

# Módulo 4. Introducción a la visualización



☰ Unidad 1 La visualización como pieza clave en el proceso de análisis de datos.

☰ Referencias

# Unidad 1 La visualización como pieza clave en el proceso de análisis de datos.

---

La visualización de datos se trata una de las principales formas para interpretar la información, por su facilidad de comprensión, flexibilidad y utilidad para simplificar los resultados que queremos comunicar (Midway, 2020).

Como Sport Scientists y siguiendo con el esquema de análisis de datos, esta última parte del proceso es fundamental para poder llevar nuestros análisis a la aplicación real, es decir al proceso de toma de decisiones. Como se ha hecho énfasis durante el curso, el análisis de datos provee contexto al entorno en el que nos encontramos y la visualización permitirá ser ese vínculo entre nuestro trabajo y los distintos departamentos que requieren la información.

Como destaca Buchheit (2017) hay tres claves en la comunicación de los datos por parte de los Sport Scientist:



Entender correctamente los datos, variables y análisis que realizamos:  
Este concepto corresponde a cursos anteriores y complementarios.

- Comunicar de manera efectiva: Reitera la importancia de ofrecer correctas visualizaciones, sobre las cuales hablaremos durante el curso.
- Tener las habilidades apropiadas para la comunicación oral: Este punto no se tratará en el curso, y tal como destaca el autor, es un punto muy individual.

A estos tres principios cabe añadir un punto fundamental:

- Conocer a quién va dirigida la información que queremos compartir y comunicar, no podremos realizar las mismas visualizaciones para los jugadores, entrenadores, directivos o staff de nuestro mismo departamento. El tiempo que puede destinar cada una de estas personas a procesar la información es muy variable, así como el nivel de experiencia y conocimiento en detalle de la información que aportamos. Será el Sport Scientist el que deba decidir el tipo de información que se comparte y el porqué de su objetivo.

Además de su papel fundamental en la comunicación de resultados, como vimos en cursos anteriores, la visualización es básica en las

fases iniciales del proceso, utilizar visualizaciones sencillas nos permitirá explorar los datos de manera más clara y decidir cuál es el plan de acción para los siguientes pasos.

La utilización del software RStudio y principalmente su paquete de gráficos ggplot2 nos permitirá gran flexibilidad en la elaboración de visualizaciones para conseguir nuestro objetivo dentro del marco del análisis de datos.

## Autevaluación

- 
- 1. Conocer a quién va dirigida la información que queremos compartir y comunicar, no podremos realizar las mismas visualizaciones para los jugadores, entrenadores, directivos o staff de nuestro mismo departamento
  
  - 2. Conocer a quién va dirigida la información que queremos compartir y comunicar no es un factor clave en el proceso de comunicación, ya que podremos realizar las mismas visualizaciones para los jugadores, entrenadores, directivos o staff de nuestro mismo departamento.

SUBMIT

## **La visualización en el contexto del Sport Scientist.**

Debemos tener en cuenta las particularidades en el campo del rendimiento deportivo. Se trata de un contexto a menudo muy complejo y nos encontramos con estructuras y fuentes de datos muy diversas, por lo que será importante tener un esquema claro de lo que queremos comunicar.

### **PREGUNTAS QUE PODEMOS HACERNOS PARA COMUNICAR LOS RESULTADOS:**

- ¿Queremos realizar una comparativa entre jugadores?
- ¿Queremos conocer la evolución de un jugador a lo largo de la temporada?
- ¿Queremos valorar esa evolución, pero en dos momentos distintos del año?
- ¿Queremos comparar como se relacionan valores de distintas fuentes de datos (GPS vs test de salto)?
- ¿Queremos describir la situación de nuestro equipo comparado con el resto de los equipos de nuestra competición?

A partir de estas preguntas, y teniendo en cuenta el contenido de análisis del curso anterior y las consideraciones de cada tipo de visualización que veremos

a continuación, podemos seleccionar qué tipo de gráfico conviene para dar respuesta a ellas.

### **Fundamentos y principios.**

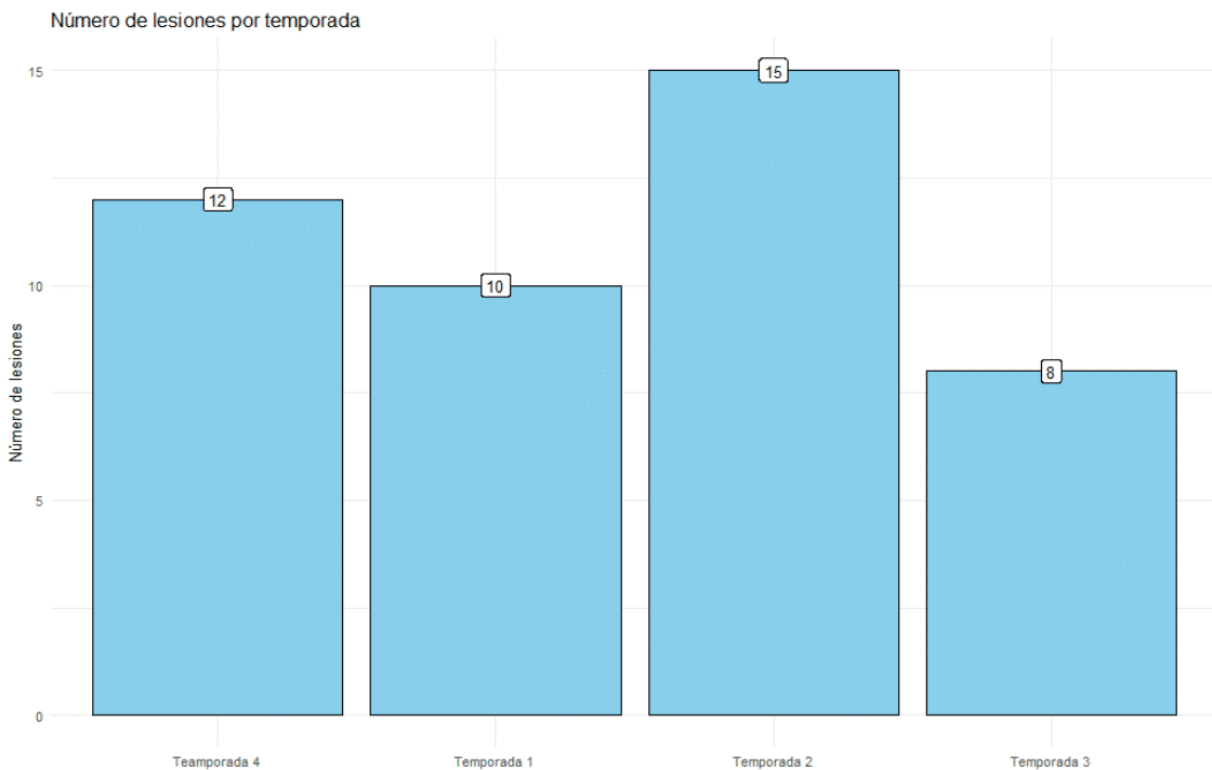
Como hemos destacado anteriormente, el primer paso que debemos realizar antes de visualizar los resultados es escoger qué datos queremos comunicar, y a partir de esta decisión seguir el primer principio:

