

## UpSet Plots

Los UpSet plots son un tipo de visualización de datos que se utiliza para representar y analizar intersecciones y relaciones entre conjuntos de datos. Los UpSet plots son especialmente útiles para comparar conjuntos de datos de gran tamaño o complejos, donde la representación tradicional de diagramas de Venn no es práctica o eficiente.

La principal característica de los UpSet plots es que representan conjuntos y sus intersecciones como matrices de puntos o barras, lo que permite una visualización clara y estructurada de las interacciones entre múltiples conjuntos de datos. Se usan comúnmente en estudios genómicos y de biología molecular, donde los investigadores pueden estar interesados en encontrar genes comunes entre diferentes experimentos o condiciones.

Los UpSet plots pueden ser creados y personalizados utilizando varias herramientas y bibliotecas de visualización, como R, Python y JavaScript. Por ejemplo, en Python, se puede utilizar la biblioteca `UpSetPlot` para crear UpSet plots.

### Uso de los UpSet plots

Se muestra una matriz de UpSet en la parte izquierda, donde cada fila corresponde a una intersección de múltiples conjuntos, que se muestran en las columnas. Se destacan en color tres intersecciones (solo en B; en A y B, pero no en C; y en todos A, B y C) y se muestran los segmentos correspondientes en un diagrama de Venn. El diagrama de Venn muestra con el mismo color, las intersecciones que están etiquetadas en el diagrama de la izquierda, para mostrar cuál es el conjunto al cual se hace referencia:

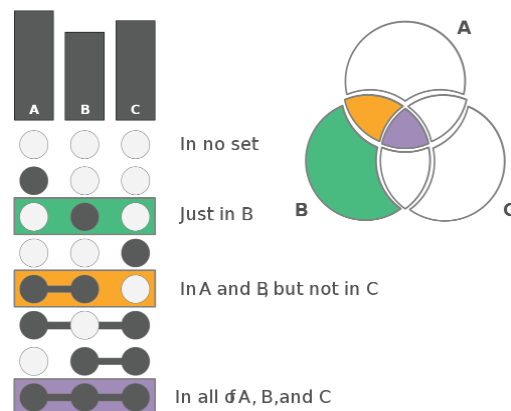
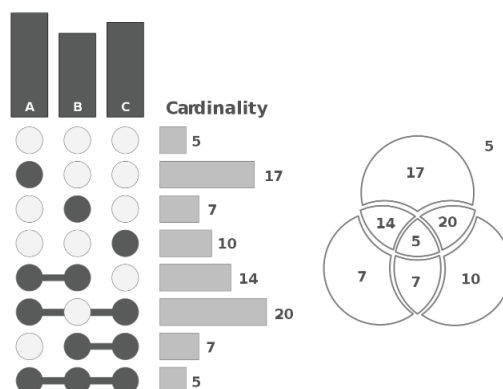


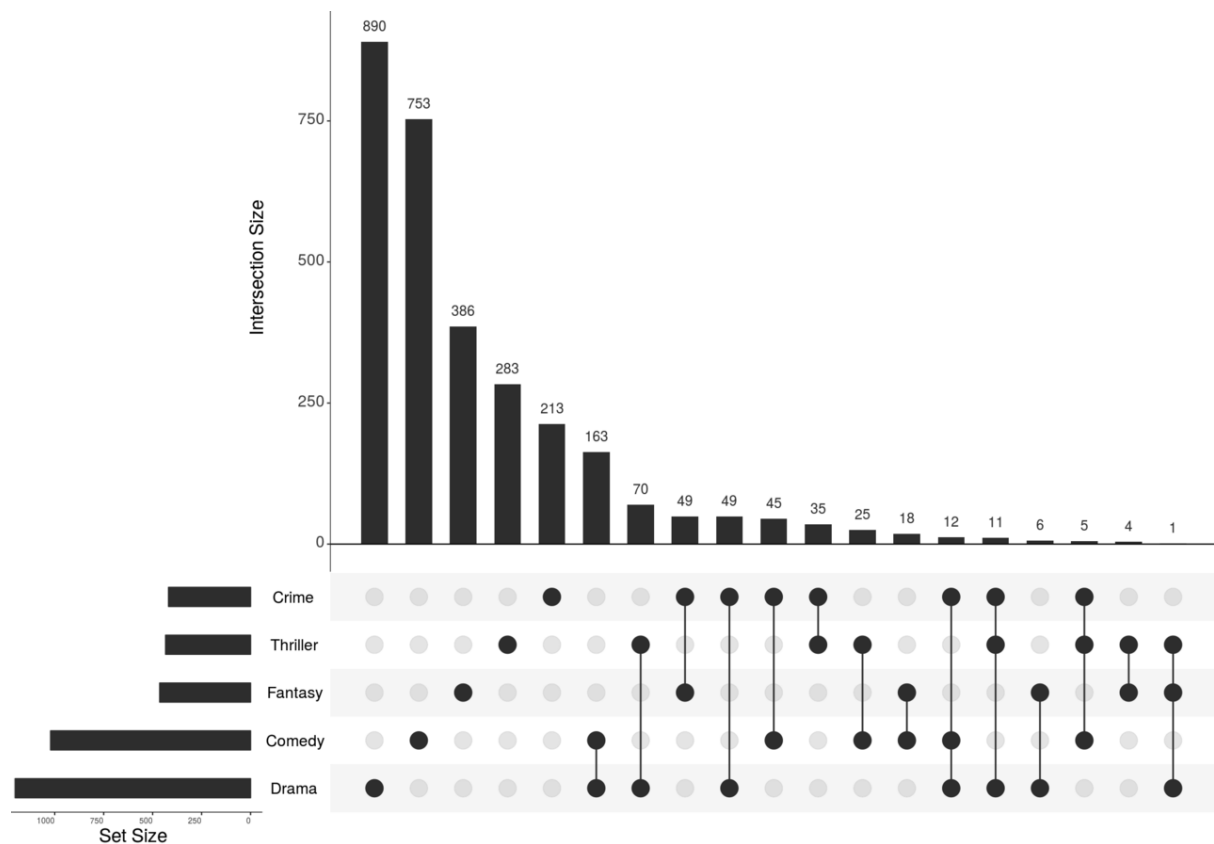
Ilustración que muestra cómo las filas de un gráfico UpSet se corresponden con los segmentos de un diagrama de Venn.

Los gráficos de UpSet visualizan intersecciones entre conjuntos en una matriz. En un gráfico de UpSet vertical, las columnas de la matriz corresponden a los conjuntos y las filas corresponden a las intersecciones. Para cada fila, las celdas que forman parte de una intersección se rellenan. Si hay varias celdas rellenas, se conectan con una línea para enfatizar la dirección de lectura del gráfico. Como los conjuntos varían en tamaño, el tamaño del conjunto se representa como gráficos de barras en la parte superior de las columnas. El tamaño de las intersecciones (cardinalidad) se muestra alineado con las filas, también como gráficos de barras. Este diseño facilita la comparación entre los tamaños de las intersecciones individuales, ya que el tamaño de las barras es fácil de comparar. Los gráficos de UpSet se pueden usar horizontal y verticalmente.



Una gráfica de UpSet que muestra tres conjuntos y el diagrama de Venn correspondiente

Los gráficos de UpSet se pueden ordenar de varias maneras. Un enfoque común de ordenación, por ejemplo, es ordenar por cardinalidad (el tamaño de una intersección), lo que coloca las intersecciones más grandes en la parte superior. Las clasificaciones alternativas son por grado de intersección o por conjuntos.



Una trama UpSet que muestra cómo se cruzan los géneros cinematográficos. La intersección más grande de dos conjuntos es la intersección Comedia-Drama.

