Casals, M. & Daunis-i-Estadella, P. (2023). Violinboxplot and enhanced radar plot as components of effective graphical dashboards: an educational example of sports analytics. *International Journal of Sports Science & Coaching*, *18* (2), 572-583.

https://doi.org/10.1177/17479541221099638

Title of the Article:

Violinboxplot and enhanced radar plot as components of effective graphical dashboards: An educational example of Sports Analytics

Martí Casals,[1,2,3] Pepus Daunis-i-Estadella[4]

[1] Sport and Physical Activity Studies Centre (CEEAF), Faculty of Medicine, University of Vic-Central University of Catalonia (UVic-UCC), Barcelona, Spain

[2] Sport Performance Analysis Research Group, University of Vic-Central University of Catalonia (UVic-UCC), Barcelona, Spain

[3] National Institute of Physical Education of Catalonia (INEFC), University of Barcelona, Barcelona, Spain.

[4] Department of Computer Science, Applied Mathematics, and Statistics, University of Girona, Girona, Spain.

Martí Casals

Sport and Physical Activity Studies Centre (CEEAF), University of Vic – Central University of Catalonia (UVic-UCC), C. Dr. Antoni Vilà Cañellas, s/n, 08500 Vic, Spain.

ORCiD: https://orcid.org/0000-0002-1775-8331

Twitter: @CasalsTMarti

His research interests are Statistics of Education, Sports Statistics, Biostatistics and Statistical Thinking.

Pepus Daunis-i-Estadella

Department of Computer Science, Applied Mathematics, and Statistics, University of Girona, Girona, Spain.

ORCiD: https://orcid.org/0000-0001-6134-9255

Twitter: @enpepus

Abstract

A statistical graph can offer an alternative compelling approach to statistical thinking that focuses on important concepts rather than procedural formulas. Nowadays, visualizing multidimensional/multivariate data is essential but can also be challenging. In sport analytics, the exploration and descriptive analysis of data using visualization techniques has increased in recent years to, for example, describe possible patterns and uncertainty of player performance. These visualization techniques have been used so far with different purposes by various professionals in the sport industry, such as managers, coaches, scouters, technical staff, journalists, and researchers. The abuse of graphs, such as the radar plot, and their frequent misinterpretation in the world of sports and possible implications for coaching decisions has led us to create more informative and accurate visualizations. Here, we propose new, more educational visualizations we have termed violinboxplots and enhanced radar plot for their use in the sports analytics and other fields. These allow us to visualize, besides distribution and statistical summaries, the extreme data values that can be fundamental in performance studies and allow us to benchmark.

Keywords: sports analytics; statistical thinking; visualization; team sports; data science; statistical education

1. Introduction

Education and introduction to data visualization are important in statistics and data science courses [1–3]. Data visualization is a perfect vehicle for teaching both students and our society the importance of context in statistics while also teaching them graphical techniques [3]. One of the recommendations of the guidelines for assessment and instruction in statistics education (revised GAISE College report) is focused on multivariable thinking through examples of data visualization that we can find in the real world [4].

Data visualization is also increasingly being used in the world of sports and is attracting interest from various professionals, such as managers, coaches, analysts, video analysts, data journalists, doctors, physical trainers, physiotherapists, psychologists, nutritionists, players themselves and above all future students. Probably, the community of education of statistics and sports is needed without knowing it.

Sports statistics are often known as sports analytics, and their use has increased exponentially thanks to advances in computer science and a growing culture of applying these data to trigger organizational changes in sports teams. This higher use is also due to the need to bring industry and science closer, as other authors have described [5,6], and it has also become popular thanks to films such as *Moneyball*.

The evolution of sports analytics has led to more immediate data capture and optimization. Together with the big data phenomenon, it has led sports professionals to new needs in order to obtain, analyze, and, above all, interpret and communicate data. A current issue is the use of information and big data with different interests: business, science, or sports club-specific internal use. Obtaining useful information from the large amount of data available is key. This process will require the skills of multidisciplinary teams in the world of sports, rather than lone superstars, especially those who are experienced working with data. Software skills using tools such as Tableau, Power BI, or Qlik to create dashboards (final products) are necessary, but having an in-depth understanding of the problem or question at hand and knowing which methodologies are most appropriate depending on the type of data (taking into account its volume, variety, veracity, and value) are key to make decisions before arriving at a final product. Statistical science deals with this [7].

Sports analytics have so far focused mainly on metric analysis or descriptive statistics and, therefore, largely on exploration, data visualization, or descriptive graphs [8]. In this sense, despite major advances in the field of data visualization [9,10], statistics education is key to avoid misuse of descriptive statistics and to understand good practices before making decisions. If the academy knows typical concepts and errors in data analysis (e.g., spurious correlations, p-hacking, bias sample), so as not to make bad decisions or make what claims, the professionals in the world neither should the sport sciences [11]. In science, especially in the health field, following responsible research conduct and the use of guidelines are encouraged before undertaking research studies [12] in order to promote the reliability and value of the literature, even more so for its dissemination to a non-specialist audience.

Many scientific journals have recently introduced policies that encourage or require authors to use more informative figures that show data points or distributions. Misuse of the bar chart for displaying continuous data has been one of the most common display problems in some journals. When choosing between different types of graphs, it is important to consider study design, sample size, and data distribution to avoid representation problems [13]. In this sense, some works regarding the good use of data visualization have been made without becoming a statement [14–20]. Sports data tends to be hypervariate, temporal, relational, hierarchical, or a combination thereof, which leads to some fascinating visualization challenges [21,22]. However, recent excessive use and misinterpretation of radar plots in the field of sports analytics have led us to carry out this work to explain their pros and cons and to show some improved alternatives with real examples. This work could help to improve the education of statistics in the field of sport and above all can contribute to the critical reflection of the practice of coaching and its staff.

2. Use of graphical visualizations

Box plots, bar plots, and histograms are probably the most commonly used graphs at the descriptive level, whether dealing with discrete or continuous data.

In the medical field, various methods are used for visualizing complex data, such as Kaplan–Meier curves, forest plots, funnel plots, violin plots, waterfall plots, heat maps, and network analysis [23]. In the face of this great variety, sometimes, correct interpretation and communication of data via visualization is not given enough importance. We must bear in mind the difficulty that visualization presents when we are dealing with more dimensions, when there is heterogeneity, and when different formatting structures can be presented. This has led some scientific journals to adopt new visualization policies for correctly representing data [24]. Various initiatives in the field of biomedical science have been promoted to improve visualization practices by the collaboration between data scientists, computer scientists, science communicators, and graphic designers [25]. A statistical graph can offer an alternative compelling approach to statistical thinking that focuses on important concepts rather than procedural formulas. Nowadays, visualizing multidimensional/multivariate data is essential but can also be challenging [1]. Multivariate data visualization is a classic topic, for which many solutions have been proposed, each with its own strengths and weaknesses, for example, scatterplot matrices, radar charts, hyperboxes, timewheels, and parallel coordinates plot [26].