Escoger la visualización correcta para el tipo de dato que queremos representar (Wilke, 2019):

- Visualizar cantidades.
  - Gráfico de Barras: Es el tipo de visualización más común para este caso, aunque también se puedan utilizar puntos. Es útil cuando se quieren realizar comparaciones entre distintos sujetos. Hay que destacar las limitaciones de este tipo de gráficos; si queremos representar distribuciones o si esas cantidades que queremos mostrar engloban a más de un sujeto, será mas apropiado escoger gráficos que veremos más adelante.

- Un ejemplo aplicado al contexto del Scientist podría ser el número de lesiones cada mes dentro de nuestro equipo. En este caso no estamos comparando sujetos sino meses, pero el propósito es el mismo. Las cantidades no engloban a múltiples sujetos por lo que es adecuado representar los datos de esta manera.

**Figura 1: Gráfico de barras**

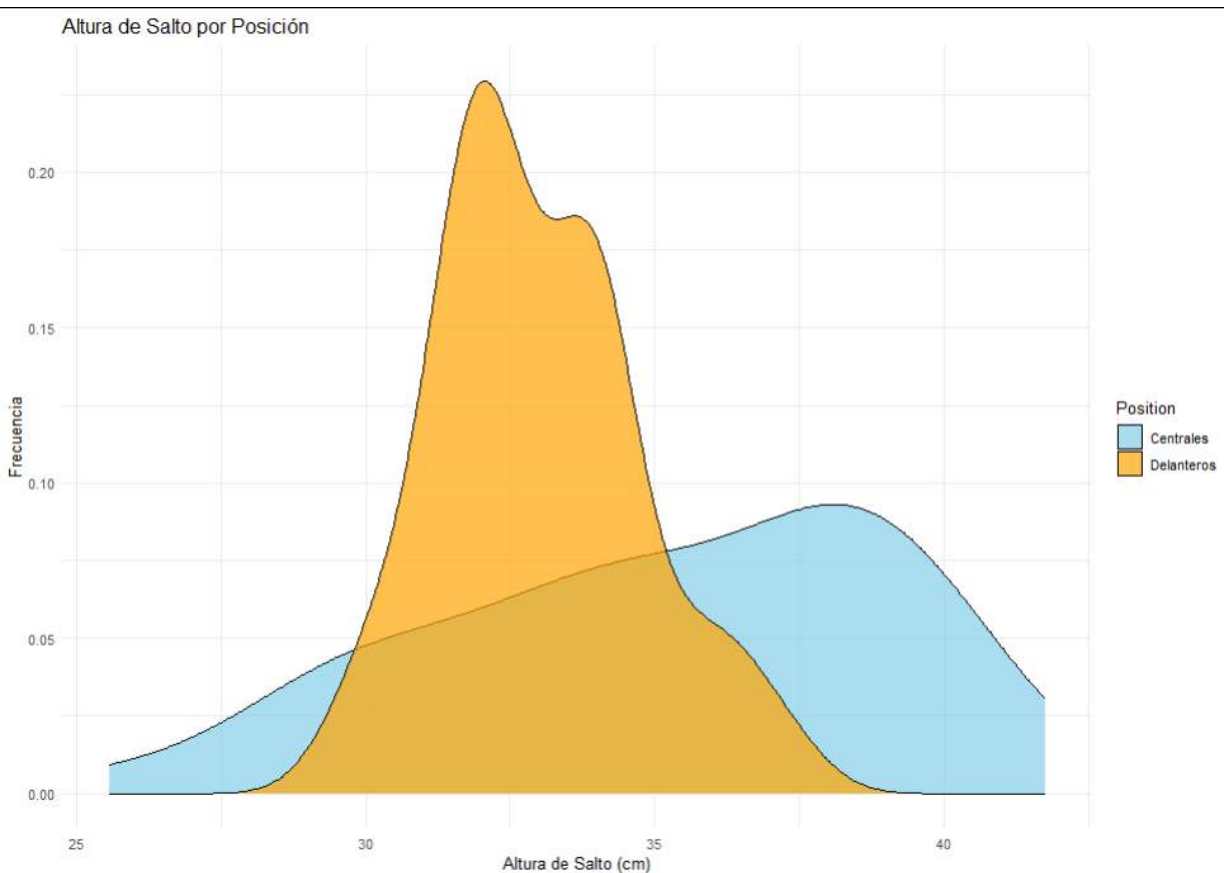


Fuente: Elaboración propia

- Visualizar distribuciones:

- Gráfico de Densidad/Histograma: Este tipo de visualizaciones busca representar cómo se organizan los datos a lo largo una serie de valores, son de gran utilidad para ver qué valores son más comunes dentro de nuestra población así como ver a simple vista los outliers o valores extremos.