Para crear un diagrama de este tipo con Python, podemos escribir algo como:

```
# Instalar la librería "upsetplot", usando el comando:
# !pip install upsetplot
import matplotlib.pyplot as plt
from upsetplot import from_contents, plot

# Conjuntos de datos de ejemplo
set1 = {1, 2, 3, 4, 5}
set2 = {4, 5, 6, 7, 8}
set3 = {5, 6, 9, 10}

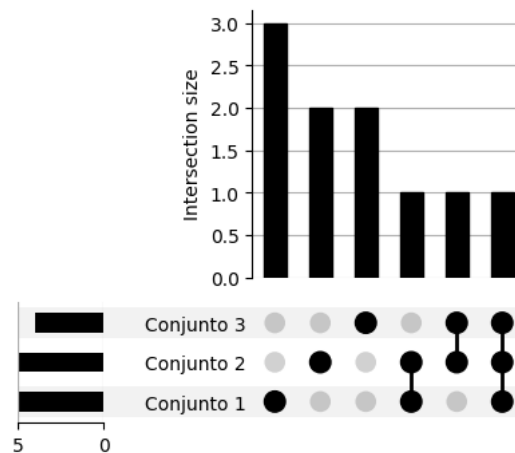
# Asignar nombres a los conjuntos de datos y crear un diccionario
data = {
    'Conjunto 1': set1,
    'Conjunto 2': set2,
    'Conjunto 3': set3,
}

# Crear conjuntos de intersecciones y conteos
example = from_contents(data)

# Crear un UpSet plot usando la biblioteca UpSetPlot
plot(example)

# Mostrar el gráfico
plt.show()
```

Donde se obtiene un gráfico como el siguiente:



- La columna 1 tiene tamaño 3, ya que el conjunto 1 tiene los elementos {1, 2, 3} que están solamente en dicho conjunto.
- La columna 2 tiene tamaño 2, ya que el conjunto 2 tiene los elementos {7, 8} que están solamente en dicho conjunto.
- La columna 3 tiene tamaño 2, ya que el conjunto 3 tiene los elementos {9, 10} que están solamente en dicho conjunto.
- La columna 4 tiene tamaño 1, ya que solamente tienen un elemento en común los conjuntos 1 y 2 que no está en 3 (elemento {4}).
- La columna 5 tiene tamaño 1, ya que solamente tienen un elemento en común los conjuntos 2 y 3 que no está en 1 (elemento {6}).
- La columna 6 tiene tamaño 1, ya que solamente tienen un elemento en común los conjuntos 1, 2 y 3 (elemento {5}).

#### Beneficios y limitaciones

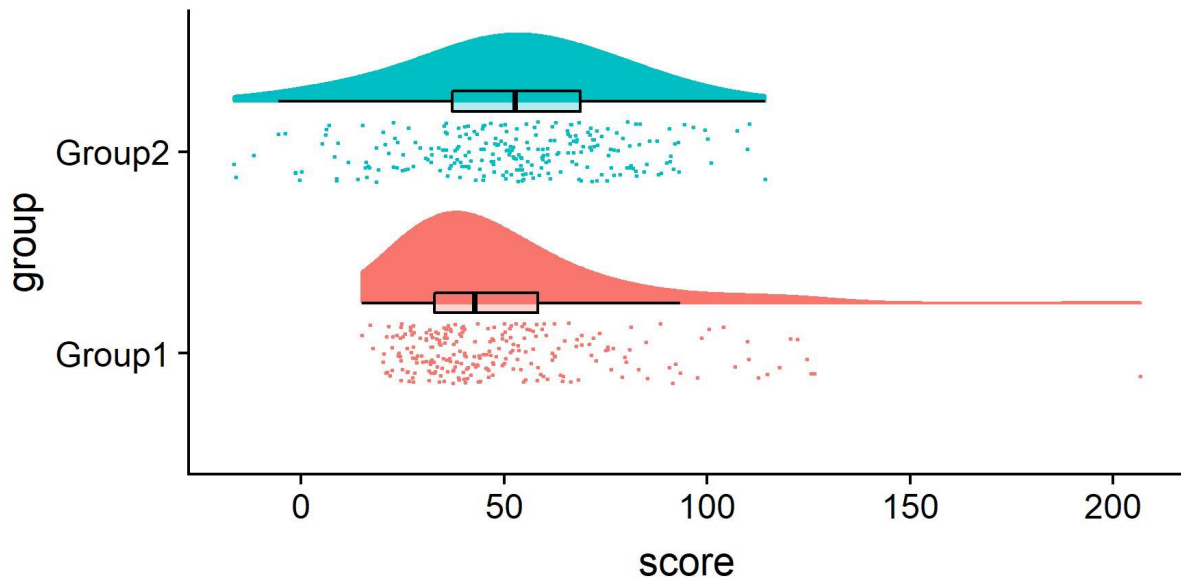
Los diagramas de UpSet tienden a funcionar mejor que los diagramas de Venn para un mayor número de conjuntos y cuando es deseable mostrar también información contextual sobre las intersecciones de conjuntos. Para visualizar diagramas con menos de tres conjuntos, o cuando solo hay unas pocas intersecciones, generalmente se prefieren los diagramas de Venn y Euler, porque tienden a ser más familiares e intuitivos de leer.

Los gráficos UpSet se limitan a mostrar 20-30 conjuntos, aunque los detalles dependen de los datos reales. Un enfoque alternativo para conjuntos de datos más grandes es mostrar un mapa de calor de co-ocurrencia, aunque estos no pueden mostrar intersecciones de orden superior.

#### Raincloud plots (diagrama de nubes de lluvia)

Adaptación de "[Raincloud plots: a multi-platform tool for robust data visualization](#)"

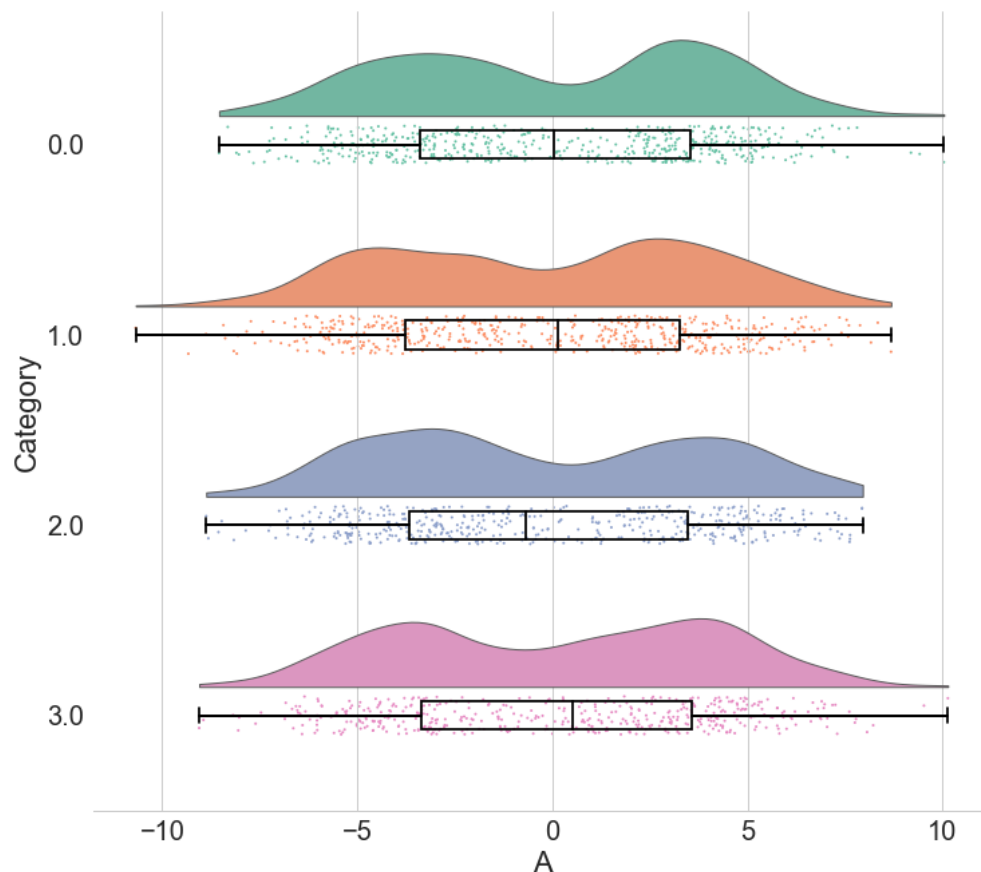
La visualización efectiva de datos es esencial para la interpretación y comunicación del análisis de datos. La herramienta "raincloud plot" es una solución de código abierto y multiplataforma que combina diferentes enfoques de visualización para ofrecer una representación de datos más robusta y comprensible. A diferencia de los gráficos de barras, que pueden ser propensos a distorsiones y malinterpretaciones, el gráfico de nube de lluvia combina elementos de gráficos de violín, puntos de datos sin procesar con dispersión aleatoria y una visualización estándar de tendencia central (como la media o mediana) y error, como un diagrama de caja, aprovechando técnicas conocidas para lograr una representación más clara de los datos.



Ejemplo de gráfico de nubes de lluvia. El diagrama de nubes de lluvia combina una ilustración de la distribución de datos (la "nube") con datos sin procesar alterados (la "lluvia"). Esto se puede complementar aún más agregando diagramas de caja u otras medidas estándar de tendencia central y error.