In the field of sport, as in others, one of the graphs that has been used recently is the radar plot [27–31].

2.1. The radar plot: What is it and what information does it provide?

The radar plot, also known as a spider plot or web plot, is a graph that consists of a series of equiangular axes, called radii, where each represents a variable. These radii are usually graduated from the smallest values in the center, to the largest values at the extreme radii. Radar plots are useful for displaying multivariate observations with arbitrary numbers of data, that is, they are graphs that help distinguish the pattern in multidimensional data. In this type of graph, a large number of variables can be represented, with different scales of measurement. These graphs are typically used for showing the degree of similarity between different groups, or displaying or comparing two or more items or groups on various characteristics [27].

2.2. Use and abuse of radar plots

Most people see radar plots as a useful tool. First, a key advantage of them is the square aspect ratio which aligns to target real estate availability in many cases. This is in contrast with a technique such as parallel coordinates. A second advantage advocated by some is that they can lead to particular polygonal patterns which become easily recognizable. However, Stephen Few, among other authors, have been outspoken critics of this tool [32–34].

Recently, in the field of sports analytics, the statistician and influential analyst Luke Boorn pointed out the abuse of the use of radar plot [35]. Many journalists, managers, scientists, and even data analysts in the field of sports, use this graph and the information that can be drawn from it, with little knowledge of the subject. Sometimes, its misuse is due to having a cursory knowledge of this graphical tool. Before using these graphs, there are some details that should not be overlooked in order to avoid misuse:

- The pattern, and not the plot area, should be the main focus. Filling in the background of the plot encourages the viewer to focus on the area when interpreting the graph and also to not be able to correctly see particular values or measures (Figure 1). A limitation of the plot of the radar plot is, also, that it suggests a connection between the axes when, often, they are independent.
-

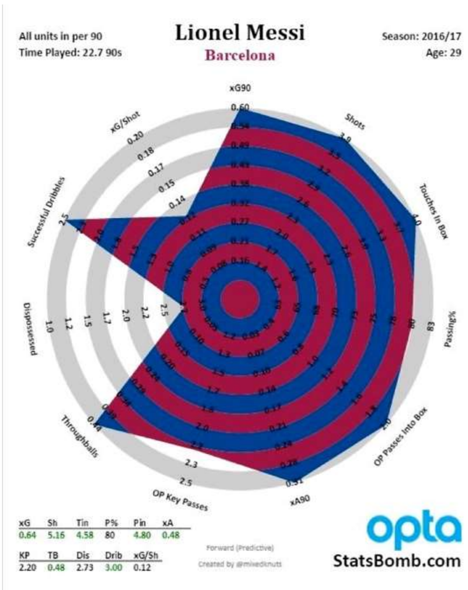

Figure 1: Example of a radar plot of the features of Lionel Messi from Opta and StatsBomb

- The scales of the variables should be taken into account when the minimum values are of interest, since the direction of the axis graduation must be reversed in these cases. An example of bad practice is shown in Figure 2. Unlike Figure 1, the decreasing scale is not taken into account and the scale order is not reversed for the “missed shots” variable, therefore, it can lead to possible misinterpretation. The values of the scales must also be taken into account. The area will depend on how we adjust the scale: if it ranges from 20 to 60 or from 0 to 100, the output will be different. This results in the graph being distorted.

Figure 2: Example of a radar plot omitting axis inversion. Taken from https://blog.fansbet.com/european-transfer-window-watch-january-23-2019/

- The measure that is being represented on the axis should be reported, whether it is value, value of a percentage or index, or measures of centralization (e.g., mean, median, mode). Frequently, the plots do not report what is being depicted, leading to misinterpretation (Figure 3).

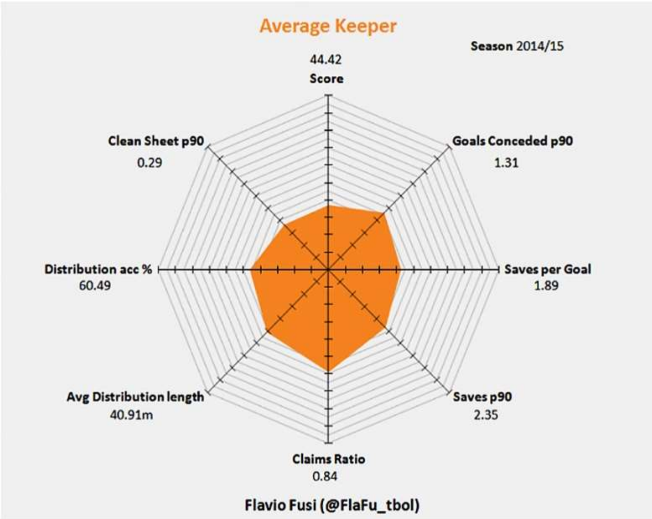

Figure 3: Example of a radar plot of the characteristics of a goalkeeper with variables that are not clearly defined. Excerpted from Flavio Fusi, “Introducing Goalkeeper radars and GK-score - Radar Love, flaws and conundrums” https://fmtrovaspazi.files.wordpress.com/2015/09/average-keeper3.png

- The possible spurious correlation that may exist between the different variables that are being represented should be taken into account before interpreting the plot [36]. This may be due to size (not the nature of the data), a common causal variable, or simply due to chance. In addition, when using large volumes of data, the use of techniques to reduce dimensions, such as principal component analysis or t-SNE, must be considered beforehand [37–40].

- We must keep in mind that our expertise in visual perception can be instrumental in being able to perceive differences in evaluated sets of variables over individual ones or similar patterns. This is apparent in simple graphs, such as boxplots, but is key with complex multivariate graphs such as radar plots. Our visual ability to distinguish a dataset on a radar plot is probably easier with similar and/or extreme data.

2.3. **Alternatives to radar plots to visualize data**

There are several effective alternatives to radar plots, depending on the information, the size of the data, or the risk of overlap that may arise.

