Silver *et al.* (2017) that shows us what *Data Science* encompasses and also its “dangers”.

**Figure 10: De Venn’s *Data Science* diagram.**



Source: Silver, A. 2017.

We see that defining *Data Science* is more complex than it seems, and the following approaches would be more real:

- A data scientist is someone who **uses data to solve problems**.
- **A multidisciplinary field.** Data scientists come from a variety of academic backgrounds: computer science, physics, statistics, and many others. What matters most is **having a creative mind coupled with first rate critical thinking skills**.
- Intersection of **three areas: maths/statistics, computation, particular domain** (e.g.: Sports) (Conway, 2010; Blei and Smyth, 2017).

- Intersection of **six areas called “Greater Data Science”**: 1- Data gathering, preparation and exploration; 2- Data representation and transformation; 3- Computing with data; 4- Data modelling; 5- Data visualization and presentation (Donoho, 2017).
- **I do not know.** Data science is an awesome profession, but there are definitely some serious frustrations that come with it.
- **“Data Science” could disappear, to be eclipsed by the next buzzword.**

One of the main differences between *Science* and *Data Science* is that “**Science** starts with a **question**; **data science** starts most of times with the data” (Carmichael and Morrison, 2019; Donoho, 2017). *Data Science* is a global process and its profession is subspecialising in *data scientist* engineers, statisticians, or *data scientist* analysts, etc. More information on this topic can be found in the literature of Carmichael, I. and Marron, J. S. (2018), Donoho, D. (2017), or even in Healy if we go back to 1978. There is currently good information on the vision of *statistical thinking* in *Data Science* (Ollé, Casals and Fernandez, undated).

#### 4. Can having knowledge and making good use of social networks help me?

According to the current literature, it is important to know how to make good use of social networks and share knowledge, since you can share knowledge with different groups of people related to your interest and create synergies and even obtain opportunities for a new job.

This presentation via *slideshare* (see: <https://es.slideshare.net/xavierlasauca/science-dissemination-20-social-media-for-researchers-126297683>), by Xavier Lasauca, shows its real possibilities as an example.

#### 2.4. Analysis of the 2<sup>nd</sup> paragraph of the CASE

Let us read carefully the following paragraph of the CASE. Now that we know how this PBL methodology works, we will try to ask some questions. After a few minutes, you can see below some of them (it does not mean they have to be exactly the same), but they can probably help.

“It is already my third interview and right now I am on Floor 3 of the F.C. Barcelona Ciutat Esportiva, in front of 7 responsible people (head of the area, data scientist, tactical analyst, academic *sports scientist*, head of *scouting*, head of the medical area *Innovation*

*Hub* and a data journalist) from the *Sports Analytics* department of the **sports science** area. At the beginning of the interview, the person in charge of this area asks me my opinion about football clubs that are starting to have **Data Science departments** and how I think **statistics** can help them."

Based on the paragraph the following questions arise:

**5. What is *FCB Innovation Hub*?**

**6. What is *Sports Analytics? Soccer Analytics*?**

**7. What is *Sports Sciences? Science and research? Scientific method*?**

**8. What does having a *Data Science* department mean?**

**9. What is statistics? And how can it help?**

**5. What is *FCB Innovation Hub*?**

*FCB Innovation Hub* is a great sports laboratory for the future of sports. If you have registered for this course, you have probably visited the web platform (<https://barcainnovationhub.com/es/>) and its blog and news that are being updated.

**6. What is *Sports Analytics? Soccer Analytics*?**

We believe that, in the previous module, it has become clear what *Sports Analytics* represents. If we focus on soccer or *football analytics*, we realize that each sport has its own complexity. In 2016, Masoud Nikravesh mentioned *Sports Analytics* as:

The processes that identify and acquire the knowledge and insight about potential players' performances based on the use of a variety of data sources such as game data and individual player performance data. These advanced and sophisticated type of analytics should be able to extract valuable actionable insights for the coaches and managers to utilize.

And about *Soccer Analytics* as:

It is the art of creating insights and actionable decisions using soccer related data. While predictive analytics uses big data to determine the probability or the likelihood of a certain outcome, intelligent descriptive analytics looks at big data and analyses it using machine learning and artificial intelligence methods to come off with suggestions that will improve the likelihood of a desired outcome.

Books like the one written by Ben Alamar (2013) probably bring us closer to these two phenomena. In any case, if we focus on *football* or *soccer analytics*, in the last two decades there has been a change in how performance in sports is observed, recorded,



and analysed and, obviously, also in football. The process from video analyst to quantitative analyst is very well detailed by Rob Carroll on his website: (<http://thevideoanalyst.com/contact/>).