**Figura 2: Gráfico de densidad**

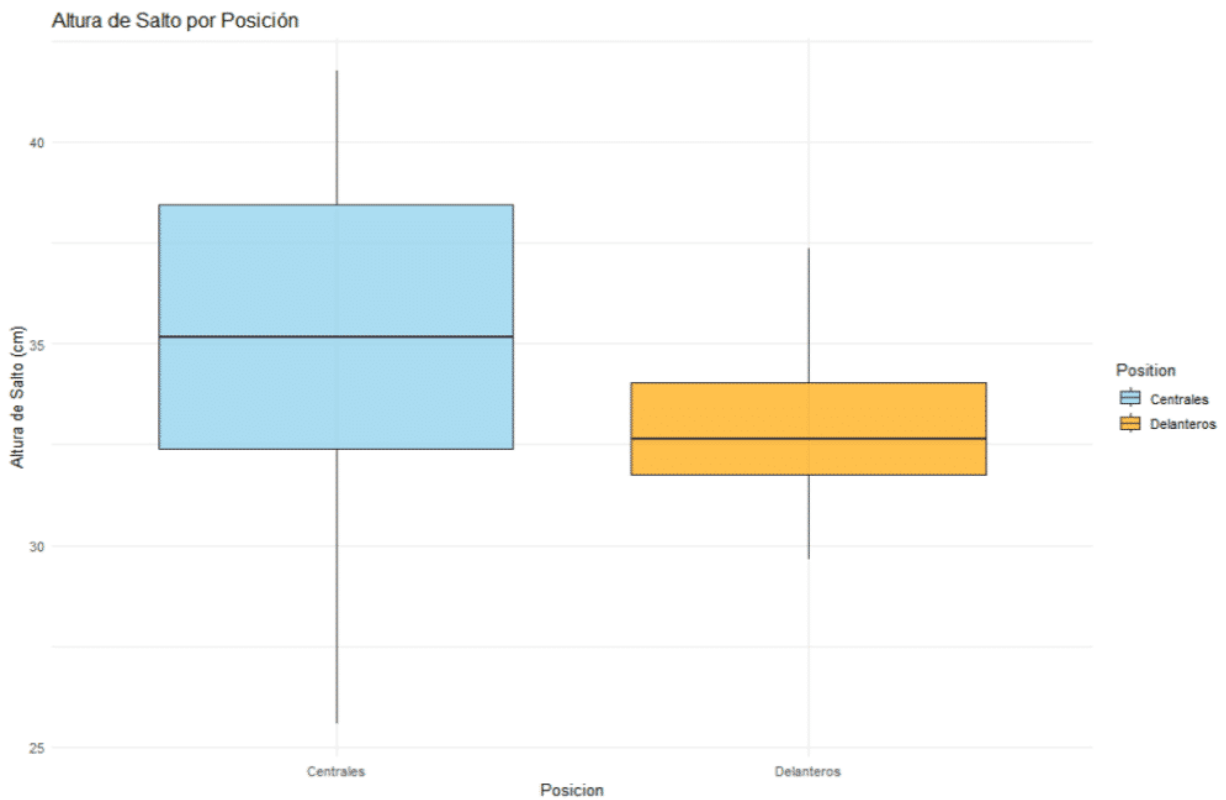


Fuente: Elaboración Propia.

---

- - Gráfico de puntos: Esta puede ser una alternativa a el gráfico de barras en el caso que queramos hacer una representación más certera de cómo se distribuyen los resultados considerando todos los individuos de los cuales queremos destacar el valor acumulado total (Torres-Ronda & Curtis, 2024).
  - Gráfico de cajas o boxplot: Se trata de una simplificación de un gráfico de densidad, pero es muy útil para hacer comparaciones ya que muestra el punto central de la distribución y el habitualmente el rango deseado, es decir, entre qué valores se acumula más cantidad de resultados.