Los gráficos de nube de lluvia (raincloud plots) tienen ventajas sobre otros tipos de gráficos tradicionales, como los gráficos de barras o de líneas, debido a varias razones:

- Representación de datos sin procesar: Los raincloud plots incluyen la visualización de puntos de datos individuales con dispersión aleatoria (jitter), lo que permite apreciar la variabilidad y las diferencias dentro de los datos. Esto es útil para detectar patrones inesperados y valores atípicos que podrían pasar desapercibidos en otros tipos de gráficos.
- Mayor información estadística: Estos gráficos combinan diferentes enfoques de visualización, como gráficos de violín y de caja, lo que proporciona información adicional sobre la distribución, la tendencia central y la dispersión de los datos. Esta combinación de elementos ofrece una visión más completa de la naturaleza de los datos en comparación con otros gráficos tradicionales.
- Flexibilidad y modularidad: Los raincloud plots permiten a los usuarios adaptar el gráfico a sus necesidades específicas y preferencias personales. Por ejemplo, pueden incluir o excluir ciertos elementos, como gráficos de violín, puntos de datos individuales o diagramas de caja, según la información que deseen destacar en su presentación.



Las gráficas de Raincloud dejan poco librado a la imaginación: Al reemplazar la distribución de probabilidad reflejada de manera redundante con un gráfico de caja y puntos de datos sin procesar, el gráfico de nubes de lluvia brinda al usuario información sobre observaciones individuales y patrones entre ellos (como estratificación o agrupación) y tendencias generales en la distribución. Como se ilustra aquí, incluso un diagrama de caja más datos sin procesar pueden ocultar la bimodalidad u otras facetas cruciales de los datos.

Para hacer este tipo de gráficos en Python podemos hacer algo como en el siguiente ejemplo:

```
import numpy as np
import matplotlib.pyplot as plt

# Función para generar datos con ruido
def generate_data(values, probabilities, size=500):
    data = np.random.choice(values, p=probabilities, size=size)
    noise = np.random.uniform(low=-1, high=1, size=len(data))
    return data.astype(float) + noise

# Generar dos conjuntos de datos con distribuciones específicas
x1 = generate_data(np.arange(0, 11), [0.01, 0.01, 0.15, 0.19, 0.05, 0.11, 0.2, 0.16, 0.10, 0.01, 0.01])
x2 = generate_data(np.arange(6, 18), [0.01, 0.01, 0.15, 0.23, 0.14, 0.06, 0.05, 0.10, 0.12, 0.11, 0.01, 0.01])
data_x = [x1, x2]

# Preparar el gráfico
fig, ax = plt.subplots(figsize=(8, 4))

# Colores personalizados para boxplots, violin plots y scatter plots
boxplots_colors = ['yellowgreen', 'olivedrab']
violin_colors = ['thistle', 'orchid']
scatter_colors = ['tomato', 'darksalmon']

# Dibujar boxplots
bp = ax.boxplot(data_x, patch_artist=True, vert=False)
for patch, color in zip(bp['boxes'], boxplots_colors):
    patch.set_facecolor(color)
    patch.set_alpha(0.4)

# Dibujar violin plots
vp = ax.violinplot(data_x, points=500, showmeans=False, showextrema=False, showmedians=False, vert=False)
for idx, b in enumerate(vp['bodies']):
    b.get_paths()[0].vertices[:, 1] = np.clip(b.get_paths()[0].vertices[:, 1], idx + 1, idx + 2)
    b.set_color(violin_colors[idx])

# Dibujar scatter plots
```

```

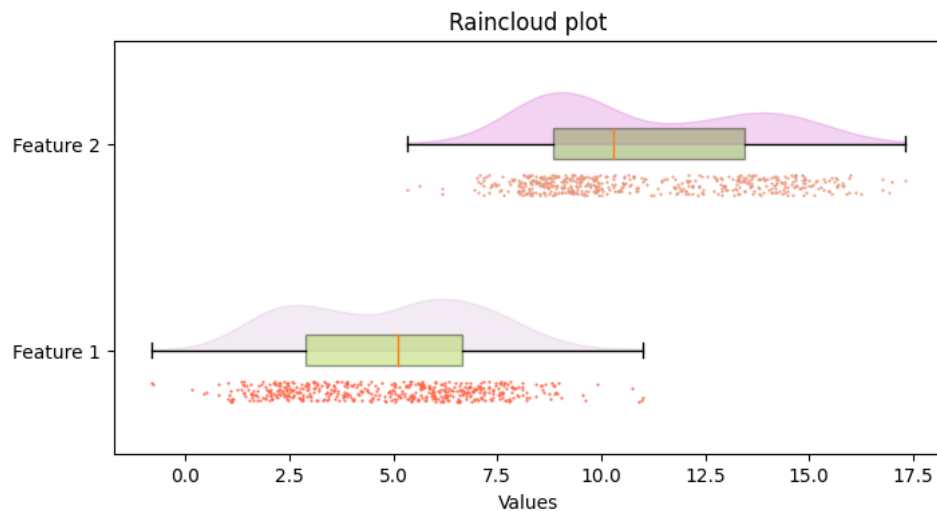
for idx, features in enumerate(data_x):
    y = np.full(len(features), idx + .8) + np.random.uniform(low=-.05, high=.05, size=len(features))
    plt.scatter(features, y, s=.3, c=scatter_colors[idx])

# Configurar las etiquetas y el título del gráfico
plt.xticks(np.arange(1, 3, 1), ['Feature 1', 'Feature 2'])
plt.xlabel('Values')
plt.title("Raincloud plot")

# Mostrar el gráfico
plt.show()

```

Y el resultado será algo como:



## Diagramas de cuerdas (Chord diagrams)

Los diagramas de cuerdas, son una creación de un científico y artista llamado [Martin Krzywinski](#). También se los conoce como "Circos", debido a una herramienta que permite generarlos.

Los diagramas de cuerdas (chord diagrams) son una representación visual utilizada para mostrar relaciones o conexiones entre elementos dentro de un conjunto de datos. Estos diagramas son útiles para visualizar datos multidimensionales y se utilizan comúnmente para representar relaciones complejas, como correlaciones, similitudes, flujos y conexiones entre diferentes nodos o entidades.

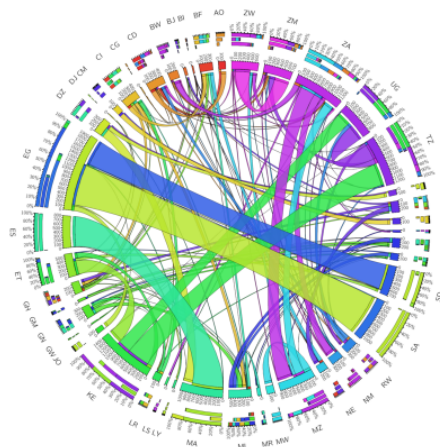
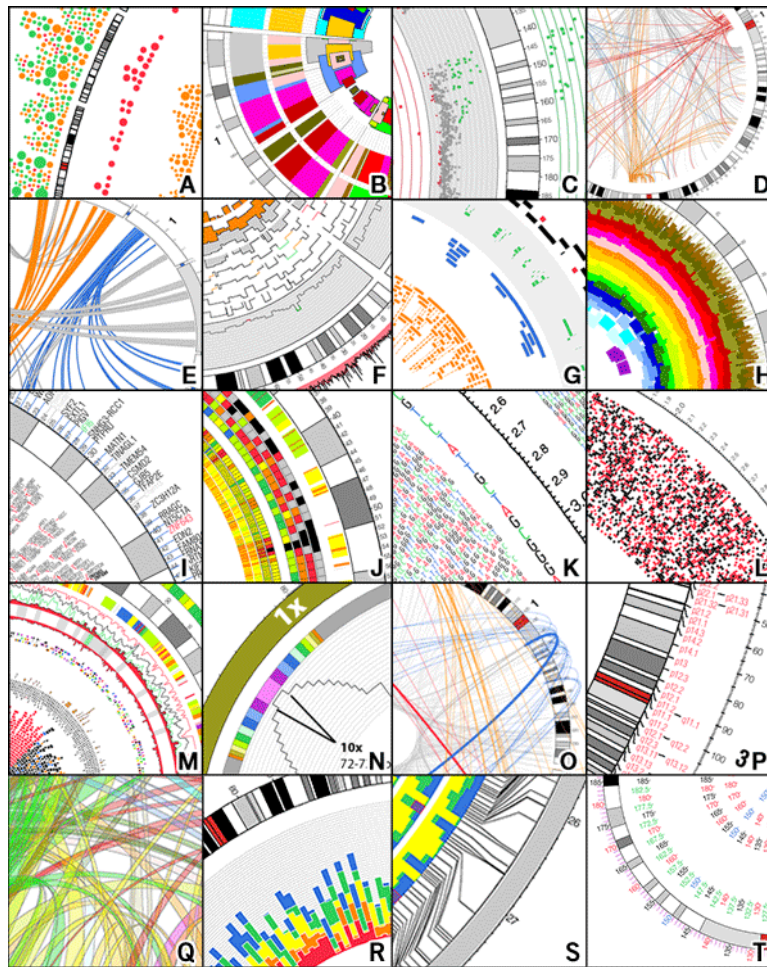


Diagrama de cuerdas que muestra el comercio de electricidad acumulado (2015-2065) entre países africanos (TWh)

Un diagrama de cuerdas se representa típicamente como un círculo dividido en segmentos proporcionales a la cantidad de nodos en el conjunto de datos. Cada segmento representa un nodo o entidad individual, y las conexiones entre estos nodos se muestran como arcos o cuerdas (chords) que conectan los segmentos. La intensidad o grosor de estos arcos puede variar en función de la fuerza o magnitud de la relación entre los nodos.

