# Module 4. Introduction to visualization

BARÇA INNOVATION HUB
Universitas

- Visualization as a key step in the data analysis process.
- References

# Visualization as a key step in the data analysis process.

Data visualization is one of the main ways to interpret information due to its ease of understanding, flexibility, and usefulness in simplifying the results we want to communicate (Midway, 2020).

As Sport Scientists and following he data analysis scheme, this last part of the process is essential to put our analyses to real use, that is, the decision-making process. As has been emphasized during the course, data analysis provides context to our environment and visualization will allow us to be the link between our work and the different departments that require the information.

**AS BUCHHEIT (2017) POINTS OUT, THERE ARE THREE KEY ELEMENTS IN THE COMMUNICATION OF DATA BY SPORT SCIENTISTS:**

**TO THESE THREE PRINCIPLES A FUNDAMENTAL POINT MUST BE ADDED:**

- Correctly understanding the data, variables and analysis we carry out: This concept is dealt with in previous and complementary courses.
- Effective communication: This concept highlights the importance of offering correct visualizations, which we will discuss during the course.

- Appropriate oral communication skills: This will not be covered in the course, and as the author points out, it is a highly individual point.

**AS BUCHHEIT (2017) POINTS OUT, THERE ARE THREE KEY ELEMENTS IN THE COMMUNICATION OF DATA BY SPORT SCIENTISTS:**

**TO THESE THREE PRINCIPLES A FUNDAMENTAL POINT MUST BE ADDED:**

- Knowing who the information is aimed at - we cannot make the same visualizations for players, coaches, managers or staff of our department. The time that each of these people can spend processing the information varies, as does the level of experience and detailed knowledge of the information we provide. It will be the Sport Scientist who decides the type of information that is shared and the reason behind that decision (their objective).

In addition to its fundamental role in communicating results, as we saw in previous courses, visualization is essential in the initial phases of the process. Using simple visualizations will allow us to explore the data more clearly and decide what the action plan will be for the next steps.

The use of the RStuido software and mainly its ggplot2 graphics package will allow us great flexibility in producing visualizations to achieve our objective within the framework of data analysis.

Enter a question title here...

- ○ Knowing who the information we want to communicate is aimed at means we can not make the same visualizations for players, coaches, managers or staff of our department
- ○ Knowing who the information we want to communicate is aimed at is not a key factor in the communication process, since we can use the same visualizations for players, coaches, managers or staff of our department

SUBMIT

**Visualization in the context of the Sport Scientist.**

We must take into account the peculiarities of the field of sports performance. This is often a very complex context and we find ourselves with very diverse structures and data sources, so it is crucial to have a clear outline of the information we want to communicate.

Questions we can ask ourselves to communicate the results:

- Do we want to make a comparison between players?
- Do we want to know the evolution of a player throughout the season?
- Do we want to assess this evolution, but at two different times of the year?
- Do we want to compare how values from different data sources are related (GPS vs jump test)?
- Do we want to describe our team compared to the rest of the teams in the same competition?

Based on these questions, and taking into account the analysis content of the previous course and the considerations of each type of visualization that we will see below, we can select which type of graph is suitable to respond to them.

**Principles and foundations.**

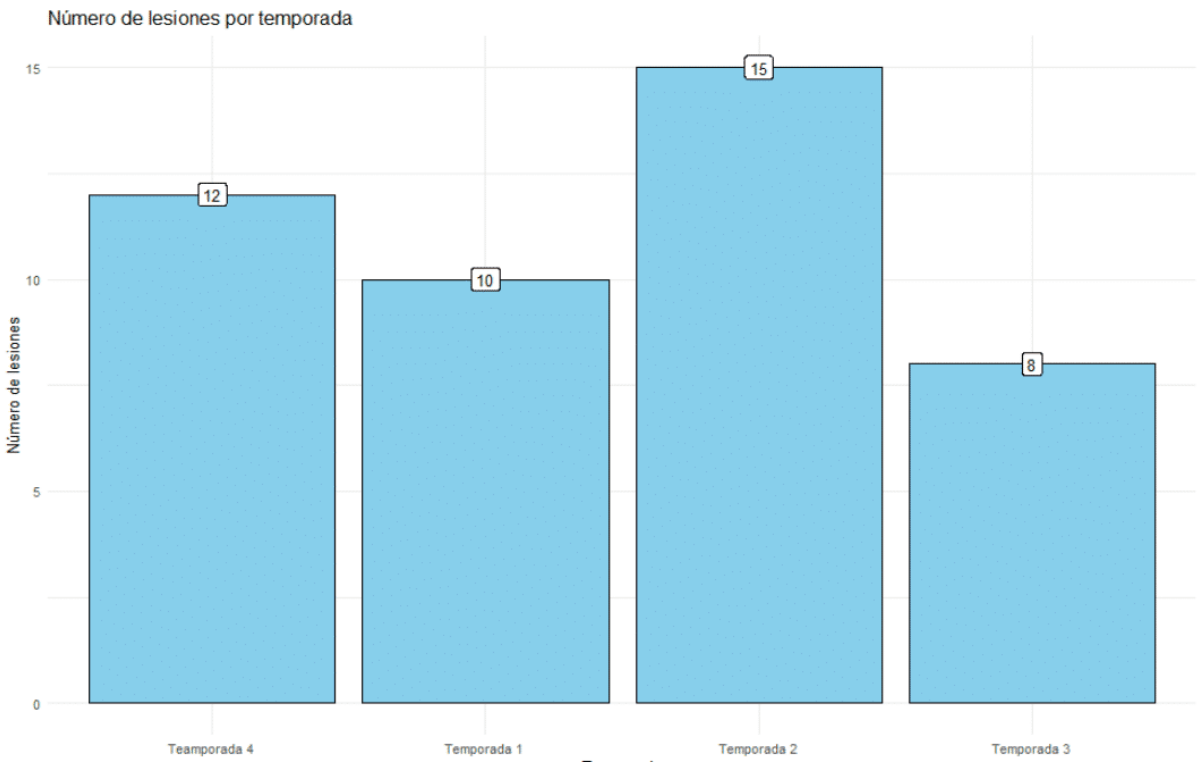
As we have highlighted above, the first step before visualizing the results is to choose what data we want to communicate, and from this decision follows the first principle:

Choose the correct visualization for the type of data we want to represent (Wilke, 2019):

- Visualizing quantities.
    - Bar Chart: This is the most common type of visualization for this case, although dots can also be used. It is useful when you want to compare different subjects. The limitations of this type of graphics

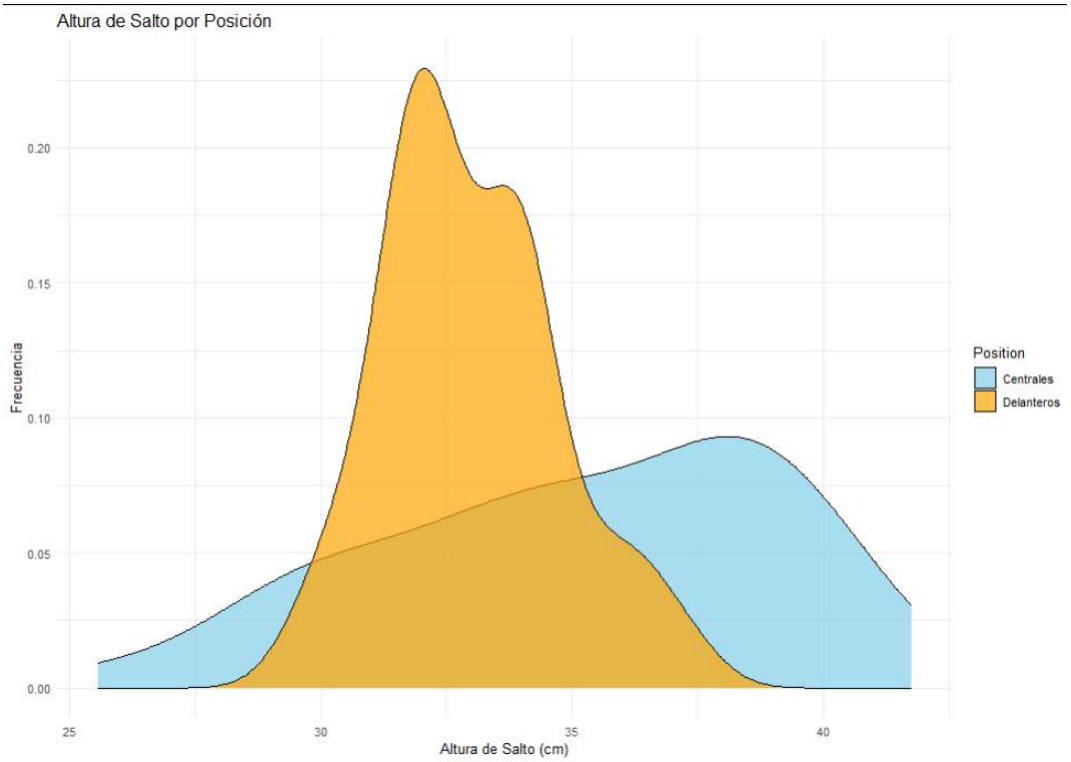
should be highlighted; If we want to represent distributions or if the quantities we want to show include more than one subject, it will be more appropriate to choose different types of graphs.

- An example applied to the context of the Sports Scientist could be the number of injuries each month within our team. In this case we are not comparing subjects but months, but the purpose is the same. Quantities do not encompass multiple subjects, so it is appropriate to represent the data in this way.