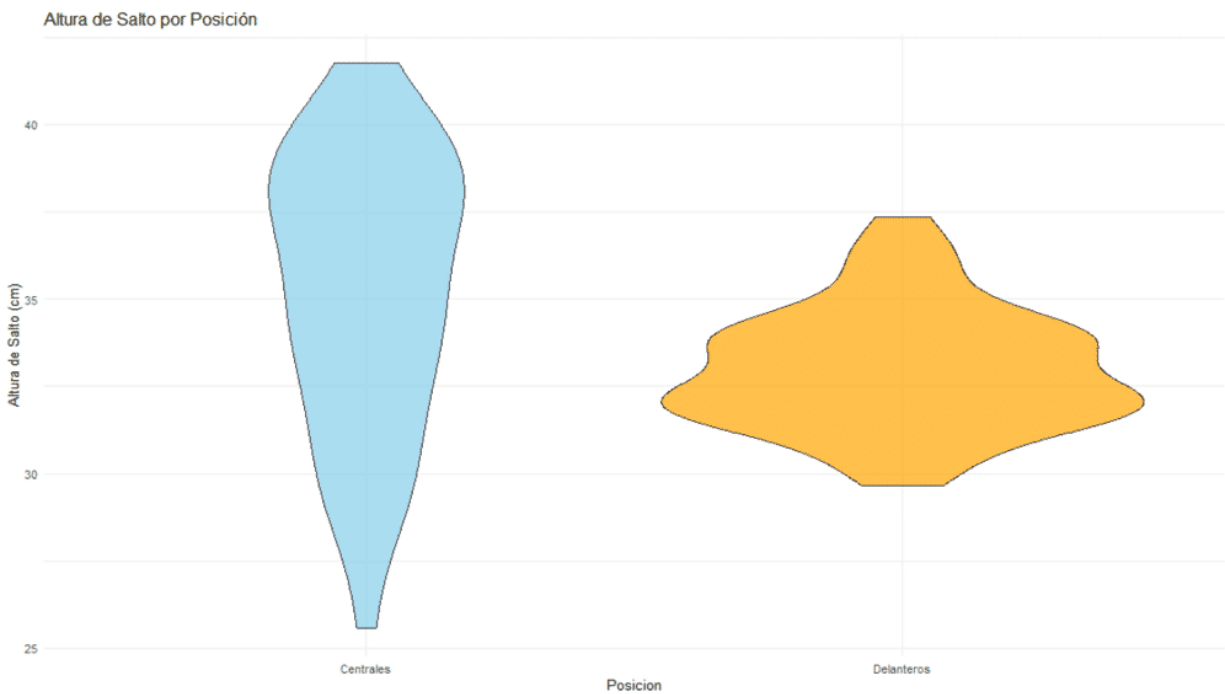
**Figura 3: Gráfico de cajas o boxplot**



Fuente: Elaboración propia

- Gráfico tipo violín: Es una versión entre el boxplot y un gráfico de densidad ya que nos da información de la forma de la distribución pero también permite ver valores descriptivos de la variable que queremos analizar.

#### Figura 4: Gráfico tipo violín



Fuente: Elaboración propia.

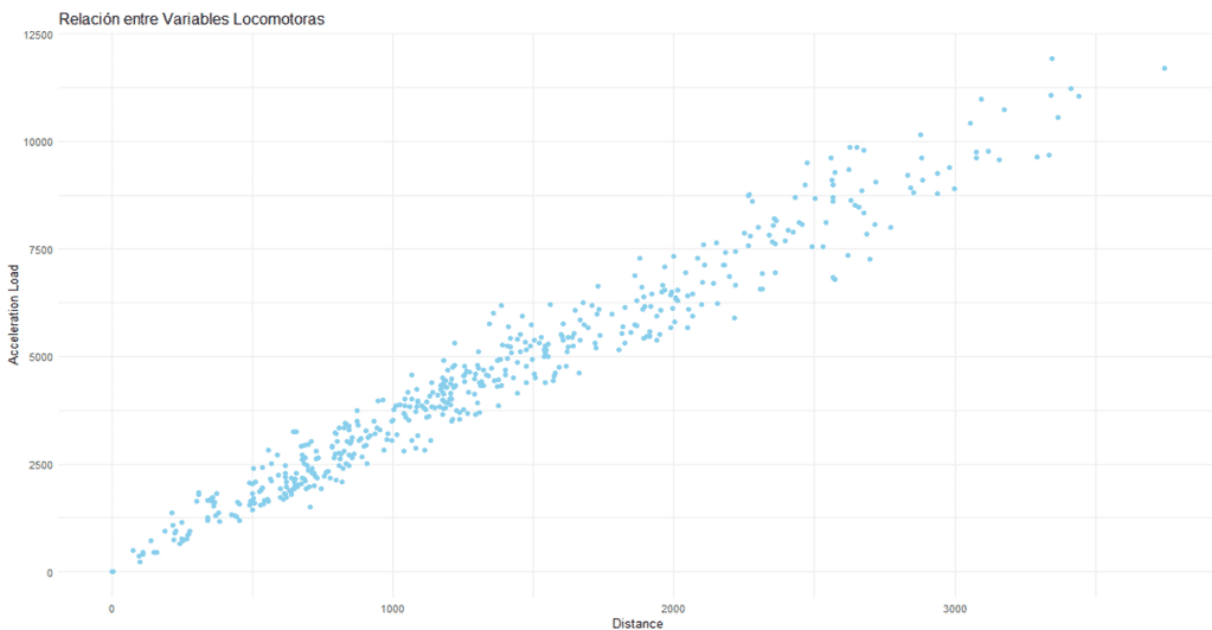
- Los tres últimos gráficos contienen los mismos datos, pero cada uno de ellos proporciona información distinta debido a las particularidades de cada tipo de visualización.
- Visualizar proporciones:
  - Gráfico de tarta: Habitualmente utilizados para representar porcentajes de un total. Tienen un gran hándicap que es su difícil lectura en muchas ocasiones cuando hay una gran cantidad de variables y las proporciones son pequeñas por lo que deberíamos optar por otro tipo de visualización

- Columnas apiladas: El mismo concepto se aplica a este tipo de gráfico. La visualización debe ser clara para cumplir su objetivo.

- Visualizar asociaciones:

- Gráfico de puntos/scatterplot: Este tipo de gráficos se utiliza cuando queremos representar la relación entre dos variables cuantitativas.