Now we are less used to hearing about notational analysis, there is more talk of *analytics*. In 2011, Ben Alamar comments that *Sports Analyst* has three components: *Data management, Predictive models, and Information systems*. Jeremy Abramson, already in 2014, defines *Sports Analytics* as “the discovery and communication of meaningful patterns in data”. Bill Gerard, in 2016, understands *Sports Analytics* as “analysis of tactical data to support tactics-related sporting decisions”. And late, Ben Alamar, in 2018, defines it as “Data and analysis are reducing the subjectivity of player valuation, strategic decisions, and player development”. Another illustrious *basketball analytics*, Dean Oliver answers Ben Alamar saying: “Reducing the subjectivity is a good phrase. You don't eliminate subjectivity (and don't really want to), but there are issues where analytics definitely limits errors subjectivity can make”. These first definitions and messages of these authors lead us to think that, probably, it is a phenomenon in continuous development. Bill Gerard gives us some tips to take into account: (<https://winningwithanalytics.com/2016/12/05/bridging-the-gap-improving-the-coach-analyst-relationship-part-2/>).

- **Start simple** when first introducing data analytics as a **coaching** tool.
- Analytical results are usually presented most effectively to coaches by using **data visualization and story-telling**.
- Data analytics is **only one input into decision making by coaches**, albeit a potentially very important one if **used effectively**.
- Data analytics is **suffering from a fixation with big-data analytics**. Big-data, context-generic statistical analysis **must be translated into practical solutions to small-data**.
- **Sports analytics is most effective when the analyst understands the specific operational context of the coach**, produces relevant data analysis, and translates that analysis into practical recommendations.

In summary, the current reflections for the *soccer analyst* and according to Brian Prestidge would be:

**The analysts are getting smarter.** Instead of relying on generic data — such as shots on target, possession statistics or heat maps— from companies such as Prozone and Opta, **they are creating their own**



**metrics and building statistical models/algorithms to fit the club's philosophy and tactics.** It is this approach that could be described as a brains race.

### 7. What is *Sports Sciences*? Science and research? Scientific method?

**Science** can be defined as “the process of observing, asking questions, and finding things, out about the world around you”.

In order to research, it is important to have the ability to ask oneself questions and use the **scientific method**. Therefore, no intuitive knowledge or unsupported assumption will be useful. The scientific method is the succession of steps (observation, induction, hypothesis; test or test of the hypothesis by experimentation; demonstration or refutation of the hypothesis; and thesis or scientific theory) that we must take to discover new knowledge and to verify unknown hypotheses up to now. Research and the stages of the scientific method are related to the main sections that we often find in a scientific article (see Chart) (Mabrouki, & Bosch, 2007).

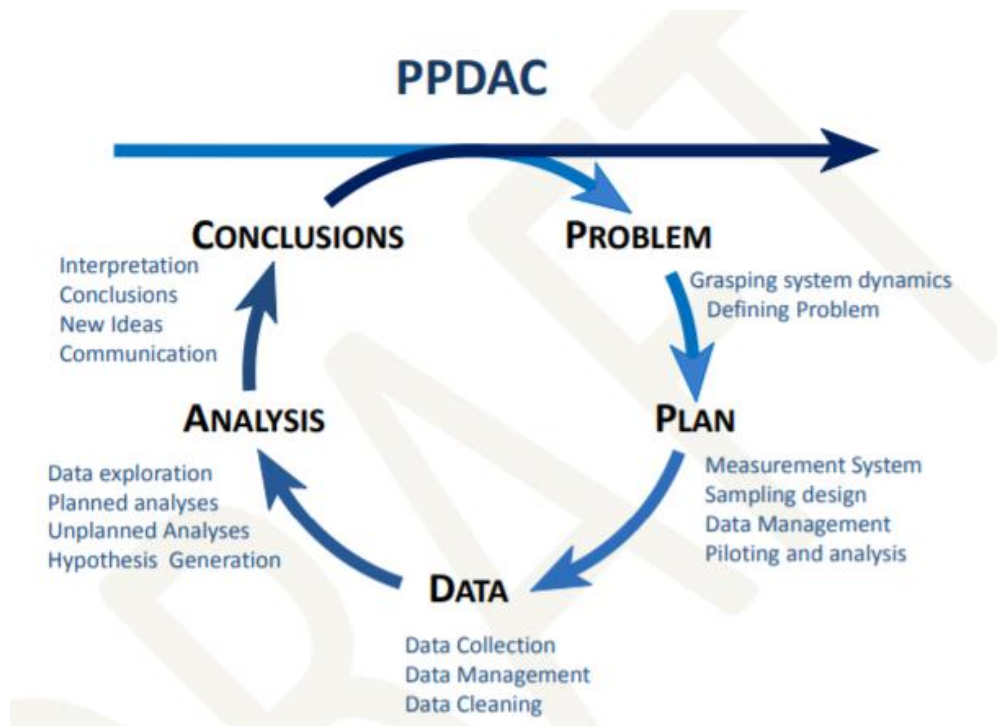
**Chart: Stages of the Scientific Method.**