Los diagramas, son muy flexibles, ya que se pueden combinar con diagramas de otros tipos, por ejemplo de barras, de área, de densidad, etc.



Algunos ejemplos de tramas de Circos. (A) glifo (B) resaltado con control de profundidad (C) dispersión (D) ubicación emparejada (E) cinta (F) histograma (G) mosaico (H) resaltado con profundidad automática (I) texto con organización automática (J) mapa de calor (K) texto de alta densidad (L) glifo de alta densidad (M) compuesto de varios tipos (N) control de escala variable (O) control de geometría fina (P) texto flexible y colocación de elementos (Q) cintas transparentes (R) histograma apilado (S) conectores (T) anillos de graduación.

Existen decenas de aplicaciones, por ejemplo: [http://circos.ca/intro/published\\_images/](http://circos.ca/intro/published_images/)

Para crear un diagrama de este tipo en Python, podemos usar la librería `pycircize`:

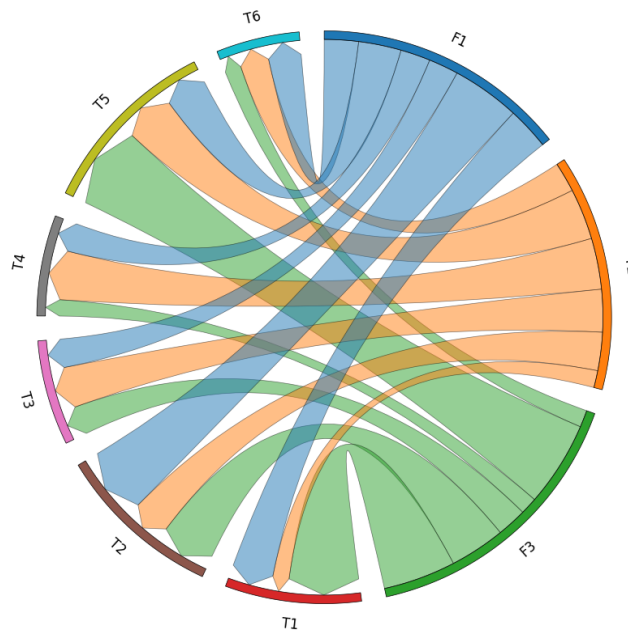
```
from pycircize import Circos
import pandas as pd

# Crear un DataFrame (3 x 6) como matriz de datos
row_names = ["F1", "F2", "F3"]
col_names = ["T1", "T2", "T3", "T4", "T5", "T6"]
matrix_data = [
    [10, 16, 7, 7, 10, 8],
    [4, 9, 10, 12, 12, 7],
    [17, 13, 7, 4, 20, 4],
]
matrix_df = pd.DataFrame(matrix_data, index=row_names, columns=col_names)

# Inicializar Circos a partir de la matriz para representar un diagrama de cuerdas
circos = Circos.initialize_from_matrix(
    matrix_df,
    space=5, # Espacio entre segmentos
    cmap="tab10", # Mapa de colores
    label_kws=dict(size=12), # Atributos de las etiquetas
    link_kws=dict(ec="black", lw=0.5, direction=1), # Atributos de los enlaces (cuerdas)
)

# Guardar el diagrama de acordes como imagen
circos.savefig("example02.png")
```

Y obtendremos un diagrama como el siguiente:



## Gráficos interactivos y dashboards

Traducción de "Interactive Data Visualization with Python: Present your data as an effective and compelling story" (Abha Belorkar, Sharath Chandra Guntuku, Shubhangi Hora, Anshu Kumar) - 2020

### Introducción

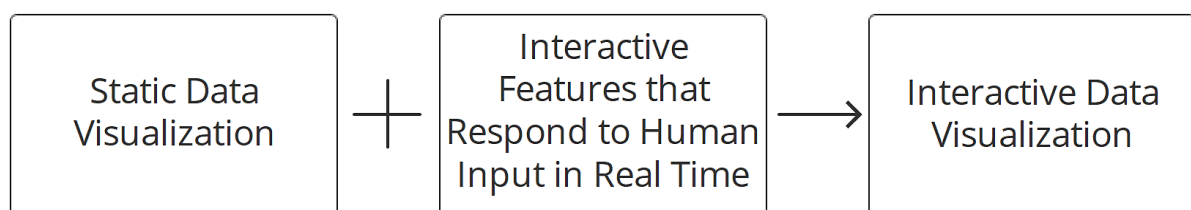
Las visualizaciones de datos son representaciones gráficas de información y datos. Su propósito es extraer valores de múltiples filas y columnas de números y datos que de otro modo serían difíciles de comprender, y representarlos en diagramas gráficamente atractivos. Como resultado, las visualizaciones de datos pueden proporcionar información clave sobre los datos de un vistazo. Esto es algo que los datos sin procesar, e incluso los datos analizados en forma tabular, no pueden hacer. Las visualizaciones vistas anteriormente se tratan de datos estáticos que están estancados, y donde la audiencia no puede modificarlos ni interactuar con ellos en tiempo real.

Las visualizaciones de datos interactivas van un paso por delante de las estáticas. La definición de interactivo es algo que involucra la comunicación entre dos o más cosas o personas que trabajan juntas.

Por lo tanto, las visualizaciones interactivas son representaciones gráficas de datos analizados (estáticos o dinámicos) que pueden reaccionar y responder a las acciones del usuario en el momento. Son visualizaciones estáticas que incorporan características para aceptar entradas humanas, mejorando y aumentando así el impacto que tienen los datos.

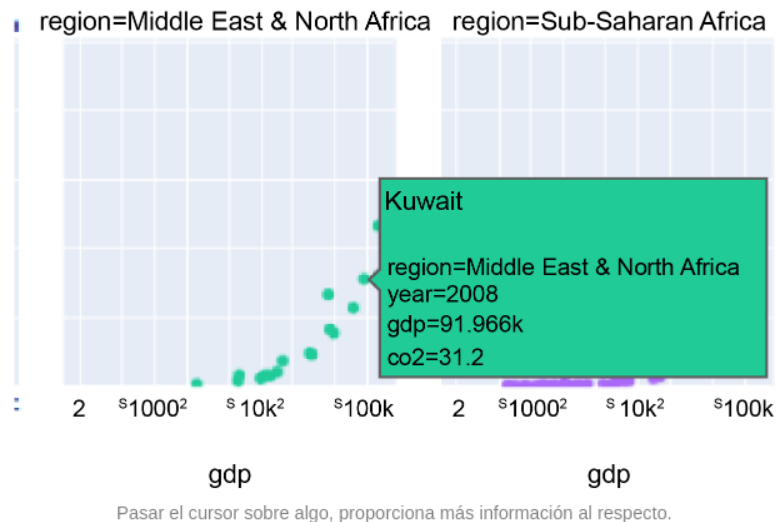
La capacidad de un gráfico para brindar más información sobre un punto de datos, cuando hay una acción del usuario (como el mouse sobre él), es lo que lo hace interactivo.