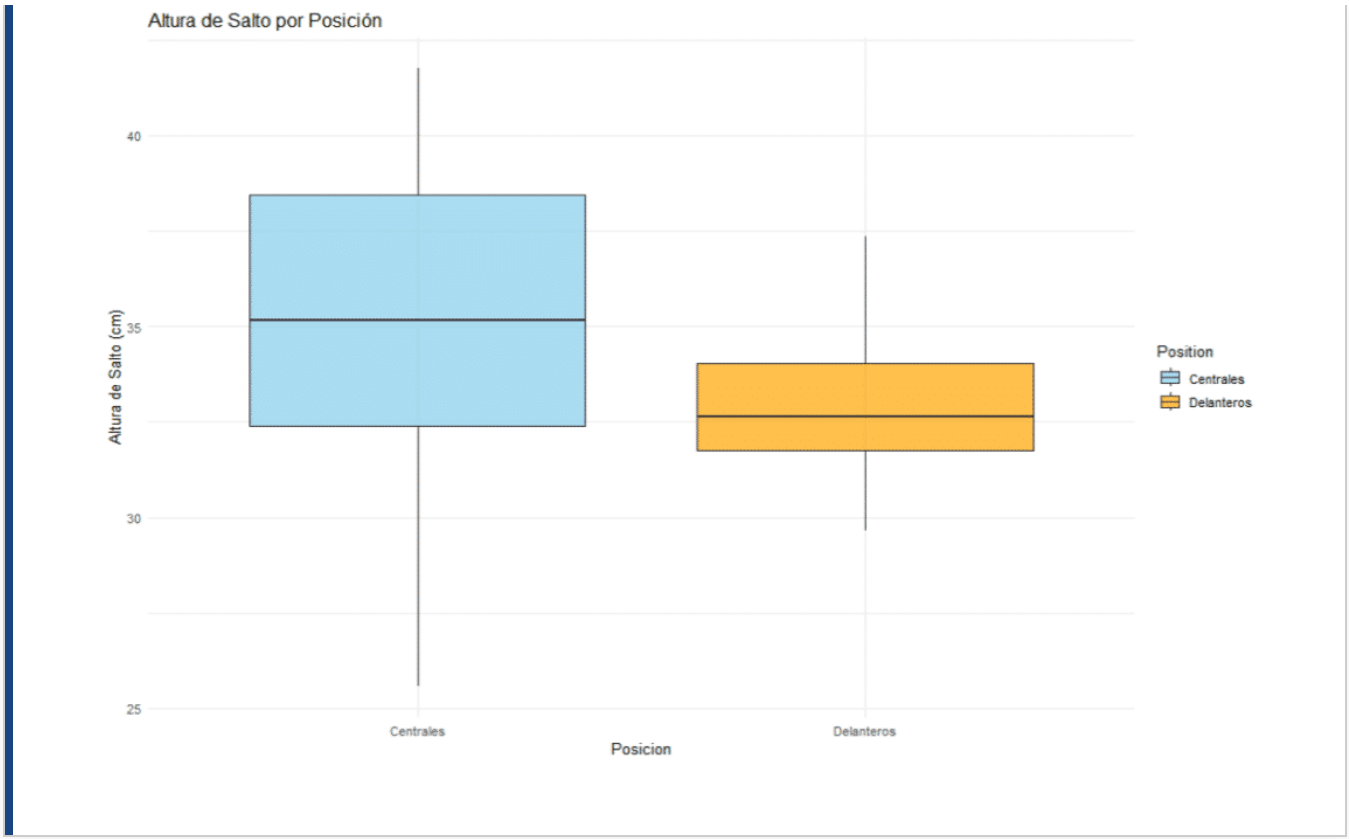

**Visualizing distributions:**

Density/Histogram Graph: This type of visualization seeks to represent the way the data is organized along a series of values. It is very useful to see which values are most common within our population as well as to see at a glance the outliers or extreme values.