**Figura 5: Gráfico de puntos**



Fuente: Elaboración propia

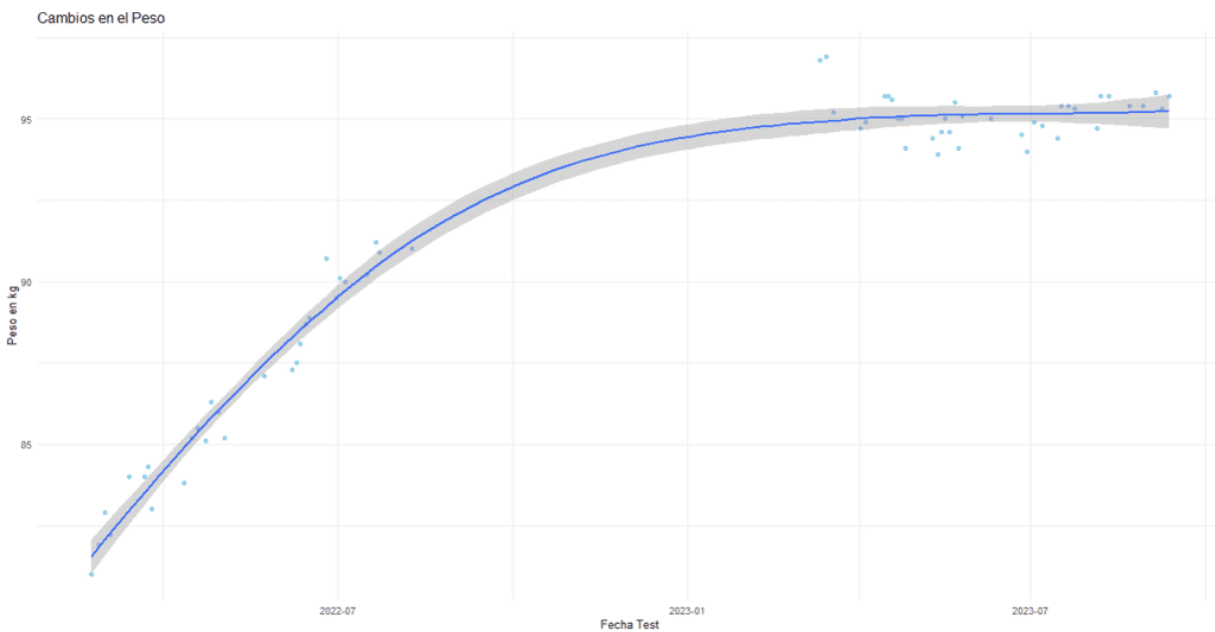
---

- Visualizar series temporales: Hablamos de series temporales cuando queremos representar el cambio

de una variable a lo largo de intervalos de tiempo determinado.

- Visualizar Tendencias: Puede tratarse de una visualización complementaria a la anterior o utilizarla sin ser necesaria la presencia de una variable tiempo. Cuando tratamos con gran cantidad de datos, cuyas observaciones pueden estar muy juntas y queremos ver en qué dirección se van moviendo los datos a lo largo del tiempo, las tendencias nos permiten simplificar las observaciones.

**Figura 6: Tendencias**



Fuente: Elaboración propia.

---

**Autevaluación:** Resaltado en amarillo la opción correcta

El primer paso que debemos realizar antes de visualizar los resultados es escoger qué datos queremos comunicar, y a partir de esta decisión seguir el primer principio. ¿Cuál es ese principio?

- Escoger la visualización correcta para el tipo de dato que queremos representar

#### CONSIDERACIONES DEL FORMATO DE LAS VISUALIZACIONES.

Aunque el punto fundamental sea qué tipo de visualización escoger para cada tipo de datos y la información que queremos comunicar, se debe tener en consideración que la forma en que llegue al receptor final tiene que ser apetecible, sin errores y de fácil asimilación. Para cumplir este objetivo es necesario tener en cuenta una serie de principios (Irizarry,2019):

- Evitar Distracciones dentro del gráfico: Mantener la información sencilla para que el mensaje llegue directamente. Podemos poner de ejemplo un gráfico de barras pero en formato 3D. Este tipo de gráfico presenta la misma información que un gráfico de barras como el que hemos visto anteriormente, pero añadiendo el 3D podemos provocar confusión al añadir perspectivas y volúmenes.
- Elección de colores: Si queremos representar varios grupos dentro de nuestra visualización, el número de esos grupos debe ser el menor posible. Cuantos más colores distintos, más difícil de visualizar los detalles de cada uno de ellos. Por ejemplo, si hablamos de un equipo de fútbol. Escoger grupos

posicionales más generalizados (Delantero es mejor opción que Extremo - Punta – Media Punta).

- Observaciones solapadas: Es muy común en gráficos de puntos como el que hemos visto anteriormente. En ese caso particular el objetivo era representar una asociación entre dos variables y no había distinción entre varios grupos o colores por lo que esas observaciones que quedan “escondidas” debajo de otras no sería necesario diferenciarlas. En el caso del gráfico de distribución, si no se hubiera escogido difuminar los colores, perderíamos información de como se distribuye cada grupo. Esta sería una de las opciones, otras opciones pueden ser dispersar los puntos. RStudio y sus paquetes de visualización permiten usar estas opciones.

- Considerar los rangos de los ejes: Este apartado es fundamental en el caso de los gráficos de barras. Es importante que el eje y siempre empiece en 0, de esta manera podemos tener una idea real de cuáles son las cantidades reflejadas y su magnitud. En el caso de otro tipo de gráficos como los scatterplots, los ejes estarán escalados a el rango de las observaciones, lo que permitirá ver más detalle entre observaciones.

- Evitar redundancias en los colores o información dentro del gráfico: En el caso del boxplot del ejemplo anterior, se ha escogido dar color a cada una de las cajas para mantener el formato del resto de gráficos, pero vemos como ya existe información del tipo de población en el eje x, por lo que podría ser omitido. Si añadiéramos además una leyenda describiendo los colores, se trataría de un gráfico muy redundante.