Using alternative graphs to radar plots is important so that we can observe if the same patterns can be identified with the same data. An alternative graph is the parallel coordinates plot or spaghetti plot [27], where the different variables can be observed in parallel, allowing to better identify the patterns, since they are not analyzed areas. This graph is valuable when there is a small amount of data.

Another alternative would be the swarm plot, which represents a categorical point diagram without overlap. Box plots and violin plots [41] would also be good complements; besides showing the visual differences through the boxes or violins, they include more information, such as measures of centralization, position and dispersion, and they identify outliers. In the case of the violin plot, the distribution or density function of the data is also displayed. Outliers, atypical data, have values that clearly differentiate them from the others and should not be confused with anomalous data, which contain incorrect values. McBean and Rovers state an outlier may be ignored if a physical reason is available but, if not, then the value must be included [42]. It should be kept in mind that statistics make the facts easier to see but do not interpret them. An improved alternative that integrates most of the plots described above to the problem of overlap is described in two works applied to soccer, where cluster information is used to show the density distribution using new visualization techniques. These papers demonstrate that parallel coordinate plots can be a good alternative for displaying patterns for an in-depth analysis of clustered or ungrouped data [43,44].

3. New proposals: Violinboxplot and enhanced radar plot

Here, we present a proposal based on a new graph called violinboxplot and the improvement of the radar plot, in what we have called enhanced radar plot.

We will illustrate this proposal with the application of two real datasets: the first of which uses data from professional football competitions with European teams, and the second uses load and injury prevention data from professional basketball players.

All analyses were performed using version 3.6 of statistical software R (Vienna, Austria; http://www.rproject.org/) and with the package ggplot2 for graphical representations [45]. The R code used and the data from one of the examples is open access and can be found at https://github.com/marticasals. This work seeks to fulfill the transparency and reproducibility of scientific research.

3.1. Presentation of the data

The first example data set is soccer data from all the matches (n = 1,168) of the top 13 European clubs in the 2016/17 season: Premier League (Arsenal, Chelsea, Liverpool, Manchester United, Manchester City, Tottenham), La Liga (FCB, Real Madrid, Atlético Madrid), Ligue 1 (PSG), Bundesliga (Bayern, Borussia Dortmund), and Serie A (Juventus). It also includes the Champions League matches of the selected teams when they participated.

The variables used in this example are “competition” and “18-yard box shot” (i.e., the percentage of shots in the penalty area excluding those made within the well-directed goal area). The data used in this work was obtained from the publicly accessible website www.whoscored.com owned by Opta Sports.

The second data set is training load and injury data collected from the work of Caparrós et al. [46]. The study was conducted between October 2014 and April 2017 for three consecutive leagues, with a total of 2,613 observations and 246 matches from 33 different male professional basketball players. Tracking data were obtained from the STATS web site (http://stats.com) and injury information was obtained from the public repositories (www.rotoworld.com, www.cbs.com and www.basketball-reference.com), following the same procedure with performance data (http://stats.com and www.basketball-reference.com). The variables we have selected to exemplify in this work are: “injury (Yes/No)”, and “external load factors” (deceleration, distance, and minutes).

3.2. Violinboxplot: definition and its use with soccer data

One of the advantages of the boxplot is that, with a few summary statistics, we can define atypical data in a non-parametric context. Data farther from 1.5 interquartile amplitudes (IQR) above the third quartile or below the first quartile are considered atypical data. This calculation is made without taking into account the mean or the deviation values, statistics that are not robust in the presence of atypical data (Figure 4a). However, boxplots are not as easy to understand, interpret, or connect with other statistical representations of the same data.

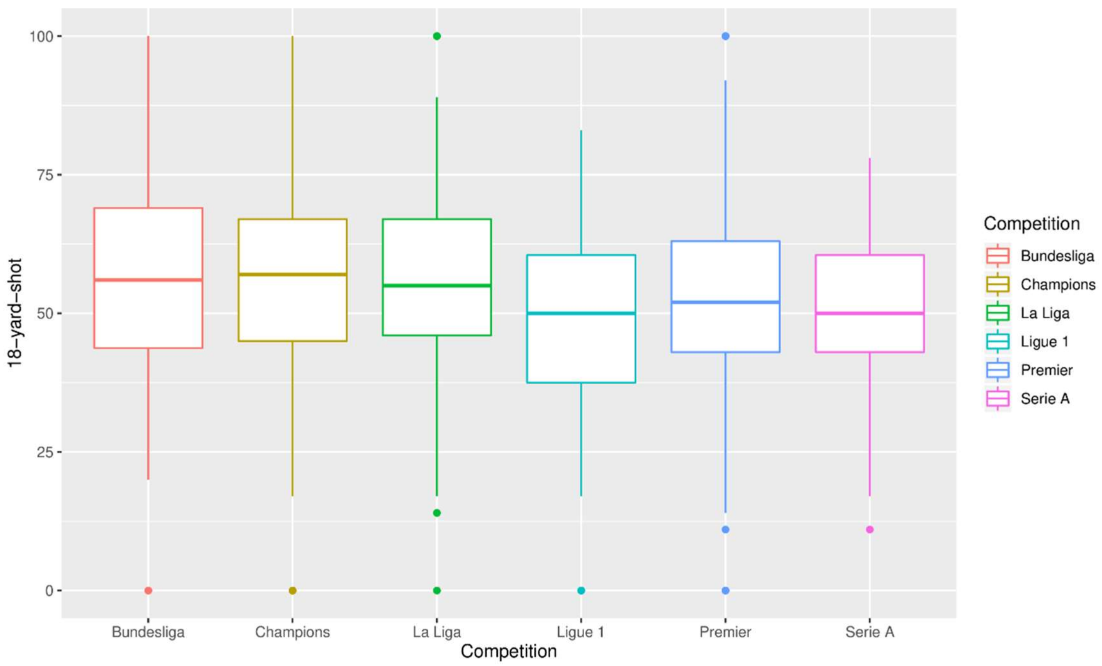

Figure 4a. Boxplots of the "18-yard shot" variable based on the competition, where the quartiles, maximum, minimum, and atypical data are highlighted; these data may be of great interest in sports analytics.

One drawback is that, between quartiles, there is no information on how the data is distributed [47], whether the distribution is unimodal or multimodal, which is why in some areas they are being replaced by violin plots.

Using kernel density graphs, such as violin plots, allows us to provide information about the distribution of the data, and not just the quartiles. However, they adapt to atypical data, since density takes into account all values, including atypical ones. The violin plot is smoothed towards the extreme, potentially towards atypical data (Figure 4b).