Un ejemplo de esto se puede ver en los siguientes diagramas:



Visualización interactiva de datos





Las visualizaciones interactivas a menudo también se basan en datos dinámicos. La palabra dinámico se usa para referirse a algo que cambia constantemente, y cuando se usa con respecto a las visualizaciones de datos, significa que los datos de entrada en los que se basa la visualización cambian constantemente a diferencia de los datos estáticos, que están estancados y no cambian. Un ejemplo de visualización de datos interactivos con datos dinámicos, son las visualizaciones que muestran las fluctuaciones en las tendencias de las acciones. Los datos de entrada que se utilizan para crear estos gráficos son dinámicos y cambian constantemente en tiempo real, por lo que las visualizaciones son interactivas. Los datos estáticos son más para la inteligencia empresarial, como cuando las visualizaciones de datos se utilizan como parte de un proceso de ciencia de datos/aprendizaje automático.

Para comprender las capacidades reales de la visualización interactiva, comparémosla directamente con la visualización estática.

### Visualización estática frente a visualización interactiva

Si bien las visualizaciones de datos estáticos son un gran paso adelante hacia el objetivo de extraer y explicar el valor y la información que contienen los conjuntos de datos, la adición de la interactividad lleva estas visualizaciones un paso adelante.

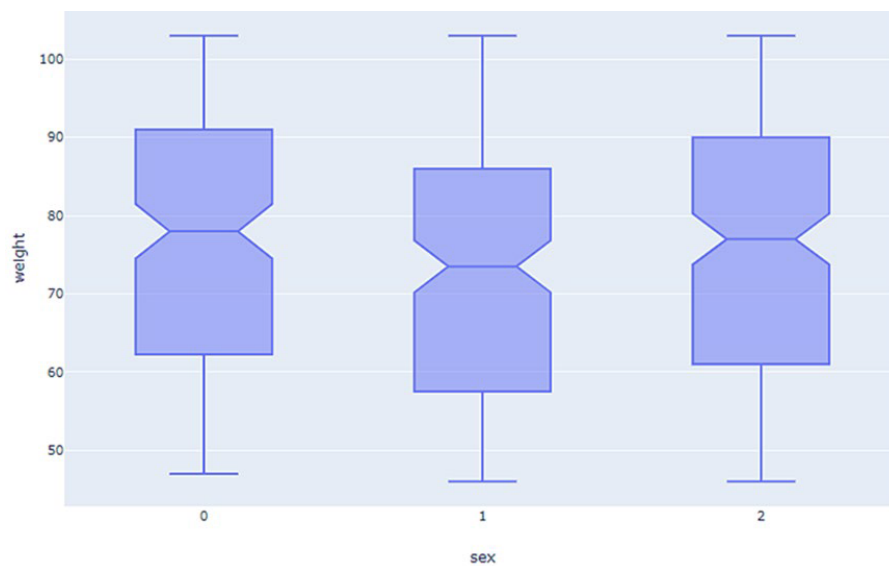
- Las visualizaciones de datos interactivos tienen las siguientes cualidades:  
Son más fáciles de explorar ya que le permiten interactuar con los datos cambiando colores, parámetros y gráficos.
- Se pueden manipular de manera fácil e instantánea. Como puede interactuar con ellos, los gráficos se pueden cambiar frente a usted. Por ejemplo, en los ejercicios y actividades de este capítulo, creará un control deslizante interactivo. Cuando se modifica la posición de este control deslizante y cambia el gráfico que ve, también podrá crear casillas de verificación que le permitan seleccionar los parámetros que desea ver.
- Permiten el acceso a datos en tiempo real ya la información que proporcionan. Esto permite el análisis eficiente y rápido de las tendencias.
- Son más fáciles de comprender, lo que permite a las organizaciones tomar mejores decisiones basadas en datos.
- Eliminan el requisito de tener varios gráficos para la misma información, un gráfico interactivo puede transmitir los mismos conocimientos.
- Te permiten observar relaciones (por ejemplo, causa y efecto).

Comencemos con un ejemplo para comprender lo que podemos lograr a través de la visualización interactiva. Consideremos un conjunto de datos para miembros inscritos en un gimnasio:

	age	weight	sex
0	29	88	2
1	45	96	1
2	35	91	0
3	37	79	1
4	27	62	0

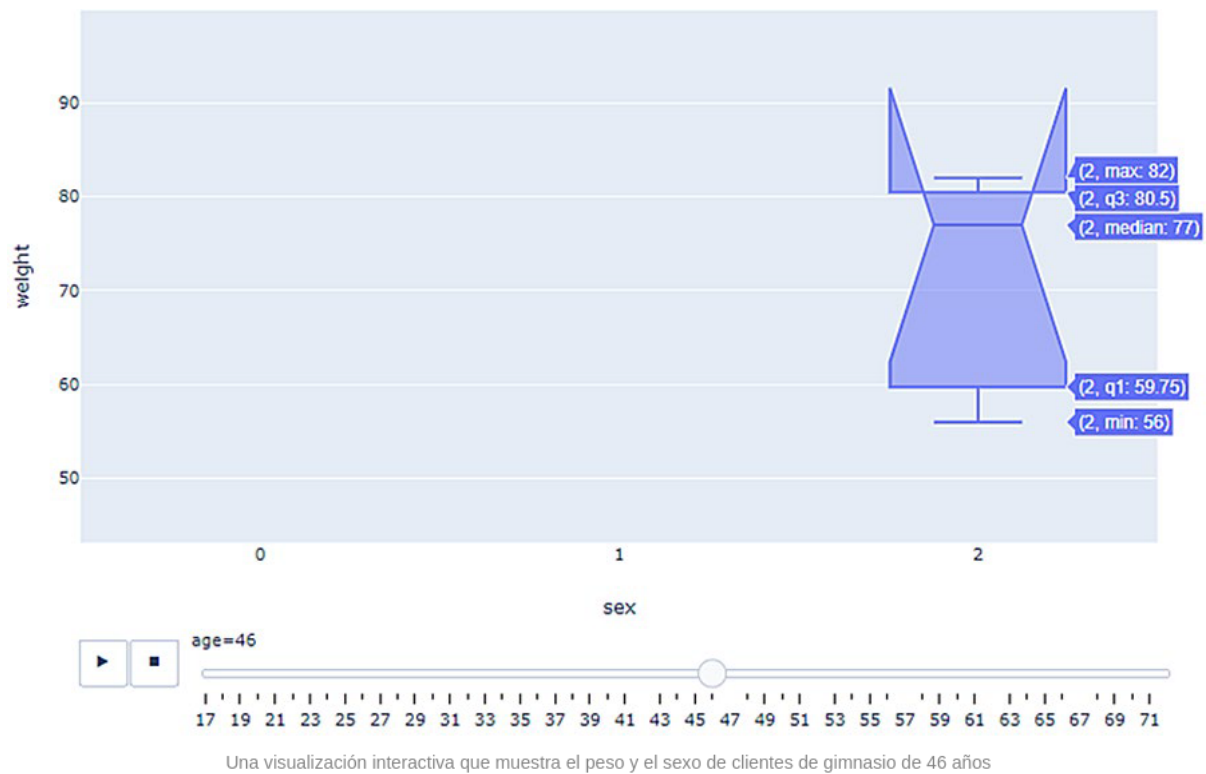
Dataset de clientes de gimnasio

La siguiente es una visualización de datos estáticos en forma de diagrama de caja que describe el peso de las personas clasificadas por su sexo (0 es hombre, 1 es mujer y 2 es otro):



Una visualización estática que muestra el peso frente al sexo de los clientes del gimnasio