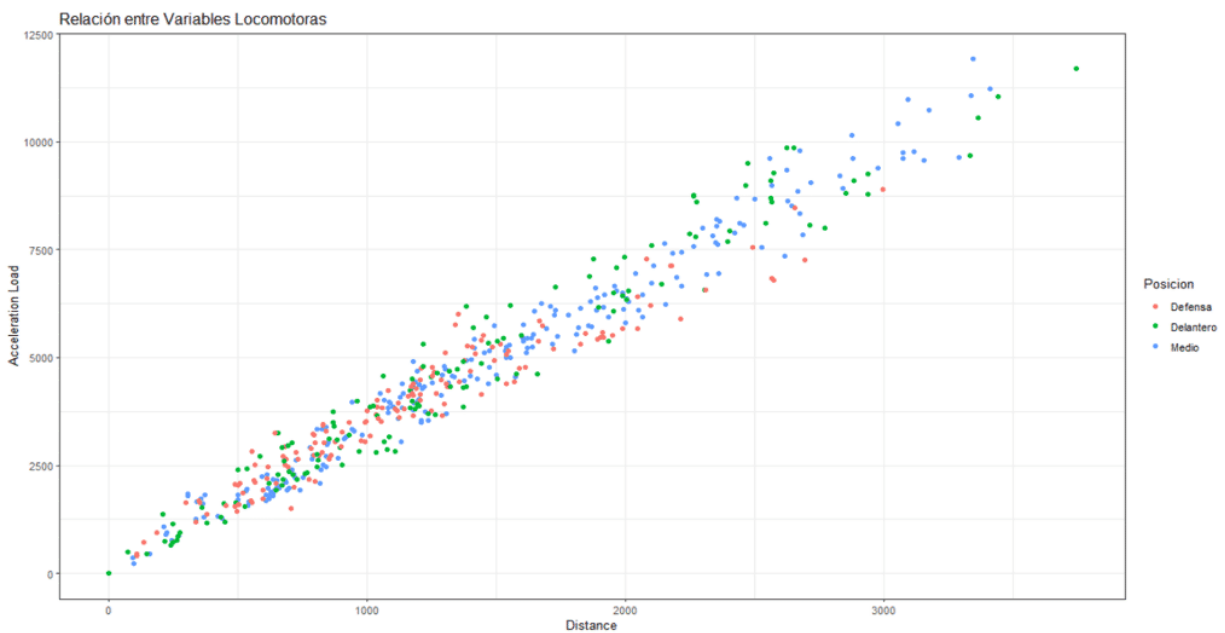
- o Lo mismo ocurre con las formas, si en un scatterplot damos un color por cada uno de los grupos, no sería necesario dar también una forma diferente (triángulo, cuadrado y rombo).

- Texto: Mantener el texto de los ejes claro y visible, así como el de los títulos.

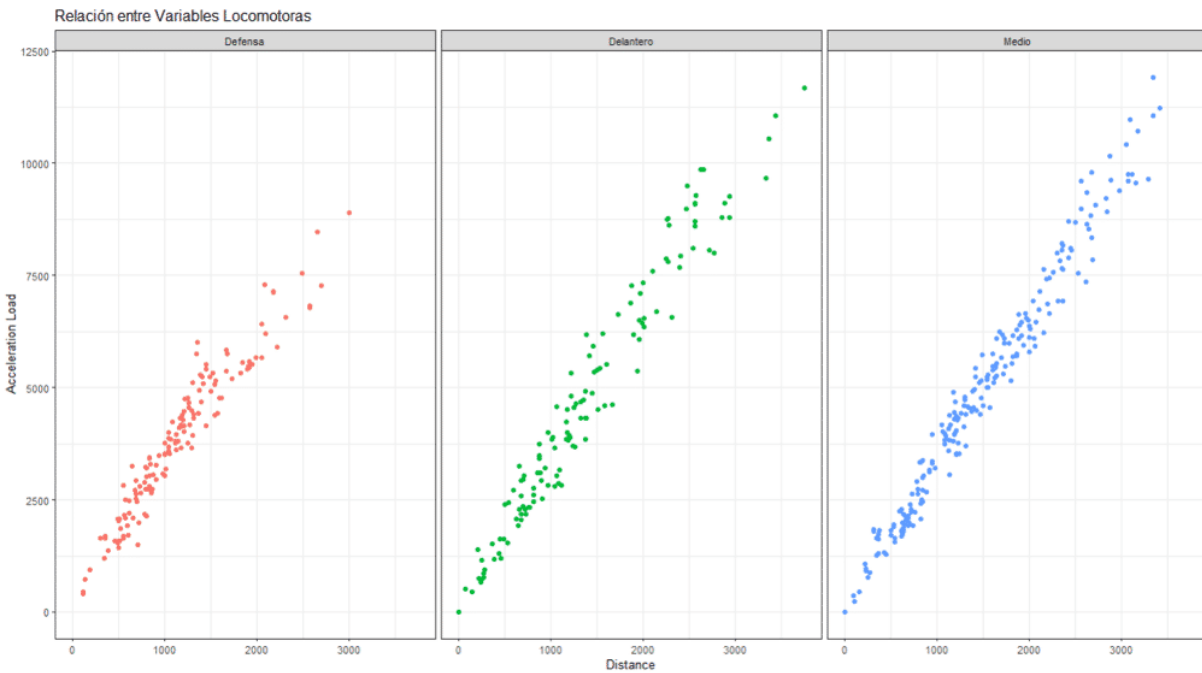
o Podemos añadir texto del gráfico para anotaciones que puedan facilitar la comprensión y el seguimiento de la visualización.