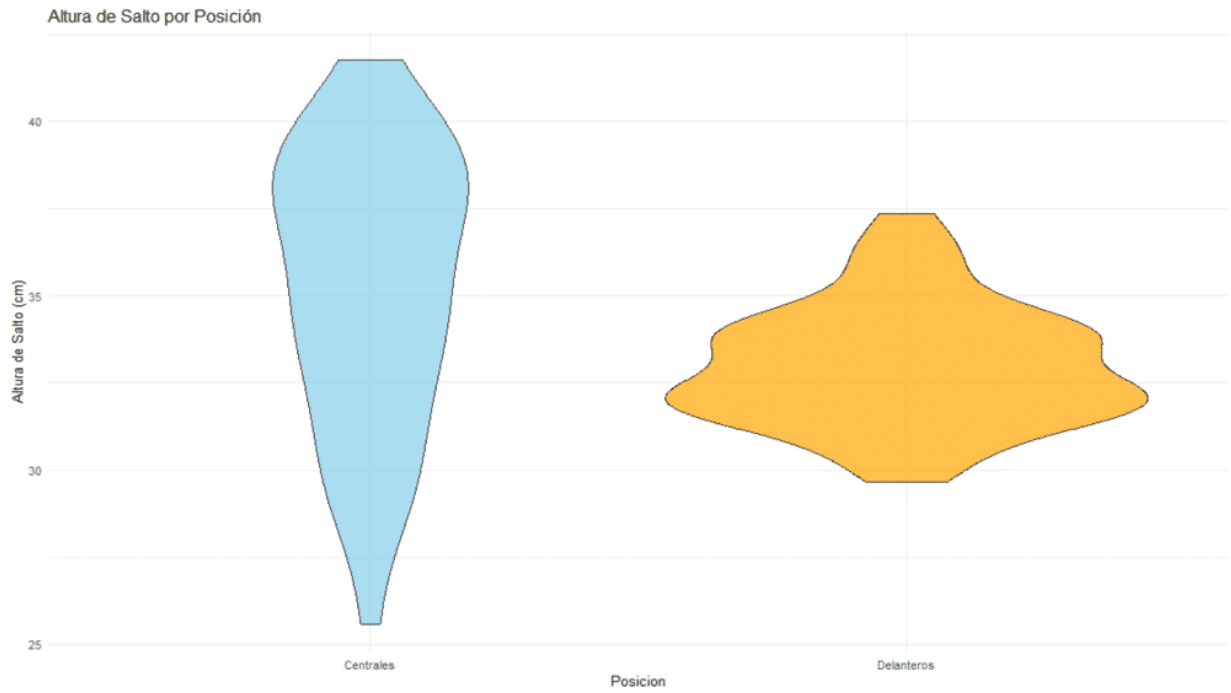



**Visualizing distributions:**

- Dot graph/plot: This can be an alternative to the bar graph in case we want to make a more accurate representation of how the results are distributed regarding all the individuals of which we want to highlight the total accumulated value (Torres-Ronda and Curtis, 2024).
- Box plot: This is a simplification of a density graph, but it is very useful for comparisons as it shows the central point of the distribution and the usual desired range, that is, between which values the most results are accumulated.




- Violin graph: It is a version between the box plot and a density graph since it provides information on the shape of the distribution but also allows us to see descriptive values of the variable we want to analyse.




Source: Author's own production

- The last three graphs present the same data, but each of them provides different information due to the particularities of each type of visualization.

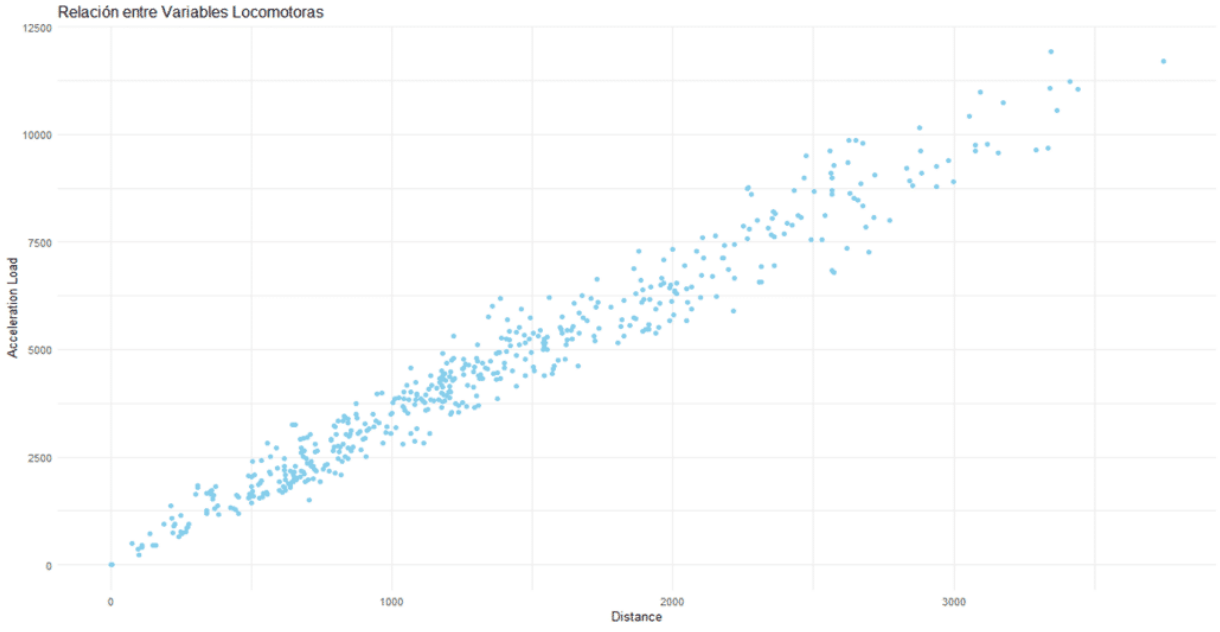
- Visualizing proportions:

    - Pie chart: It is commonly used to represent percentages of a total. A great disadvantage is their difficult reading on many occasions when there are a large number of variables and the proportions are small, in which case we should opt for another type of visualization.

- Stacked columns: The same concept applies to this type of graph. The visualization must be clear to meet our objective.