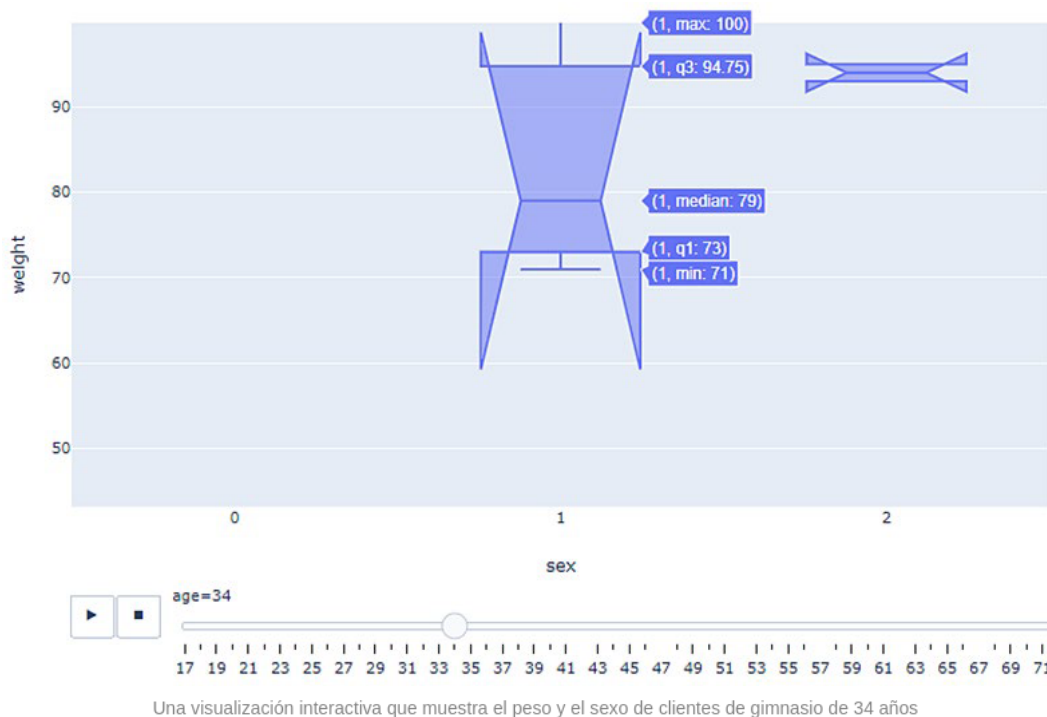
La única información que podemos obtener de este gráfico es la relación entre el peso y el sexo: los clientes masculinos que visitan este gimnasio pesan entre 62 kg y 91 kg, las clientes mujeres pesan entre 57 kg y 86 kg, y las/los clientes que se identifican como otras/os pesan entre 61 kg y 90 kg. Sin embargo, hay una tercera característica presente en el conjunto de datos que se usa para generar este diagrama de caja: la edad. Agregar esta característica al gráfico estático anterior puede generar confusión en términos de comprensión de los datos. Entonces, estamos un poco atascados con respecto a mostrar la relación entre las tres características usando una visualización estática. Este problema se puede resolver fácilmente creando una visualización interactiva, como se muestra aquí:



En el diagrama de caja anterior, se introdujo un control deslizante para la función de edad. El usuario puede deslizar manualmente la posición del control deslizante para observar la relación entre peso, sexo y edad en diferentes valores de edad. Además, hay una herramienta flotante que permite al usuario obtener más información sobre los datos.

El diagrama de caja anterior describe que, en este gimnasio, los únicos clientes de 46 años son lxs que se identifican como otrxs, y el más pesado de 46 años pesa 82 kilogramos, mientras que el más liviano pesa 56 kilogramos.

El usuario puede deslizar a otra posición para observar la relación entre peso y sexo a una edad diferente, como se muestra en el siguiente gráfico:



El gráfico anterior describe los datos a la edad de 34 años: no hay clientes masculinos en el gimnasio; sin embargo, la cliente más pesada de 34 años pesa 100 kilogramos mientras que la más liviana pesa 71 kilogramos.

Pero aún hay más aspectos a considerar al diferenciar entre visualizaciones estáticas e interactivas. Miremos la siguiente tabla:

	<i>Static Data Visualizations</i>	<i>Interactive Data Visualizations</i>
<b>Target Media/Fields</b>	Best-suited for print media and presentations	Social media applications and websites, business intelligence, and so on
<b>Cost to Create</b>	Low	High
<b>Connection to Data Source</b>	Not required	Required in cases where data is dynamic; results in complicated systems involving online databases
<b>Viewing</b>	Render easily and can be saved as images	Might require advanced UI designs
<b>Popular Python Libraries</b>	Matplotlib, Seaborn	Bokeh, Plotly

En última instancia, las visualizaciones de datos interactivas transforman la discusión de los datos en el arte de contar historias, lo que simplifica el proceso de comprensión de lo que los datos intentan decirnos. Benefician tanto a las personas que crean las visualizaciones (ya que los mensajes y la información que intentan transmitir se transmiten de manera eficiente y visualmente agradable) como a quienes las ven (ya que pueden comprender y observar patrones e ideas casi al instante). Estos aspectos son los que separan las visualizaciones interactivas de las visualizaciones estáticas.

Librerías Python

En Python tenemos librerías alternativas a la popular Matplotlib, que incorporan capacidades interactivas, por ejemplo:

- **Plotly:** <https://plotly.com/python/>
- **Bokeh:** <https://demo.bokeh.org/>

También tenemos librerías que permiten crear aplicaciones web con facilidad desde Python, que permiten construir visualizaciones interactivas, como por ejemplo:

- **Streamlit:** <https://streamlit.io/gallery>
- **Plotly Dash:** <https://plotly.com/examples/>
- **Gradio:** <https://gradio.app/demos/>

## KPI (Indicadores Clave de Performance)

Un **KPI** (*key performance indicator*), conocido también como **indicador clave o medidor de desempeño** o **indicador clave de rendimiento**, es una medida del nivel del rendimiento de un proceso. El valor del indicador está directamente relacionado con un objetivo fijado para la toma de decisiones previamente y normalmente se expresa en valores porcentuales.<sup>1</sup>

Un KPI se diseña para mostrar cómo es el progreso en un proceso o producto en concreto, por lo que es un indicador de rendimiento. Existen KPI para diversas áreas de una empresa: compras, logística, ventas, servicio al cliente, etc. Las grandes compañías disponen de KPI que muestran si las acciones desarrolladas están dando sus frutos o si, por el contrario, no se progresa como se esperaba.

Los indicadores clave de desempeño son mediciones financieras o no financieras utilizadas para cuantificar el grado de cumplimiento de los objetivos; reflejan el rendimiento de una organización y generalmente se recogen en su plan estratégico. Estos KPI se utilizan en inteligencia empresarial para reflejar el estado actual de un negocio y definir una línea de acción futura.

El acto de monitorear los indicadores clave de desempeño en tiempo real se conoce como «monitorización de actividad de negocio». Los indicadores de rendimiento son frecuentemente utilizados para "valorar" actividades complicadas de medir, como los beneficios de desarrollos líderes, el compromiso de los empleados, el servicio o la satisfacción.

### Objetivos

Los KPI tienen como objetivos principales medir el nivel de servicio, realizar un diagnóstico de la situación, comunicar e informar sobre la situación y los objetivos, motivar a los equipos responsables del cumplimiento de los objetivos reflejados en el KPI y, en general, evaluar cualquier progreso de manera constante.

### Usos Frecuentes

Usado para calcular, entre otros:

- Usado como herramienta.
- Nivel de la satisfacción del cliente.
- Tiempo de mejoras de asuntos relacionados con los niveles de servicio.
- Impacto de la calidad de los recursos financieros adicionales necesarios para realizar el nivel de servicio definido.
- Rentabilidad de un proyecto (Retorno de la Inversión ROI)
- Calidad de la gestión de la empresa (rotación del inventario, días de cuentas por cobrar [DCC], y por pagar [DCP])
- Monitorear el rendimiento de los equipos de trabajo en Campo, por ejemplo en el sector de las telecomunicaciones

### Cualidades de los KPI

Cualquier organización debe poder identificar sus propios KPI. Las claves para esto son:

- Tener predefinido un proceso de negocio.
- Tener claros los objetivos o el rendimiento requeridos en el proceso de negocio.
- Tener una medida cuantitativa/cualitativa de los resultados, y que sea posible su comparación con los objetivos.
- Investigar variaciones y ajustar procesos o recursos para alcanzar metas a corto plazo.

Cuando se definen KPI se suele aplicar el acrónimo SMART, ya que los KPIs tienen que ser:

- ESpecíficos (*Specific*)
- Medibles (*Measurable*)
- Alcanzables (*Achievable*)
- Relevantes (*Relevant*)
- OporTunos (*Timely*)

Es importante que:

1. Los datos de los que dependen los KPI sean consistentes y correctos.
2. Estos datos estén disponibles a tiempo.

### KPIs estándares

Procesos industriales: Los KPI de producción son esenciales a la hora de realizar un seguimiento de la producción y de los resultados de una empresa. En donde según el caso, se pueden encontrar diferentes indicadores de producción. Sin embargo, existen una serie de KPI que, por su importancia, van a ser comunes a todas las compañías, por lo que es esencial contar con ellos en el análisis de datos.

Cuáles son los principales KPI de producción?

La elección de los KPI de producción que se toman como referencia puede variar de una empresa a otra. Sin embargo, vamos a encontrar algunos comunes entre todas debido a la especial importancia que tienen a la hora de realizar el análisis de datos y de valorar los resultados obtenidos.