- Dividir una visualización en múltiples: Si nuestra visualización tiene muchos grupos, podemos dividirla en múltiples, para resaltar los detalles de cada una de ellas.

### Figura 7: Múltiple visualización



Fuente: Elaboración propia.



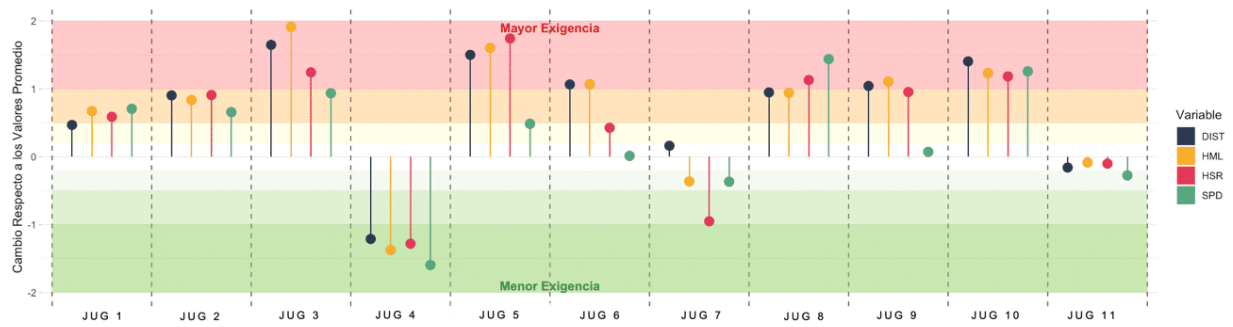
Fuente: Elaboración propia.

## Ejemplos de Comunicación de datos en Rendimiento Deportivo.

Para plasmar algunas de las consideraciones extra en el campo de visualización vamos a utilizar ejemplos reales, pero en este caso con visualizaciones más complejas que las anteriores que pretenden dar respuesta a necesidades de un equipo u organización deportiva.

La siguiente visualización permite plasmar la importancia de la simplificación de los datos para su comunicación a distintas audiencias.

## Figura 8: Simplificación



Fuente: Elaboración propia.

Este gráfico muestra las demandas condicionales, medidas mediante dispositivo GPS, de una sesión de entrenamiento de fútbol.

Veamos las particularidades de esta visualización:

- Representamos 4 variables a la vez (Distancia, High Metabolic Load, Sprint Distance y Sprinting Distance).
  - En lugar de utilizar 4 gráficos distintos y con unidades diferentes (metros, W/kg), se decide utilizar un Z-score, es decir una estandarización de los datos. De esta manera podemos comparar todas las variables entre ellas y a su misma vez entre los jugadores.
- Colores y texto que ayudan en el seguimiento de la información.

- Los colores y texto indican de manera clara que si los puntos y líneas están en la zona roja, ha sido una sesión de mayor demanda para el jugador, si por el contrario, se encuentran por debajo, los valores de esa sesión son inferiores a los promedios del jugador. A simple vista, cualquier miembro del staff puede ver qué tipo de sesión se ha realizado.
- Niveles de detalle. Según el nivel de experiencia del profesional que esté viendo el gráfico podrá obtener más o menos detalles de los resultados y por ejemplo, ver qué métricas han sido más estimuladas en la sesión y por lo tanto si se ha cumplido el objetivo condicional de la sesión.

Los siguientes gráficos pretenden mostrar otro principio fundamental en nuestro contexto. La interacción o exploración de la visualización. Este es un punto que permite dar gran valor a los datos pero también tiene el riesgo de perder énfasis en los puntos importantes que queremos comunicar.

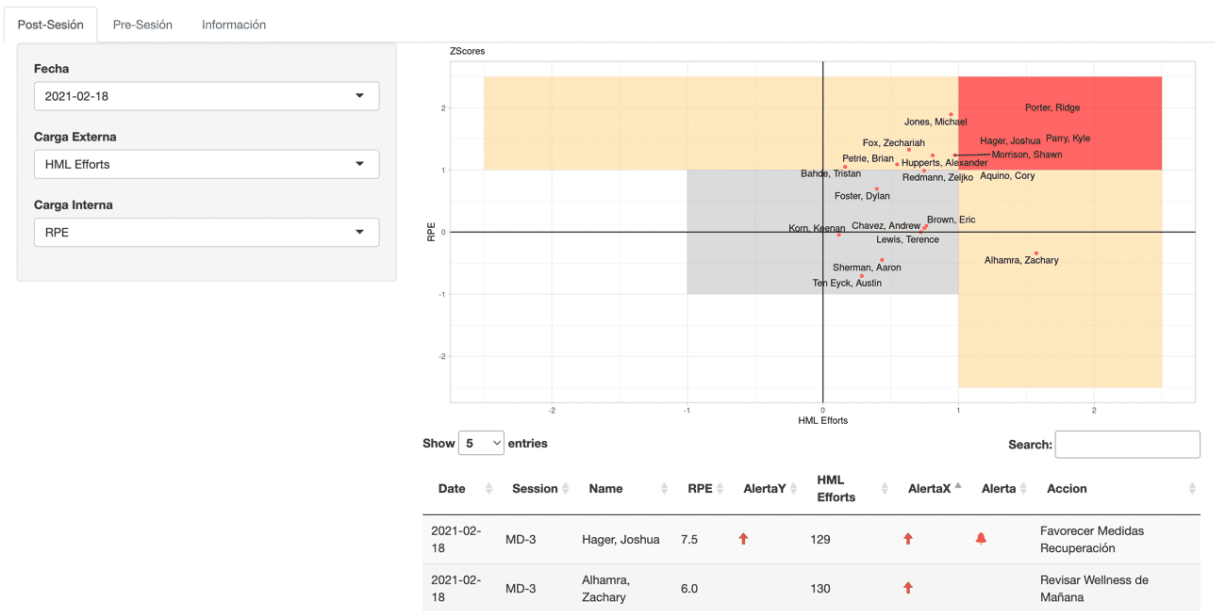
Estos gráficos pertenecen a una Shiny app, se trata de un paquete de R que permite desarrollar webs interactivas para mostrar visualizaciones y análisis a sus usuarios.