STAGES IN THE SCIENTIFIC METHOD	MAIN SECTIONS OF A SCIENTIFIC ARTICLE
Understand the problem to be studied	Introduction
Formulate a hypothesis	
Collect data	Material and Methods
Analyse data	Results
Interpret results	Discussion
Conclusions	

Source: own elaboration

The scientific method is linked to the PPDAC (*Problem-Plan-Data-Analysis-Conclusion*) that was mentioned in module 1 and that refers to the following figure by David Spiegelhalter.

Figure 11: The PPDAC cycle.



Source: <https://www.pinterest.es/pin/391320655096691083/?lp=true>

A key aspect to research is knowing how to ask good questions and, from there on, apply the scientific method, knowing the different designs well, using the correct methodology and, above all, following consensus guidelines (for example, in the field of Medicine: Strengthening the Reporting of Observational Studies in Epidemiology (STROBE), Consolidated Standards of Reporting Trials (CONSORT), Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), WHO Injury Surveillance Guidelines).

Anyway, there is still room for improvement in science, as John Ioannidis (2005) and Ben Goldacre (2010) indicated in their respective works.

Defining *Sports Sciences*, and later *Sports Scientist*, is not easy. According to Xavi Schelling, Head of Department, Applied Sports Science (San Antonio Spurs; NBA):

An applied sports scientist is a multidiscipline integrator, who uses the scientific method to extract the signal from the noise. He/she will ask the right questions to the right people (experts) and will explain the new knowledge to the decision-makers in understandable terms.

The different roles of a staff and, specifically, of a *sports scientist*, have been evolving in recent years as a result of technological development itself. In one of Casals (2019)

articles, where an analogy between basketball and a multidisciplinary research team is mentioned, several roles of a possible new staff are described, among them, the *sports scientist*.

### **8. What does having a *Data Science* department mean?**

A few years ago, when we visited the post on the *nbastuffer* website titled: "NBA Teams That Have Analytics Department", it was difficult to find more than one or two people in each NBA franchise. The *Sports Analytics* or *Data Science* departments have been growing recently, they have incorporated a specialized academy (*statisticians, mathematicians, computer scientists*) in the sports industry. NBA franchises, such as Raptors, Sixers, Spurs, among others, have helped that. It seems that in football, different European clubs are, little by little, betting heavily. Good news is the incorporation of Michael López into the NFL as director of *Sports Analytics*, as he is possibly one of the best academic *sports analyst* up to now that can help bridge the gap between the industry and the academy of sports analytics.

### **9. What is statistics? And how can it help?**

The Royal Statistical Society was already asked this question in 1838. Do we have a unique answer? Can we speak of a single discipline or profession?

Guadalupe Gómez, Professor Statistics and Operations Research (Universitat Politècnica de Catalunya-BarcelonaTECH), already answered about some points of interest that a great part of society does not know; the profession of statistician exists as such.

What do statisticians do? Collect data, analyse it, collaborate with scientists at all stages of research, and develop theories and methods.

What forms does the statistic take? It can accept a scientific model derived from a mathematical theory, or it can build a model and take it as good and correct.

What role do statisticians play? Active/passive role in planning, active role in analysis or as support staff.

Where do statisticians work? In the world of sports, economics, public health, politics, psychology, etc. No wonder statistics are hard to find as a single, unified discipline.

Defining statistics, a young science, is not so simple. If Karl Pearson, in the years 1857-1936 defined statistics as the grammar of science, today some already define it more simply as the science that uses data to understand uncertainty about a process and make informed decisions making use of the probability theory. Anyway, the American Statistical Association (ASA) defines it as: "the science of learning from data, and of



measuring, controlling and communicating uncertainty". Most definitions mention "data" and "uncertainty", although it is difficult to have a single consensus when defining statistics.

According to ASA, **statisticians:**

- Help companies make sense of the world around us by **analysing** data.
- Use data to **solve** complex
- **problems**, in fields like **business**,
- **medicine, public service, sports...**
- and more.

Are you ready to read the next paragraph and try to ask yourself good questions to learn more content of this course?