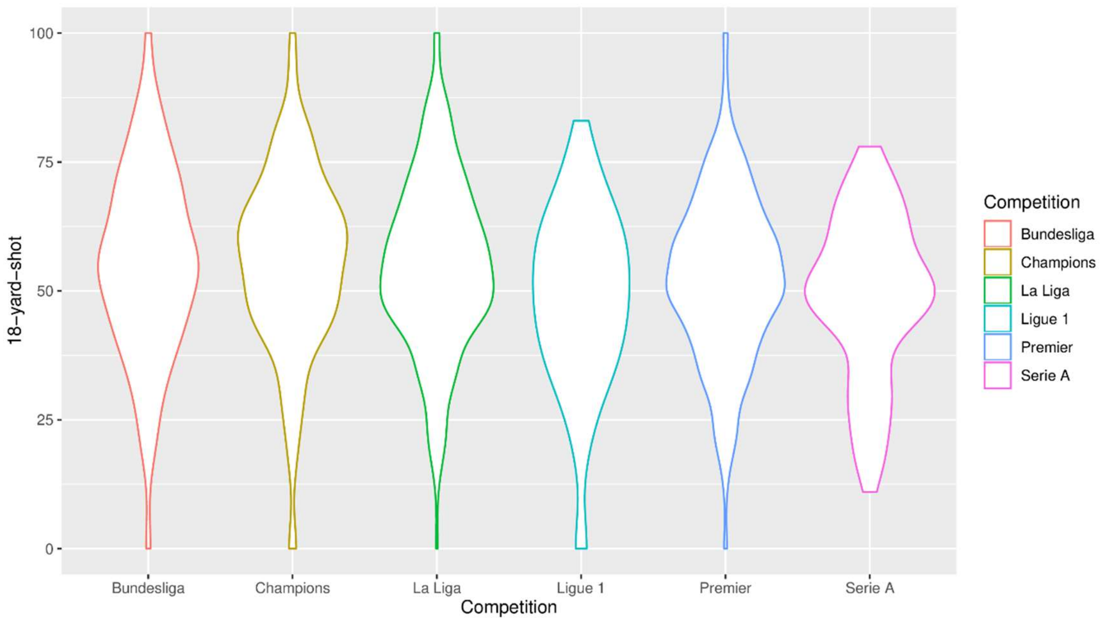

Figure 4b. Violin plots with data-driven distributions. Atypical data is integrated into the graph, representing the “18-yard shot” variable based on the competition.

We propose to use the violin plot only for non-atypical data and to mark atypical data independently, as is usually done in boxplots (Figure 5). This can help to better interpret the pattern of a graph, taking into account all the data and highlighting atypical values as well. In the case of sports, it can help to detect possible atypical data: “superstars”. We have called this new graph the violinboxplot where we show a violinplot for non-atypical data, in the boxplot sense, and after that we show the atypical data in the same axe.

This proposal for a new graphical representation goes beyond a simple superposition of two graphs preserving the concept of atypical data in the boxplot and the approximation of the kernel density of the violinplot by the non-atypical ones.

In Figure 5, in the violinboxplot corresponding to the “18-yard shot (%)” variable for all competitions, we can find several atypical data that would not be perceived in the usual violinplot (Figure 4b), (e.g: outliers in two competitions such as la Liga and the Premier (two of the most "powerful" competitions today)). They could go unnoticed in violinplots, while in the field of sports analytics they may be of great interest; and also we can compare the shape of the different distributions that is not possible in standard boxplots. For example, we noticed that it is just the Serie A that has a less wide distribution (less variability) and with a more pronounced mode, which makes us perceive that there is a more stable 18-yard-shot distribution in this competition. Figure 5, unlike the previous one (Figure 4b), shows how the distribution of the violins is not so long, since it has been smoothed taking into account the “non-atypical” range of the boxplots.

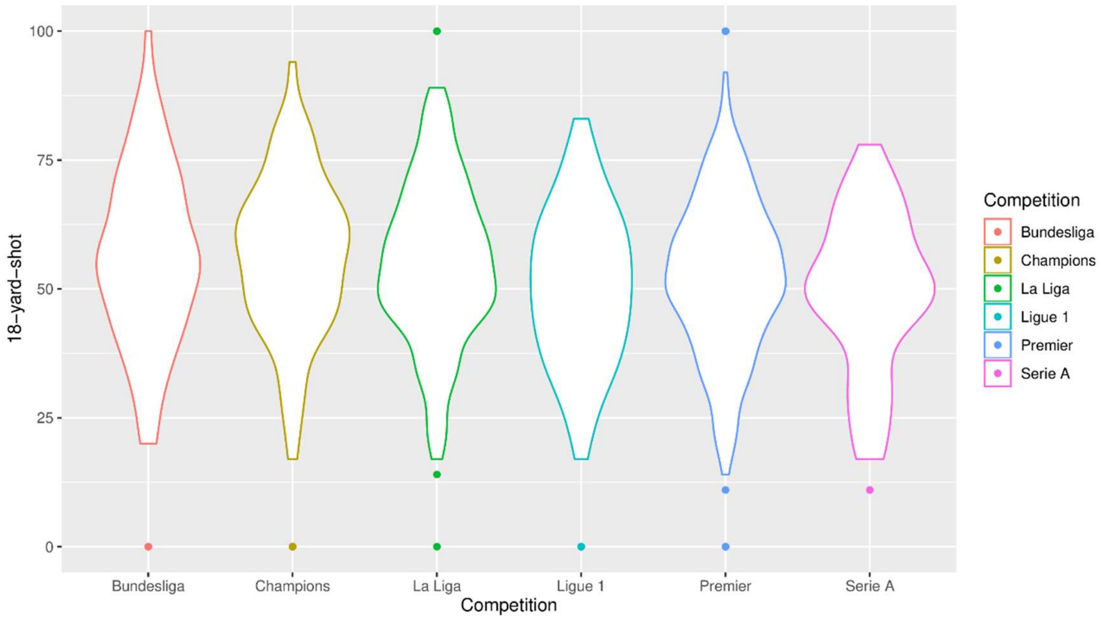

Figure 5. Violinboxplots: violinplots adapted to the presence of atypical data of the “18-yard shot (%)” variable based on the competition.

3.3. Enhanced radar plot: definition and its use with soccer data

Radar plots are used in sports analytics mainly to highlight athletes who have higher values of performance or to report on abilities, by means of certain indicators or physical parameters, which are clearly noticeable above the rest [31,48,49].