Esta aplicación está basada en el trabajo Miguel Ángel Campos (2021)

Enlace a la Shiny app:

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

**Figura 9: Shiny app**



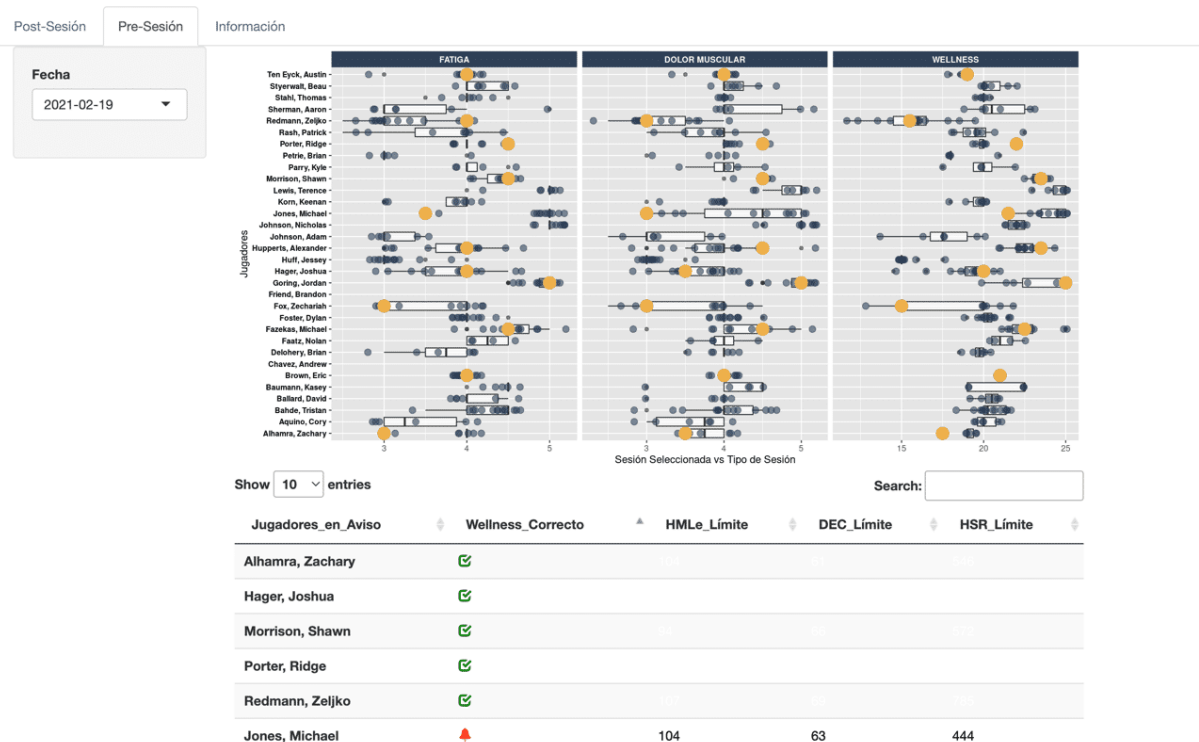
Fuente: Elaboración propia.

En la primera pestaña de la aplicación se indica que se trata del análisis post sesión, el usuario debe escoger la fecha que quiere analizar y las variables de carga externa e interna.

La selección permite ver cómo ha sido a nivel interno y externo la sesión para cada jugador, y vemos más abajo, en forma de "alertas", si esos valores han sido por encima de los valores habituales del jugador.

Cada uno de los jugadores también tendrá una recomendación al final de la sesión.

**Figura 10: Sesión**



Fuente: Elaboración propia.

Al día siguiente debemos volver a la app para observar cuáles han sido las respuestas al cuestionario de bienestar, es decir, cómo llegan al inicio de la sesión.

Tenemos también una lista con los jugadores que estaban en “alerta” después de finalizar la sesión anterior, y vemos si los resultados han



Se define como cadenas de alta densidad cuando hay partidos consecutivos con menos de 3 días de recuperación entre ellos. De esta manera podemos evaluar qué equipos tienen un calendario más ajustado, con menos descanso y por lo tanto más exigente. Este tipo de visualizaciones puede ayudar a distintos departamentos a ajustar planes de recuperación, entrenamiento y viaje.

La elección de colores, en el anterior gráfico y en el posterior permite seguir otro de los principios fundamentales, resaltar lo que queremos comunicar. En el primer caso distinguimos claramente qué equipos y en qué fechas han tenido un calendario más exigente. En el segundo gráfico, las distancias recorridas por un equipo quedan bien representadas.

### **Figura 12: Distancias recorridas**



Fuente: Elaboración propia.

Se define como cadenas de alta densidad cuando hay partidos consecutivos con:

- menos de 3 días de recuperación entre ellos
- Más de 3 días de recuperación entre ellos



Menos de 5 días de recuperación entre ellos



Más de 5 días de recuperación entre ellos

SUBMIT

CONTINUAR

## Referencias

---

**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.

CONTINUAR