- Visualizing associations:

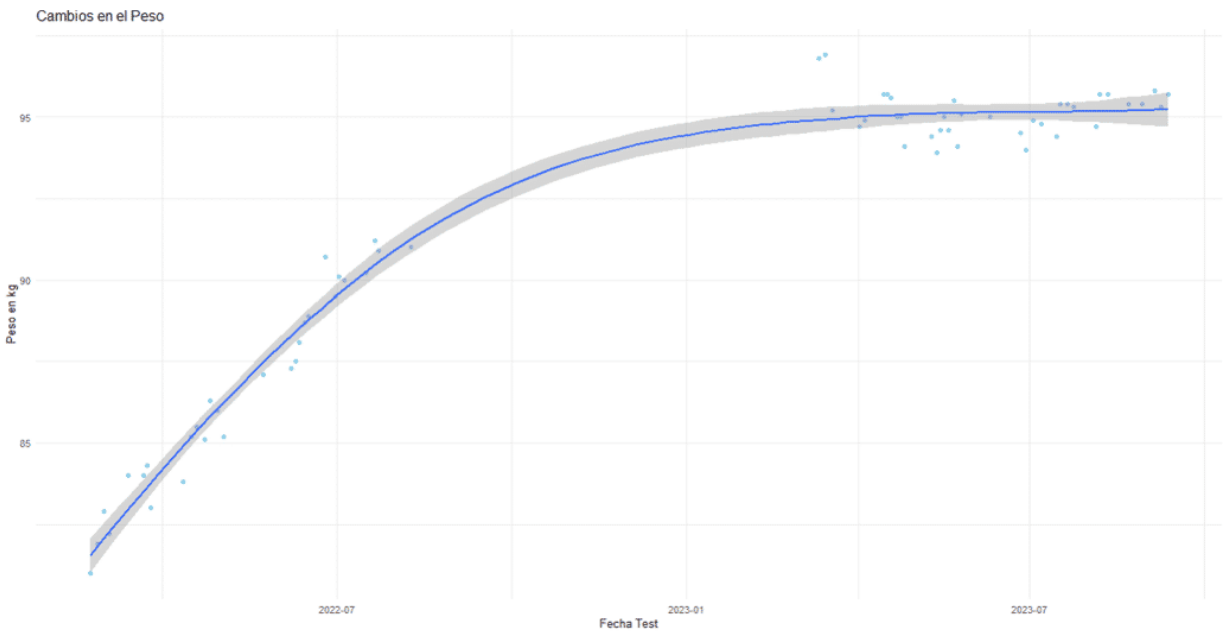
  - Scatterplot: This type of graph is used when we want to represent the relationship between two quantitative




Source: Author's own production

- Visualizing time series: We consider time series when we want to represent the change of a variable over certain time period.

- Visualizing Trends: It can be a complementary visualization to the previous one or we could use it without a time variable. When we deal with a lot of data, and its observations can be very close together, we may want to see which way the data is moving over time. Trends allow us to simplify observations.




Source: Author's own production

The first step before visualising the results is to choose what data we want to communicate, and from this decision follows the first principle. What is that principle?

Type your answer here

SUBMIT

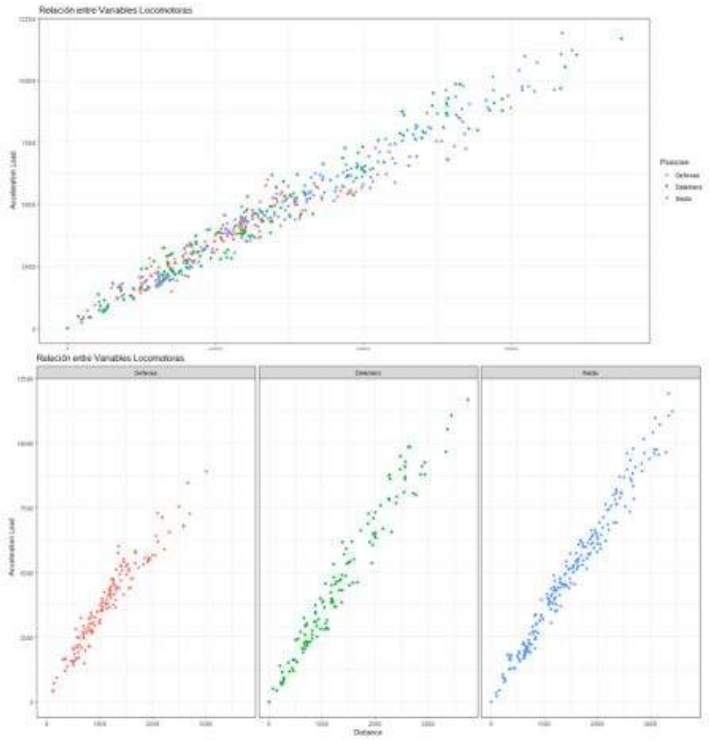
**Visualization format considerations.**

Although the fundamental point is the type of visualization to choose for each type of data and the information we want to communicate, it must be taken into consideration that the way in which it reaches the final recipient has to be appealing, error-free and easily assimilated. To achieve this, it is necessary to take into account a series of principles (Irizarry, 2019):

- Avoid Distractions within the graph: Keep the information simple so that the message reaches your audience. We can take a bar graph as an example but in 3D format. This type of chart presents the same information as a bar chart like the one we have seen above, but adding 3D can cause confusion by adding perspectives and volumes.
- Choice of colours: If we want to represent several groups within our visualization, the number of those groups should be as small as possible. The more different colours, the more difficult it is to visualize the details of each of them. For example, when describing a football team. Choose more generalized positional groups (Forward is a better option than Winger or Midfielder).
- Overlapping observations: This is very common in dot plots like the one we have seen above. In that particular case, the objective was to represent an association between two variables and there was no distinction between several groups or colours, so those observations that are "hidden" under others would not need to be differentiated. In the case of the distribution

graph, if we did not chose to blur the colours, we would lose information on how each group is distributed. This would be one of the options; another option may be to scatter the dots. RStudio and its visualization packages allow you to use these options.