We aimed to create a new, enhanced radar plot, based on the violinboxplot, that would avoid the presence of areas that often facilitate graph misinterpretation. Additionally, we wanted to leverage the symmetry of violin plots to add comparisons between the two categories of a dichotomous variable.

If polar coordinates are used instead of parallel coordinates, we can adapt the graph built in the previous section (Figure 5) and obtain the enhanced radar plot, a radar plot based on violinboxplots (Figure 6). In Figure 6 we can interpret the same as in Figure 5, but with the difference of being able to see in the form of radar, which prevents the misuse of areas in radar plots.

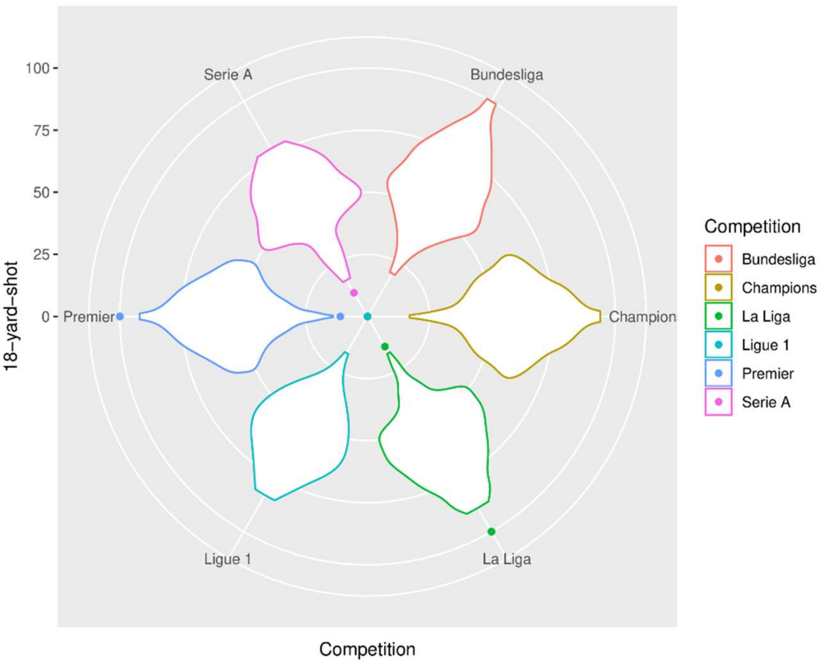

Figure 6. Enhanced radar plot for the “18-yard shot (%)” variable (%) according to the competition: violinboxplots emphasizing the presence of atypical data in polar coordinates.

3.4. Representation of supplementary elements in the enhanced radar plot with basketball data

In the enhanced radar plot, the coordinates of an athlete can be represented, and their position can be seen relative to others.

As Pascal Eduoard et al. pointed out we also need to improve figures in sports injury epidemiology [50]. Figure 7a shows an enhanced radar plot with the variables “acceleration”, “deceleration” and “minutes” from a professional basketball database. Minutes is the variable that refers to played time in minutes. We detect in our radar the minutes played by the players where there are outliers (those who play the most) and there is also a clear asymmetry and possible bimodal distribution (2 groups of players with more minutes played). Regarding external load variables, acceleration has a fairly wide distribution without outliers, while deceleration has a wider distribution with outliers (high values).

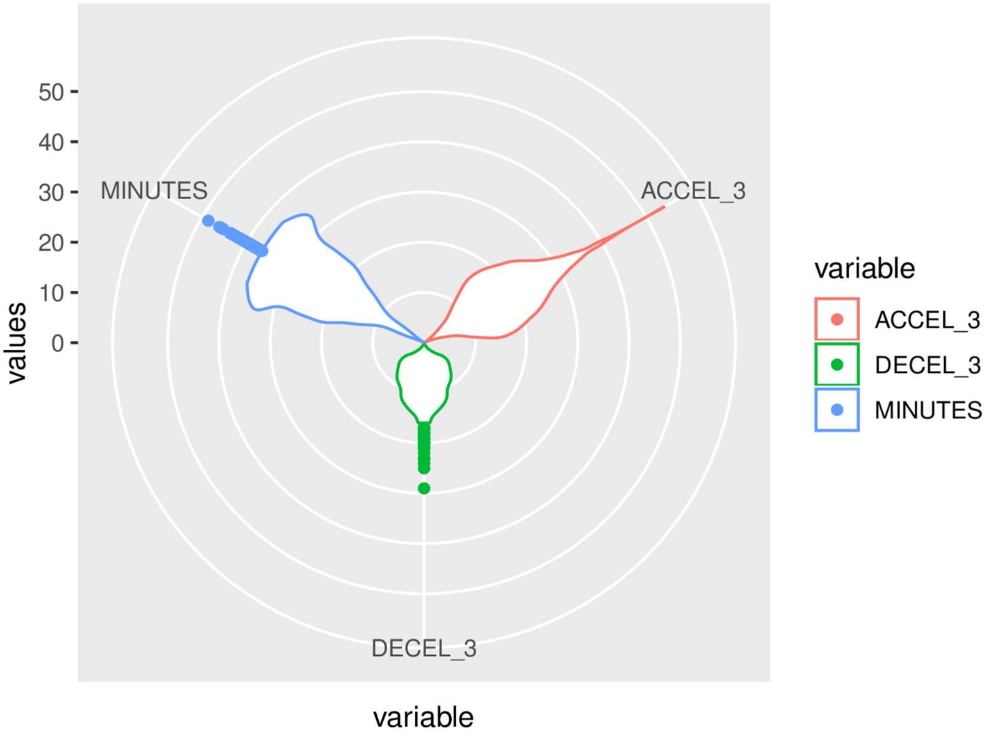

Figure7a. Representation of the enhanced radar plot of variables “acceleration”, “deceleration”, and “minutes”.

Using this, we can, for example, represent the information of these variables for a specific player (e.g., new signing). In this example, the new player has acceleration values of 40 m/s^2, deceleration of 20 m/s^2, and minutes equal to 15. The graph in Figure 7b shows two (first and third) of the three coordinates have values within the body of the violinboxplot, while the second coordinate is atypical. This type of visualization can help evaluating performance and scouting by highlighting the new player's pattern over the rest in the observed parameters. This possible player that we incorporate in the previous plot, would be a prototype of player that has a few minutes (if we see the dot in the distribution of this variable), with high values of acceleration and also deceleration (specifically would correspond to an outlier).