- **Tiempo de ciclo de producción:** Permite medir el tiempo que se tarda en producir un producto determinado teniendo en cuenta el lote de producción. Para ello se toma como referencia inicial el momento en el que se ejecuta la orden de producción y como referencia final el momento en el que el lote en cuestión puede darse por completado. En otras palabras, se toma el tiempo de producción del lote y se divide en las unidades conformes. Por que no se hace en base también de las unidades no conformes, por que la naturaleza del proceso ya lleva inmerso el porcentaje de no conformes (obviando que el proceso se debe ajustar). El tiempo que transcurre entre un momento y otro es el tiempo de ciclo de producción de cada producto.
- **Rendimiento de calidad:** En este caso, el KPI de producción se centra en analizar el porcentaje de productos que son fabricados sin errores. Es decir, el número de productos que, al terminar la cadena de producción, son perfectos desde el punto de vista de los estándares de calidad de la compañía.
- **Tasa de rechazo:** La tasa de rechazo es un KPI que mide los productos que no cumplen con los estándares de calidad determinados. Es decir, se trata del KPI opuesto al rendimiento de calidad, ya que, en este caso, lo que se mide es la cantidad de productos que son eliminados de la cadena de suministro antes de llegar a completar su producción.
- **Tasa de devoluciones:** Este KPI de producción hace referencia al seguimiento del producto una vez se ha completado la producción. Sin embargo, es esencial tenerlo en cuenta para adaptar la producción y reducir la tasa de devoluciones, ya que determina de forma directa el nivel de satisfacción del cliente. Una tasa de devoluciones elevadas cuando el producto cumple con los estándares de calidad determinados suele estar asociada a un diseño ineficaz o a una incapacidad del producto para cumplir con las expectativas del consumidor, por lo que, en muchos casos, será necesario reformular la producción desde el principio.

Proceso de mantenimiento:

Los indicadores clave de rendimiento de mantenimiento (KPI) son métricas que evalúan factores críticos para el éxito de una organización. Una amplia gama de empresas hace un seguimiento de los KPI. Sin embargo, en el mundo del mantenimiento, estas métricas controlan el rendimiento con respecto a los objetivos vinculados a cosas como los fallos de la máquina, los tiempos de reparación, los retrasos en el mantenimiento y los costos.

¿Cuáles son los principales indicadores de mantenimiento?

Aunque los KPIs tienen una relación directa con los objetivos que se pretende alcanzar, no son metas. Un KPI de mantenimiento es una métrica que sirve para evaluar, de forma cuantitativa, el rendimiento de una determinada actividad, activo o departamento. Los indicadores pueden dividirse en dos categorías: • Los que ponen de relieve el efecto del mantenimiento en el rendimiento del negocio • Los que se asocian con la fiabilidad y la disponibilidad de los activos. De esta categorización, podemos destacar como principales KPIs: downtime; backlog; MTBF; MTTR; OEE; PMP (planned maintenance percentage/tasa de mantenimiento planificado) y tasa de cumplimiento de mantenimiento preventivo. A continuación, explicaremos sus respectivas funciones, beneficios y cuáles son los patrones medios globales para cada uno de los indicadores.

- **Downtime** Esta métrica de mantenimiento, también conocida como Tiempo de Inactividad del Equipo, puede ser usada para rastrear, monitorear y evaluar la fiabilidad de un activo. El downtime corresponde al tiempo de inactividad no programada. Es decir, es el resultado de un acontecimiento imprevisto que requerirá algún tipo de intervención. Este KPI puede ser evaluado, independientemente de si ya existe (o no) un cronograma de mantenimiento para el equipo.
- **Backlog** El backlog es un indicador de tiempo que puede traducirse como "retraso en el mantenimiento". Representa la acumulación de actividades pendientes o en ejecución, por cada técnico o empleado, independientemente de si ya están en marcha o todavía están solo planificadas. En otras palabras, el Backlog es el tiempo de servicio necesario para realizar una determinada acción de Mantenimiento Correctivo, Preventivo o Predictivo; Inspecciones de Calidad; Mejoras o cualquier otra actividad para el buen funcionamiento de los activos. *MTBF – Mean Time Between Failures* Otro importante indicador de rendimiento del mantenimiento es el MTBF, también conocido como el Indicador de Confiabilidad. Mide la tasa de fallos aleatorios (no previstos), incluso si son causados por fallos de software o defectos de fabricación que comprometen su vida útil. Se excluyen los fallos que no causan downtime.
- **MTTR – Mean Time To Repair** Un indicador de mantenimiento igualmente común es el Tiempo Medio de Reparación (MTTR), que puede aplicarse a un equipo, máquina, componente o sistema. El MTTR considera el tiempo medio que tarda su equipo técnico en intervenir o resolver una avería después de que haya ocurrido. A diferencia del MTBF, el objetivo es reducir al máximo este KPI de mantenimiento. En cierto modo, la reducción del MTTR sirve de detonante para tomar decisiones que mejoren su estrategia de mantenimiento, siempre con el objetivo de maximizar los beneficios y reducir los riesgos, es ampliamente utilizado en los procesos en donde la pérdida de beneficios por una maquinaria inactiva es significativa.

- **OEE – Overall Equipment Effectiveness** Este es uno de los KPI de mantenimiento más importantes ya que mide la eficacia global de la empresa. Con este cálculo se podrá establecer si los procesos son eficientes o no. Uno de los beneficios de calcular la OEE es saber con qué frecuencia los equipos están en funcionamiento. También nos ayuda a conocer la rapidez con que se desarrolla la producción de la empresa y, por último, cuántos productos (o servicios) se han producido (o realizado) sin ningún tipo de fallo.
- **PMP – Planned Maintenance Percentage/Porcentaje de Mantenimiento Planificado** El Porcentaje de Mantenimiento Planificado considera el tiempo dedicado a las actividades programadas (ya sean de mantenimiento, reparación o sustitución) con los activos definidos. Este KPI de mantenimiento está directamente asociado con el Plan de Mantenimiento Preventivo de una empresa. Se considera la eficacia, el modo cómo ha transcurrido cada actividad, así como el tiempo necesario para completarla, el resultado del PMP indicará el grado de eficiencia de una empresa, así como su rendimiento y éxito en el sector del mercado en el que opera.
- **Tasa de Cumplimiento del Mantenimiento Preventivo** Esta métrica no podía faltar en una lista de los principales indicadores de rendimiento del mantenimiento. Analiza la conformidad de la empresa con el cronograma establecido.

Marketing Digital y SEO (Optimización de motores de búsqueda):

- **Tráfico orgánico:** Este KPI mide la cantidad de visitantes que llegan a tu sitio web a través de motores de búsqueda sin pagar por publicidad. Un aumento en el tráfico orgánico generalmente indica un buen rendimiento de SEO.
- **Posiciones de ranking de palabras clave:** Las palabras clave son esenciales para el SEO. Este KPI mide la posición de un sitio web en los resultados de búsqueda para las palabras clave seleccionadas.
- **Tasa de rebote:** Este es el porcentaje de visitantes que abandonan tu sitio web después de visitar solo una página. Una tasa de rebote alta podría indicar que el contenido de la página no es relevante para los visitantes.
- **Páginas vistas por sesión:** Este KPI mide el número promedio de páginas que un visitante ve durante una sesión. Un número más alto puede indicar que los visitantes encuentran el contenido relevante y útil.
- **Enlaces entrantes:** Los enlaces de otros sitios web a un sitio (backlinks) son un factor importante en la clasificación de los motores de búsqueda. Este KPI mide la cantidad y calidad de estos enlaces.
- **Tasa de conversión:** Este KPI mide el porcentaje de visitantes que completan una acción deseada, como hacer una compra, registrarse para un boletín informativo, descargar un eBook, etc.
- **Costo por adquisición (CPA):** Este KPI mide cuánto cuesta adquirir un nuevo cliente a través de los esfuerzos de marketing digital.
- **Retorno de la inversión (ROI):** Este KPI mide la rentabilidad de los esfuerzos de marketing digital. Se calcula dividiendo la ganancia neta de tus esfuerzos de marketing por el costo total.
- **Tasa de clics (CTR):** Este KPI mide el porcentaje de personas que hacen clic en los anuncios en comparación con el número total de impresiones.
- **Tasa de participación en las redes sociales:** Este KPI mide cómo interactúan los usuarios con tu contenido en las redes sociales, incluyendo me gusta, comentarios, compartidos, etc.

## Dashboards (Tableros de mando)

En los sistemas de información empresarial, un tablero es un tipo de interfaz gráfica de usuario que a menudo proporciona vistas de un vistazo de los indicadores clave de rendimiento (KPI) relevantes para un objetivo o proceso comercial en particular. Al proporcionar esta descripción general, los dueños de negocios pueden ahorrar tiempo y mejorar su toma de decisiones utilizando paneles.