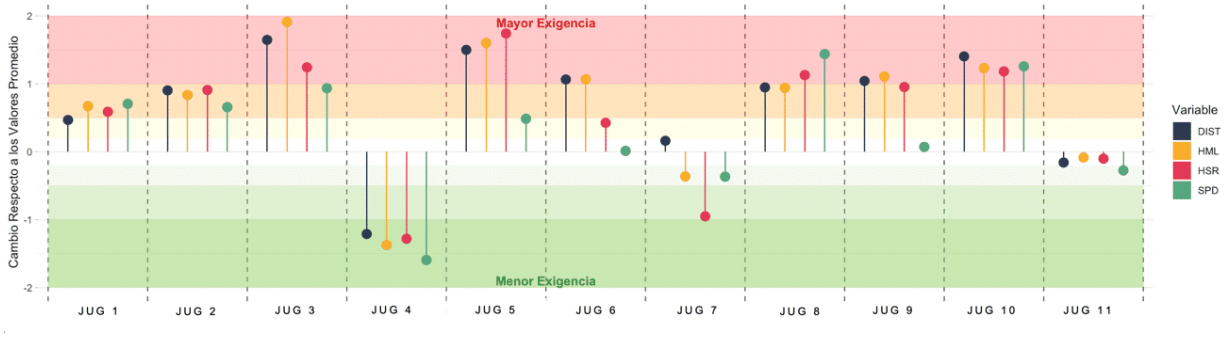
- Consider the ranges of the axes: This section is essential in the case of bar graphs. It is important that the y-axis always starts at 0, so we can have a real idea of what the reflected quantities and their magnitude are. In the case of other types of graphs such as scatterplots, the axes will be scaled to the range of the observations.
- Avoid redundancies in the colours or information within the graph: In the case of the box plot in the previous example, it has been chosen to colour each of the boxes to maintain the format of the rest of the graphs, but we see how there is already some information on the type of population on the x-axis, so it could be omitted. If we were to add a legend describing the colours, it would be a very redundant graph.
    - The same goes for shapes, if in a scatterplot we assign a colour to each of the groups, it would not be necessary to also give a different shape (triangle, square and rhombus).
- Text: Keep the text of the axes clear and visible, as well as that of the titles.
    - We can add text to the chart to make it easier to understand and track the visualization.
- Divide a visualization into multiples: If our visualization has many groups, we can divide it into multiples, to highlight the details of each of them.




**Examples of data communication in sports performance.**

To capture some of the extra considerations in the field of visualization we will share real examples, but in this case with more complex visualizations than the previous ones that aim to respond to the needs of a team or sports organization.

The following visualization allows us to capture the importance of simplifying data for communication to different audiences.




Source: Author's own production

This graph shows the conditional demands, measured by GPS devices, of a football training session.

Let us look at the particularities of this visualization:

- We represent 4 variables at the same time (Distance, High Metabolic Load, Sprint Distance and Sprinting Distance).
    - Instead of using 4 different graphs and different units (meters, W/kg), it was decided to use a Z-score, i.e. a standardization of the data. In this way we can compare variables and players simultaneously.
- Colours and text that help in tracking information.

- The colours and text clearly indicate that if the dots and lines are in the red zone, it has been a session of greater demand for the player; if on the contrary, they are below, the values of that session are lower than the player's averages. At a glance, any member of the staff can see the type of session it was.

- Levels of detail. Depending on the level of experience of the professional who is looking at the graph, they will be able to get more or less details out of the results and, for example, see which metrics have been most stimulated in the session and therefore if the conditional objective of the session has been met.