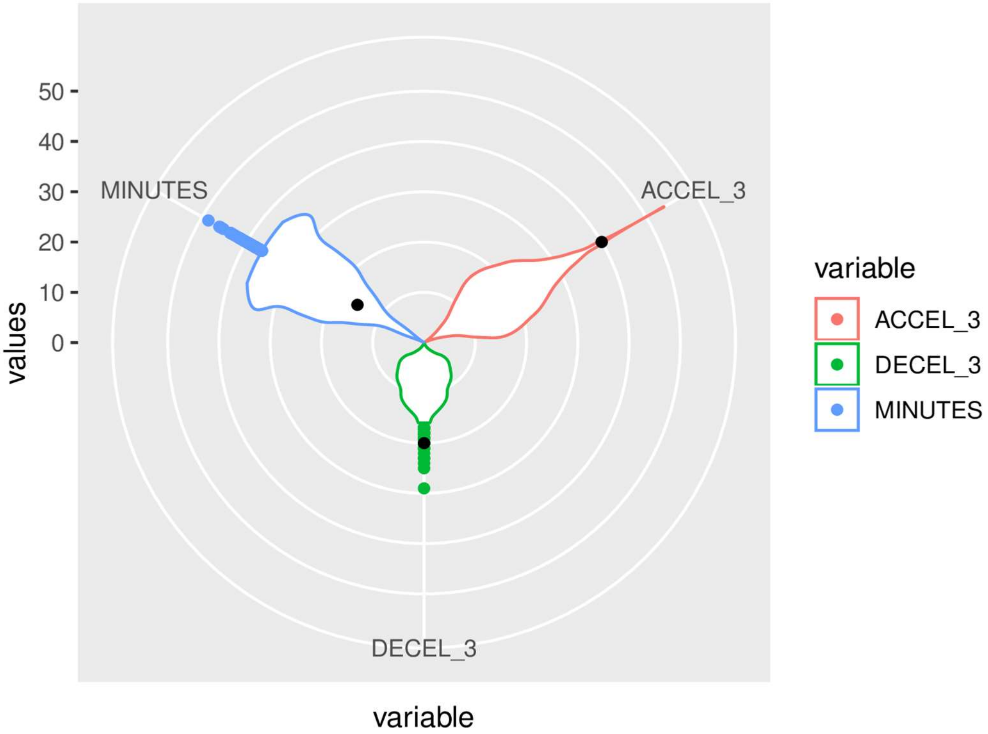

Figure 7b. Incorporation of new data (marked in black) in the enhanced radar plot to graphically evaluate these values and compare them with the distribution of the rest.

3.5. Representation of complementary dichotomous variables in the enhanced radar plot

Violinplots typically represent two symmetrical density functions on each axis, although this double symmetry has no significant contribution. If we represent only one part of the symmetric density function, we, therefore, show only one half of the violinplot. This would allow us to represent the two opposite half violins of the binary variable, resulting in a half-half violin plot or split violin plot that allows us to compare the two distributions. From this, we could obtain an enhanced radar plot and, thus, compare distributions of the categories of a binary variable for each of our variables with the corresponding half-half violinboxplots. We need to highlight that the split violin plots are only useful for 2 distributions. For more than 2, we need to overlap plots and perhaps we should use other tools.

In Figure 8, we have used the variable “injury (yes/no)” in conjunction with the 3 variables used in Figure 7a and we can see their distribution differences and similarities according to injury.

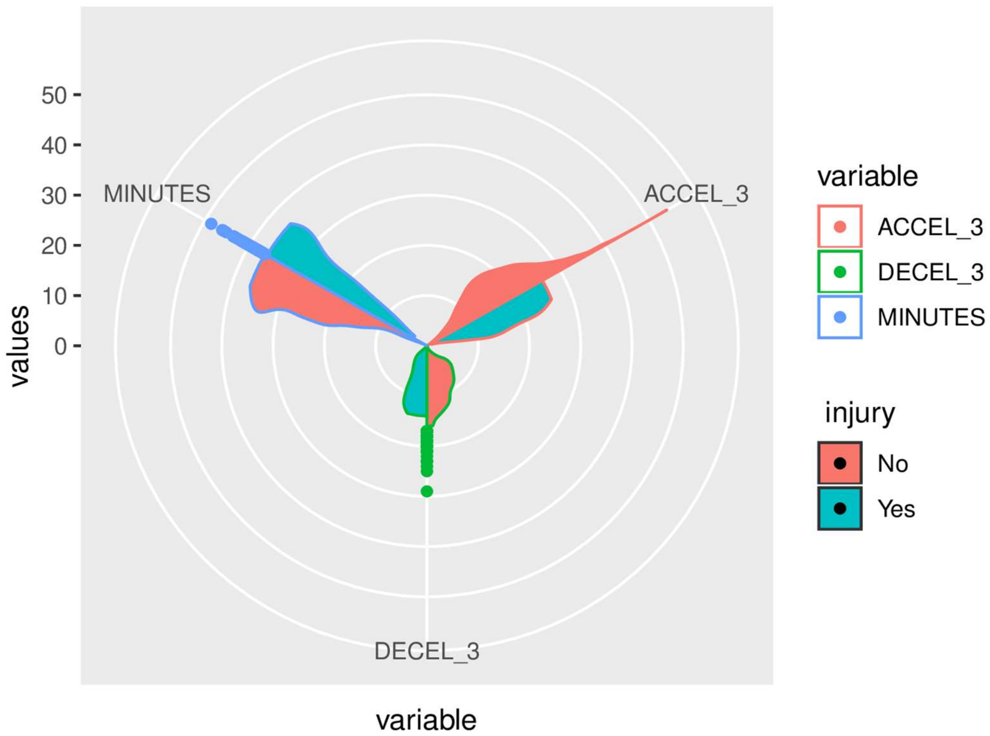

Figure 8. Use of a supplementary binary variable, such as injury, to compare the two categories tin the enhanced radar plot.

4. Conclusions

Scouts use data to find talented players, trainers use it to improve player performance, and analysts use to find even the slightest edge of a team over its rivals. With the increasing interest in sports visualization from both sports practitioners and visualization researchers, there is a need to promote transparency by improving the quality of the figures. Just like language, graphics communicate information to other people. It is crucial to keep it in mind (the audience and their knowledge) when the graphic is being constructed [51]. Academic journals and funding agencies are making lasting improvements to data visualization in the scientific literature and the sport industry.