### **Third paragraph:**

Minutes later, the data scientist talks to me about the importance of being familiar with research and how it adapts to *Big Data* and *Data Science* phenomena. He also explains me that his group meets every Tuesday to talk about some scientific articles of interest related to football. He tells me that, at the beginning of the league, he was surprised by the number of yellow cards and for this reason he would like to know how these could influence football. I quickly tell him that a study that appeared in the *Journal of Quantitative Analysis in Sports* (Anders, A. and Rotthoff, K. W. (2011) a few years ago sounds familiar to me. Yellow Cards: Do They Matter? *Journal of Quantitative Analysis in Sports*, 7(1)).

# References

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

**Anders, A. y Rotthoff, K.** (2011). Yellow Cards: Do They Matter? *Journal of Quantitative Analysis in Sports*, 7(1).

**Bik, H. y Goldstein, M.** (2013). An introduction to social media for scientists. *PLoS biology*, 11(4), e1001535.

**Blei, D. y Smyth, P.** (2017). Science and data science. *Proceedings of the National Academy of Sciences*, 114(33), 8689-8692.

**Bucheit, M.** (2017). Want to see my report, coach? Sport science reporting in the real world. *Aspetar Sports Medicine Journal*.

**Carmichael, I. y Marron, J.** (2018). Data science vs. statistics: two cultures? *Japanese Journal of Statistics and Data Science*, 1(1), 117-138.

**Caroll R.** (<http://thevideoanalyst.com/contact/>).

**Casals, M. y Nielsen, R.** (2019). Who and what can contribute to improve the statistical thinking in sports injury research? A humorous analogy between basketball and members of the multidisciplinary research team. *Apunts Medicina de l' Esport (English Edition)*, 54(203), 81-84.

**Conway, D.** (2010). The data science Venn diagram. (2010). <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>.

**Conway, D.** (2013). The data science Venn diagram.  
<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

**Davenport, T. y Patil, D.** (2012). Harvard Business Review. Data Scientist: the Sexiest Job of the 21<sup>st</sup> Century. October 2012. <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>

**Donoho, D.** (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4), 745-766.



**Felt, M.** (2016). Social media and the social sciences: How researchers employ Big Data analytics. *Big Data & Society*, 3(1), 2053951716645828.

**Goldacre, B.** (2010). *Bad science: Quacks, hacks, and big pharma flacks*. McClelland & Stewart.

**Healy, M.** (1978). Is statistics a science? *Journal of the Royal Statistical Society: Series A (General)*, 141(3), 385-393.

**[Untitled image about the learning pyramid].** undated Taken from <https://eingleses.com/la-piramide-del-aprendizaje/>

**[Untitled image of some people's vision of what performance analyst represents].** 2012. Taken from <http://thevolleyballanalyst.blogspot.com/2012/02/performance-analyst-perception.html>

**[Untitled image of injuries from the English League].** (2016). Taken from <http://sportsdiscovery.net/journal/2016/08/05/interview-head-of-sports-science-at-swansea-city-afc-jonny-northeast/>

**Mabrouki, K. y Bosch, F.** (2007). *Scientific writing in biomedicine. What needs to be known*. Barcelona, ES: Prous Science.

**Mitrotasios, M. y Armatas, V.** (2014). *Analysis of goal scoring patterns in the 2012 European Football Championship*. United States Sports Academy.

**Ollé, J., Casals, M. y Fernández, D.** (s.f). The figure of data scientific - myths and professional revolution. Taken from <https://investigacondatos.com/2019-dia2-charla05-cientifico-datos-marti-casals-daniel-fernandez/>

**Paarlberg, B.** (2015). The measurement standard. Why "Data Scientist" Is Being Called The Sexiest Job of The 21st Century. July 2015. Taken from: <http://www.themeasurementstandard.com/2015/07/why-data-scientist-is-being-called-the-sexiest-job-of-the-21st-century/>

**Servant-Miklos, V.** (2018). Fifty Years on: A Retrospective on the World's First Problem-based Learning Programme at McMaster University Medical School. *Health Professions Education*.

**Silver, A. et al.** (2017). The Essential Data Science Venn Diagram. Taken from: <https://towardsdatascience.com/the-essential-data-science-venn-diagram-35800c3bef40>

**Van Noorden, R.** (2014). Online collaboration: Scientists and the social network. *Nature news*, 512(7513), 126.

**Varian, H.** (2009). Flowing Data. Google's Chief Economist Hal Varian on Statistics and Data. Taken from <https://flowingdata.com/2009/02/25/googles-chief-economist-hal-varian-on-statistics-and-data/>

**Voytek, B.** (2017). Social Media, Open Science, and Data Science Are Inextricably Linked. *Neuron*, 96(6), 1219-1222.