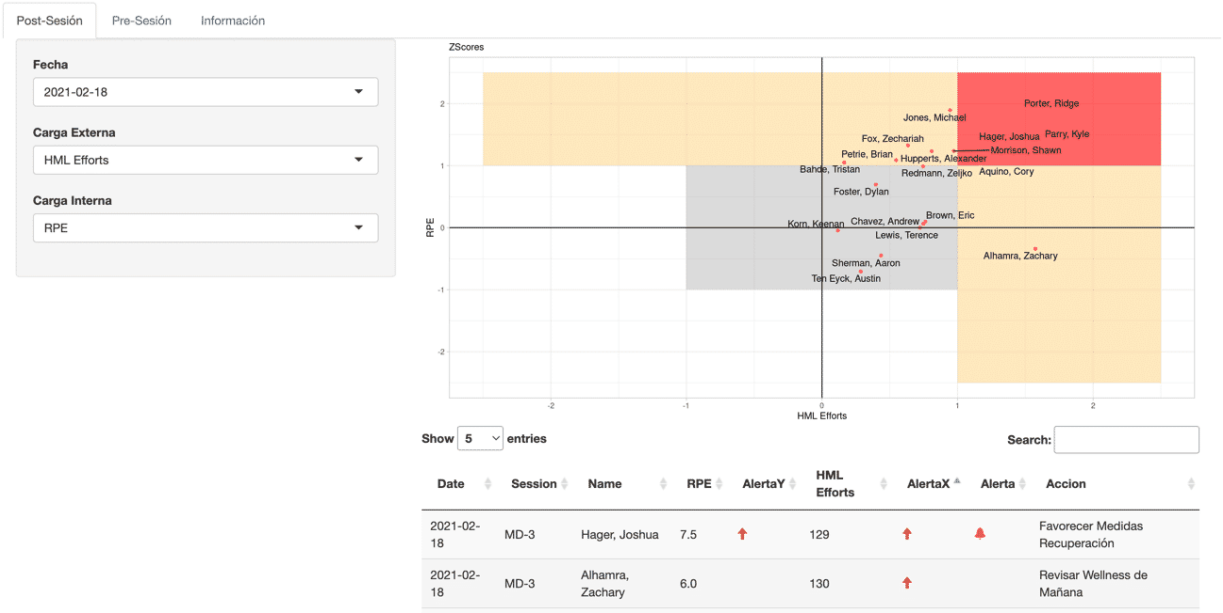
The following graphs are intended to show another fundamental principle in our context. The interaction or exploration of visualization. This allows us to give great value to data but we also run the risk of losing emphasis on the important points we want to communicate.

These graphics belong to a Shiny app, it is an R package that allows you to develop interactive websites to show visualizations and analysis to your users.

This app is based on the work of Miguel Ángel Campos (2021)

Link to the Shiny app:

https://davidpm.shinyapps.io/MonitoringCycle/

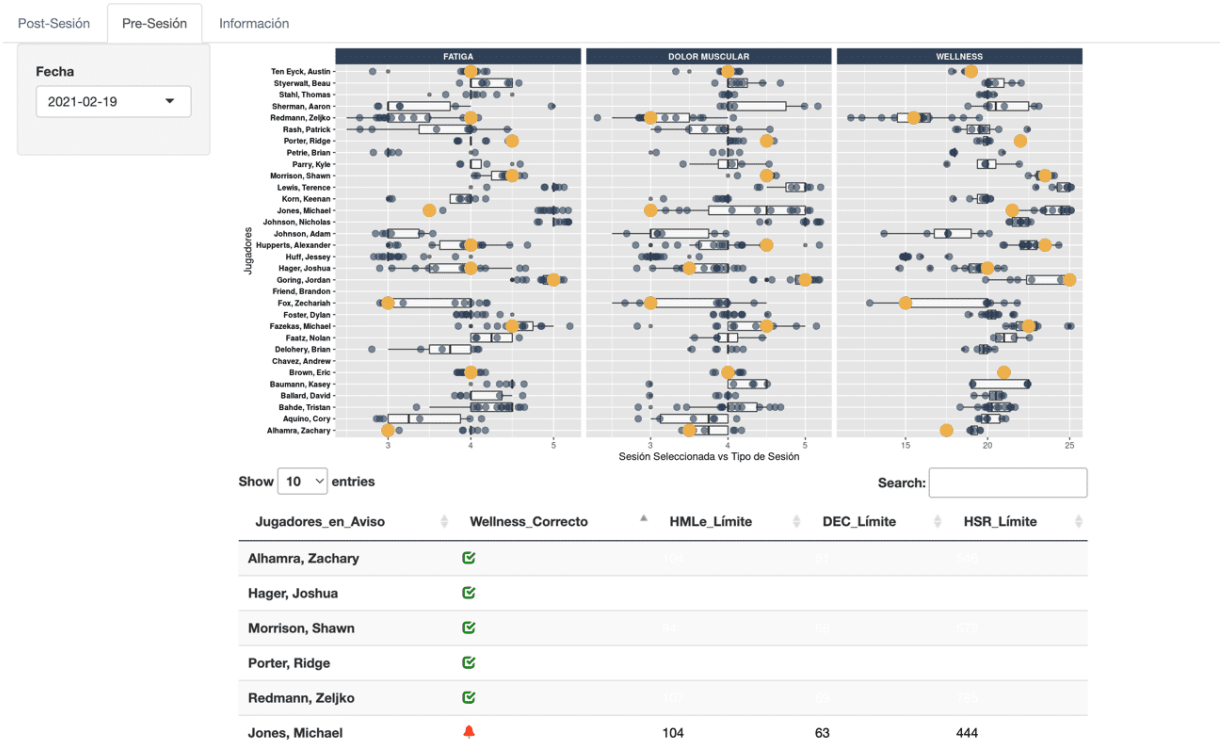



Source: Author's own production

In the first tab of the app it is indicated that it is the post-session analysis. The user must choose the date they want to analyse and the external and internal load variables.