The abuse of some graphs, such as the radar plot, and their frequent misinterpretation in the world of sports highlights the need to create more informative visualizations. Researchers and practitioners (especially scouts) can use more accurate alternatives than radar plots. In this work, we present the violinboxplot and enhanced radar plot as interesting alternatives. These are especially ideal with moderate or large volumes of data, showing possible patterns and uncertainty of the data thanks to the integration of the violin plots -representing the distribution of the data- and the integration of the box plots, representing the summaries and statistics. In addition, when representing boxplots within enhanced ones, outliers are represented (unlike in violin plots). The graphical representations were performed through the ggplot2 package that is generally perceived to be more clear in the multivariate relationship than using the base R[52].

Scientific and educational articles like this one that provide open data and the code used allow the readers to reproduce the analyses and determine if using different analytical techniques would have different results. Transparent reporting and open data are becoming increasingly important as scientists, funding agencies, and journals implement new strategies to improve scientific rigor and reproducibility.

It is our hope that these examples not only inspire the sports community, but also statistics educators and students motivated by sports examples to get closer to the statistical literacy.

In summary, we propose violinboxplots and enhanced radar plot to represent the data pattern, taking into account the statistical summaries as well as their distribution, without neglecting the interesting atypical data in the form of radar plots. These plots do not represent areas, which could lead to confusion. This work aims to be educational, from the perspective of statistical thinking, and also applied, so that the field of sports analytics (including decision makers, coaches, practitioners, and researchers) take into account these graphs when they want to observe or describe different patterns. In addition, these charts can not only be used in the field of Sports Analytics but can also be useful in other fields or areas.

Acknowledgments

We would like to thank Alfredo Ruiz Brichs, director of Data & Business Intelligence at Vinces and Partner Advisor on Sports Analytics, for transferring the football data set used in this paper. Moreover, the authors would like to thank Caparrós et al for the permission to use the basketball data set.

MC acknowledge the financial support from Ministerio de Ciencia e Innovación y Ministerio de Universidades (Ref: PID2019-104830RB-I00 / AEI DOI: 10.13039/501100011033).

References

1. Nolan D, Perrett J. Teaching and Learning Data Visualization: Ideas and Assignments. *Am Stat* 2016; 70: 260–269.

2. Rossman AJ, Horton NJ. Interview With Nicholas J. Horton. *J Stat Educ* 2018; 26: 234–245.

3. Hudiburgh LM, Garbinsky D. Data Visualization: Bringing Data to Life in an Introductory Statistics Course. *J Stat Educ* 2020; 28: 262–279.

4. Wood BL, Mocko M, Everson M, et al. Updated Guidelines, Updated Curriculum: The GAISE College Report and Introductory Statistics for the Modern Student. *CHANCE* 2018; 31: 53–59.

5. Alamar B. *Sports analytics: A guide for coaches, managers, and other decision makers*. 2013.

6. Swartz TB. Where Should I Publish My Sports Paper? *Am Stat* 2018; 1–6.

7. Spiegelhalter D. *The art of statistics: learning from data*. 2019.

8. Jayal A, McRobert A, Oatley G, et al. *Sports Analytics: Analysis, Visualisation and Decision Making in Sports Performance*. London: Routledge. Epub ahead of print 2018. DOI: 10.4324/9781315222783.

9. Yau N. *Visualize this: the FlowingData guide to design, visualization, and statistics*. 2011. Epub ahead of print 2011. DOI: 10.1002/9781118722213.

10. Cairo A. *The truthful art: data, charts, and maps for communication*. 2016.

11. Makin TR, Orban de Xivry J-J. Ten common statistical mistakes to watch out for when writing or reviewing a manuscript. *Elife*; 8. Epub ahead of print 9 October 2019. DOI: 10.7554/eLife.48175.

12. Groves T. Enhancing the quality and transparency of health research, https://www.bmj.com/content/337/bmj.a718.short (2008, accessed 19 March 2020).

13. Weissgerber TL, Winham SJ, Heinzen EP, et al. Reveal, Don't Conceal. *Circulation* 2019; 140: 1506–1518.

14. Brodlie K, Allendes Osorio R, Lopes A. A Review of Uncertainty in Data Visualization. In: *Expanding the Frontiers of Visual Analytics and Visualization*. Springer London, 2012, pp. 81–109.

15. Diong J, Butler AA, Gandevia SC, et al. Poor statistical reporting, inadequate data presentation and spin persist despite editorial advice. *PLoS One* 2018; 13: e0202121.

16. Hertel J. A Picture Tells 1000 Words (but Most Results Graphs Do Not): 21 Alternatives to Simple Bar and Line Graphs. *Clinics in Sports Medicine* 2018; 37: 441–462.

17. Weissgerber TL, Savic M, Winham SJ, et al. Data visualization, bar naked: A free tool for creating interactive graphics. *ASBMB*. Epub ahead of print 2017. DOI: 10.1074/jbc.RA117.000147.

18. Weissgerber TL, Garovic VD, Winham SJ, et al. Transparent reporting for reproducible science. *J Neurosci Res* 2016; 94: 859–864.

19. Weissgerber TL, Garovic VD, Savic M, et al. From Static to Interactive: Transforming

Data Visualization to Improve Transparency. *journals.plos.org*; 14. Epub ahead of print 22 June 2016. DOI: 10.1371/journal.pbio.1002484.

20. Weissgerber TL, Garovic VD, Savic M, et al. Correction: From Static to Interactive: Transforming Data Visualization to Improve Transparency. *PLOS Biol* 2016; 14: e1002545.

21. Basole RC, Saupe D. Sports Data Visualization [Guest editors' introduction]. *IEEE Comput Graph Appl* 2016; 36: 24–26.

22. Du M, Yuan X. A survey of competitive sports data visualization and visual analysis. *J Vis* 2021; 24: 47–67.

23. Chia PL, Gedye C, Boutros PC, et al. Current and Evolving Methods to Visualize Biological Data in Cancer Research. *J Natl Cancer Inst* 2016; 108: djw031.

24. Vandemeulebroecke M, Baillie M, Carr D, et al. How can we make better graphs? An initiative to increase the graphical expertise and productivity of quantitative scientists. *Pharm Stat* 2019; 18: 106–114.

25. O'Donoghue SI, Baldi BF, Clark SJ, et al. Visualization of Biomedical Data. *Annu Rev Biomed Data Sci* 2018; 1: 275–304.

26. Claessen JHT, van Wijk JJ. Flexible Linked Axes for Multivariate Data Visualization. *IEEE Trans Vis Comput Graph* 2011; 17: 2310–2316.

27. Saary MJ. Radar plots: a useful way for presenting multivariate health care data. *J Clin Epidemiol* 2008; 61: 311–317.

28. Pardo S, Dunne N, Simmons DA. Using Radar Plots to Demonstrate the Accuracy and Precision of 6 Blood Glucose Monitoring Systems. *J Diabetes Sci Technol* 2017; 11: 966–969.

29. Brenner DM, Dove LS, Andrae DA, et al. Radar plots: A novel modality for displaying disparate data on the efficacy of eluxadoline for the treatment of irritable bowel syndrome with diarrhea. *Neurogastroenterol Motil*; 30. Epub ahead of print 1 August 2018. DOI: 10.1111/nmo.13331.

30. Artur Serrano J, Awad H, Broekx R. Nature-inspired radar charts as an innovative Big Data analysis tool. In: *Applications of Big Data Analytics: Trends, Issues, and Challenges*. Springer International Publishing, 2018, pp. 177–183.

31. Losada AG, Theron R, Benito A. BKViz: A Basketball Visual Analysis Tool. *IEEE Comput Graph Appl* 2016; 36: 58–68.

32. Few S. Keep radar graphs below the radar-far below. *Information Management* 2005; 15: 48.

33. Odds G. A Critique of Radar Charts. *Scottlogic*, https://blog.scottlogic.com/2011/09/23/a-critique-of-radar-charts.html (2011, accessed 24 August 2020).

34. Alberto Cairo. Radar graphs: Avoid them (99.9% of the time). *thefunctionalart*, http://www.thefunctionalart.com/2012/11/radar-graphs-avoid-them-999-of-time.html (2012).

35. Boorn L. A reminder, blatantly plagiarized from @stat_sam, of why radar plots are misleading. Eye focuses on area, not length., https://twitter.com/LukeBornn/status/864856335191388162 (2017).

36. Tobías A, Casals M, Peña J, et al. FIFA World Cup and climate change: correlation is not causation Copa del Mundo FIFA y cambio climático: correlación no implica causalidad. *cafyd.com*. Epub ahead of print 2019. DOI: 10.5232/ricyde.

37. Moura FA, Martins LEB, Cunha SA. Analysis of football game-related statistics using

multivariate techniques. *J Sports Sci* 2014; 32: 1881–1887.

38. Van Der Maaten L, Hinton G. *Visualizing Data using t-SNE*, http://www.jmlr.org/papers/v9/vandermaaten08a.html (2008, accessed 19 March 2020).

39. Ho EMCH. You Down with t-SNE?, https://www.americansocceranalysis.com/home/2018/11/26/tsne (2018).

40. Amabili L. *Visualizing Algorithms of Nonlinear Dimensionality Reduction Techniques*. 2017.

41. Hintze JL, Nelson RD. Violin Plots: A Box Plot-Density Trace Synergism. *Am Stat* 1998; 52: 181–184.

42. McBean EA, Rovers FA. Estimation of the Probability of Exceedance of Contaminant Concentrations. *Groundw Monit Remediat* 1992; 12: 115–119.

43. Stein M, Janetzko H, Seebacher D, et al. How to make sense of team sport data: From acquisition to data modeling and research aspects. *Data* 2017; 2: 2.

44. Janetzko H, Stein M, Sacha D, et al. Enhancing Parallel Coordinates: Statistical Visualizations for Analyzing Soccer Data. *Electron Imaging* 2016; 2016: 1–8.

45. Wickham H. *ggplot2*. Cham: Springer International Publishing. 2016. DOI: 10.1007/978-3-319-24277-4.

46. Caparrós T, Casals M, Solana Á, et al. Low external workloads are related to higher injury risk in professional male basketball games. *J Sports Sci Med* 2018; 17: 289.

47. Matejka J, Fitzmaurice G. Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. *dl.acm.org* 2017; 2017-May: 1290–1294.

48. Fister I, Fister I, Fister D. Visualization of Sports Activities Created by Wearable Mobile Devices. In: *Adaptation, Learning, and Optimization*. Springer Verlag, 2019, pp. 223–246.

49. Whitehead T. Nylon Calculus: In defense of radar charts. *Nylon Calc*, https://fansided.com/2017/05/19/nylon-calculus-defense-radar-charts/ (2017).

50. Pascal Edouard LN. Keep it simple: how to best present the extent of a sports injury problem. *Blog BJSM*, https://blogs.bmj.com/bjsm/2019/12/19/keep-it-simple-how-to-best-present-the-extent-of-a-sports-injury-problem/ (2019).

51. Rolandi M, Cheng K, Pérez-Kriz S. A Brief Guide to Designing Effective Figures for the Scientific Paper. *Adv Mater* 2011; 23: 4343–4346.

52. Myint L, Hadavand A, Jager L, et al. Comparison of Beginning R Students' Perceptions of Peer-Made Plots Created in Two Plotting Systems: A Randomized Experiment. *J Stat Educ* 2020; 28: 98–108.

Figure captions

Figure 1: Example of a radar plot of the features of Lionel Messi from Opta and StatsBomb

Figure 2: Example of a radar plot omitting axis inversion. Taken from https://blog.fansbet.com/european-transfer-window-watch-january-23-2019/

Figure 3: Example of a radar plot of the characteristics of a goalkeeper with variables that are not clearly defined. Excerpted from Flavio Fusi, "Introducing Goalkeeper radars and GK-score - Radar Love, flaws and conundrums" https://fmtrovaspazi.files.wordpress.com/2015/09/average-keeper3.png

Figure 4a. Boxplots of the “18-yard shot” variable based on the competition, where the quartiles, maximum, minimum, and atypical data are highlighted; these data may be of great interest in sports analytics.

Figure 4b. Violin plots with data-driven distributions. Atypical data is integrated into the graph, representing the “18-yard shot” variable based on the competition.

Figure 5. Violinboxplots: violinplots adapted to the presence of atypical data of the “18-yard shot (%)” variable based on the competition.

Figure 6. Enhanced radar plot for the “18-yard shot (%)” variable (%) according to the competition: violinboxplots emphasizing the presence of atypical data in polar coordinates.

Figure7a. Representation of the enhanced radar plot of variables “acceleration”, “deceleration”, and “minutes”.

Figure 7b. Incorporation of new data (marked in black) in the enhanced radar plot to graphically evaluate these values and compare them with the distribution of the rest.

Figure 8. Use of a supplementary binary variable, such as injury, to compare the two categories tin the enhanced radar plot.