The selection allows you to see how the session has been internally and externally for each player, and you can see "alerts" if these values have been above the player's usual ones. Each of the players

will also be provided with a recommendation at the end of the session.



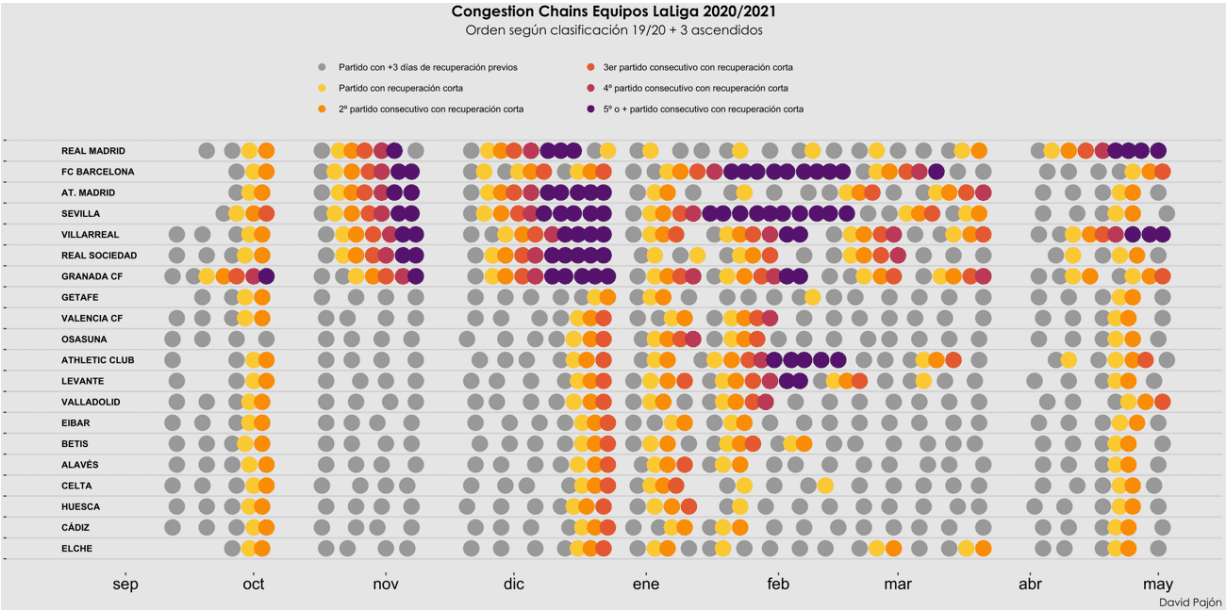

Source: Author's own production.

On the following day, we have to log into the app to see the answers to the well-being questionnaire, that is, how they have arrived to the session.

There was a list of players who were on "alert" after the end of the previous session, and we can see if the results have been within their average values or not. This will determine our training instructions in the following session.

As we can see, this is an app  with interactions which aims to aid decision-making and to provide the user with a day-to-day track.

Finally, there is an example outside the classic visualizations associated with the Sport Scientist, but which has great informative value on certain contexts. This is an analysis of the density of matches for LaLiga teams in the 20/21 season.




Source: Author's own production

High-density chains occur when there are consecutive matches with less than 3 days of recovery in between them. Through this analysis we can assess which teams have a tighter schedule, less rest and therefore more demand. These types of visualizations help different departments adjust recovery, training, and travel plans.

The choice of colours in the above and the below graphs helps to highlight what we want to communicate, which is one of the fundamental principles . In the first case, we can clearly distinguish which teams and which dates have been on a more demanding schedule. In the second case, the distances a team has covered are well represented.


Source: Author's own production

High-density chains occur when there are consecutive games with:

- More than 3 days of recovery in between them
- Less than 3 days of recovery in between them
- Less than 5 days of recovery in between them
- More than 5 days of recovery in between them

SUBMIT

CONTINUAR

# References

Buchheit, M. (2017)."Want to see my report, coach?" . *Aspetar Sports Medicine Journal, 6*. Retrieved from https://journal.aspetar.com/en/journals/volume-6-targeted-topic-straight-science/want-to-see-my-report-coach

Irizarry, R.A. (2019). Introduction to Data Science: Data Analysis and Prediction Algorithms with R (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429341830

Midway, S. R. (2020). Principles of Effective Data Visualization. Perspective, 1(9), 100141. https://doi.org/10.1016/j.patter.2020.100141

Torres-Ronda, L., & Curtis, R. M. (2024). Building the Foundations for Information Communication in Sports Science and Its Use in Decision Making. Strength and Conditioning Journal, 46(1), 74-81. https://doi.org/10.1519/SSC.0000000000000811

Wilke, C. (2019). Fundamentals of Data Visualization: A Primer on

Making Informative and Compelling Figures (1st ed.). O'Reilly Media, Inc.

CONTINUE