El tablero a menudo es accesible mediante un navegador web y generalmente está vinculado a fuentes de datos que se actualizan periódicamente.

Los tableros bien conocidos incluyen los tableros de Google Analytics, utilizados en la mayoría de los sitios web, que muestran la actividad de los visitantes de una web; como cantidad de visitas, páginas de entrada, tasa de rebote y fuentes de tráfico.

La pandemia de COVID-19 de 2020 puso en primer plano otros paneles, como el [rastreador de coronavirus de Johns Hopkins](#) y el [rastreador de coronavirus del gobierno del Reino Unido](#) como buenos ejemplos.

El término dashboard tiene su origen en el tablero del automóvil, donde los conductores controlan las principales funciones de un vistazo a través del panel de instrumentos.

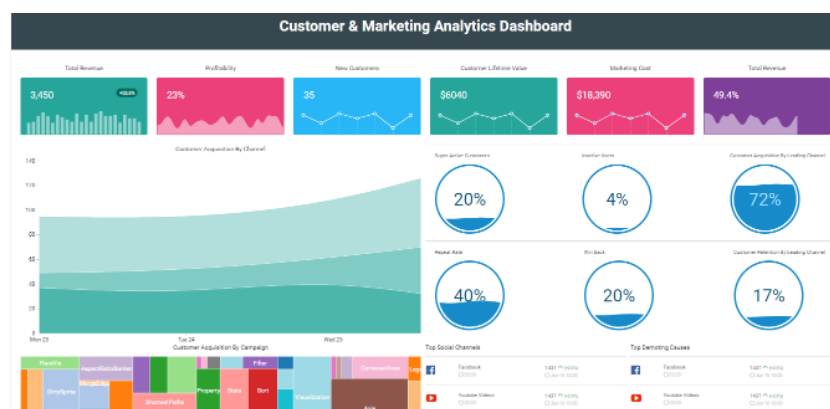
### Beneficios

Los tableros permiten a los gerentes y directores monitorear el rendimiento de los diversos departamentos en su organización, proporcionando una mirada integral del negocio.

Los beneficios de usar paneles digitales incluyen:

- Presentación visual de las medidas de rendimiento
- Capacidad para identificar y corregir tendencias negativas
- Medir eficiencias e ineficiencias
- Tomar decisiones objetivas basadas en datos

Los tableros ofrecen una visión holística de todo el negocio ya que proporcionan a la dirección de una visión panorámica del rendimiento de las ventas, el inventario de datos, el tráfico web, las analíticas de redes sociales y otros datos asociados que se presentan visualmente en un único panel. Las analíticas web juegan un papel crucial en la configuración de la estrategia de marketing de muchas empresas. Los paneles también facilitan un mejor seguimiento de las ventas y la presentación de informes financieros ya que los datos son más precisos y están en un solo área. Por último, los paneles ofrecen un mejor servicio al cliente a través del seguimiento porque mantienen tanto a los gerentes como a los clientes actualizados sobre el progreso del proyecto a través de correos electrónicos y notificaciones automatizadas.



Tablero de seguimiento de marketing

### Tableros interactivos de ejemplo

<https://www.heavy.ai/demos>

<https://www.tableau.com/learn/articles/business-intelligence-dashboards-examples>

### Herramientas de Business Intelligence (BI)

Las herramientas de Business Intelligence (BI), o Inteligencia de Negocios, son programas de software que se utilizan para analizar, procesar y visualizar datos relacionados con las actividades de una organización. En general, son las herramientas que se utilizan para construir dashboards en las empresas medianas y grandes, y pueden manejar grandes cantidades de datos desestructurados para ayudar a identificar, desarrollar y generar nuevos enfoques estratégicos de negocio.

Algunas de las características fundamentales de las herramientas de BI incluyen:

1. **Recopilación de Datos:** Las herramientas de BI pueden recolectar datos de diversas fuentes, como bases de datos, archivos CSV, hojas de cálculo de Excel, entre otros.
2. **Almacenamiento de Datos:** Algunas herramientas de BI tienen su propio sistema de almacenamiento de datos, mientras que otras pueden integrarse con sistemas de almacenamiento de datos existentes.
3. **Procesamiento de Datos:** Las herramientas de BI pueden procesar grandes cantidades de datos y convertirlos en información significativa. Esto a menudo implica funciones como la limpieza de datos, la integración de datos, la transformación de datos y la carga de datos.
4. **Análisis de Datos:** Las herramientas de BI pueden realizar una variedad de análisis, desde análisis básicos hasta complejos, como análisis de regresión, análisis de correlación, análisis de segmentación y mucho más.
5. **Visualización de Datos:** Las herramientas de BI pueden representar datos de manera gráfica, facilitando la comprensión y el análisis de los datos. Esto puede incluir gráficos, mapas de calor, gráficos de barras, gráficos de líneas, etc.



6. **Reporte y Dashboard:** Las herramientas de BI pueden generar informes detallados y tableros de control interactivos para presentar los resultados del análisis de datos de una manera fácil de entender.

Entre las herramientas de BI más populares se encuentran Microsoft Power BI, Tableau, QlikView, Looker, y SAS BI. Estas herramientas permiten a las organizaciones tomar decisiones informadas al proporcionar información útil sobre las tendencias, los problemas y los resultados que están ocurriendo en la organización.

El Cuadrante Mágico de Gartner es un informe de investigación y análisis que proporciona una visión profunda del mercado, la dirección, la madurez y los participantes de una industria específica, en este caso, la industria de la Inteligencia de Negocios (Business Intelligence, BI).

En el Cuadrante Mágico, Gartner evalúa a los proveedores de BI en base a su integridad de visión y su capacidad para ejecutar. Con base en estas evaluaciones, los proveedores son clasificados en uno de los cuatro cuadrantes:

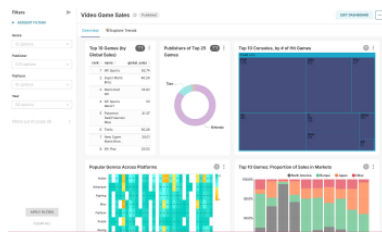
1. **Líderes:** Los líderes ejecutan bien contra su visión actual y están bien posicionados para el futuro. Tienen un sólido historial, una presencia visible en el mercado y una visión clara de la dirección del mercado.
2. **Visionarios:** Los visionarios tienen una visión clara de la dirección del mercado y están desarrollando capacidades para mantenerse al día con las tendencias, pero no pueden ejecutar su visión tan eficazmente como los líderes.
3. **Jugadores de Nicho:** Los jugadores de nicho se centran con éxito en un segmento pequeño del mercado, o están enfocados en una funcionalidad específica, pero no muestran la capacidad de innovar o de influir en la dirección general del mercado.
4. **Retadores:** Los retadores tienen la capacidad de ejecutar bien hoy en día, pero no tienen una visión tan desarrollada como los líderes. A menudo tienen una gran presencia en el mercado, pero carecen de la profundidad de funcionalidad o capacidad para mantenerse al día con las tendencias del mercado.

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms



Source: Gartner March (2023)

Existen varias soluciones de Business Intelligence (BI) de código abierto que pueden ser una excelente opción para las empresas que buscan flexibilidad y control sobre su software. Por ejemplo [Apache Superset](#), [Redash](#), [Metabase](#), [Grafana](#), [Kibana](#), entre otras.



Apache Superset



Redash



Metabase



Grafana



Kibana

## Buenas y malas prácticas de diseño y visualización

El diseño y la visualización son componentes esenciales para la interacción del usuario. El modo en que se presenta la información, el esquema de colores que se utiliza, la elección de la tipografía, la colocación de los botones y muchos otros factores influyen en cómo los usuarios perciben y utilizan una interfaz. Un buen diseño puede facilitar la comprensión, mejorar la usabilidad y aumentar la satisfacción del usuario. Sin embargo, un mal diseño puede frustrar a los usuarios, crear barreras para la accesibilidad y disminuir la eficacia general del producto.

### Uso del color

El color es una de las cosas más importantes para entender en la visualización de datos y con frecuencia se utiliza incorrectamente. No debe usar el color solo para darle vida a una visualización aburrida. De hecho, muchas visualizaciones de datos excelentes no usan color en absoluto y son informativas y hermosas.

En la Figura 1.15, vemos la visualización de Shine Pulikathara que ganó la competencia Tableau Iron Viz de 2015. Notar el simple uso del color.

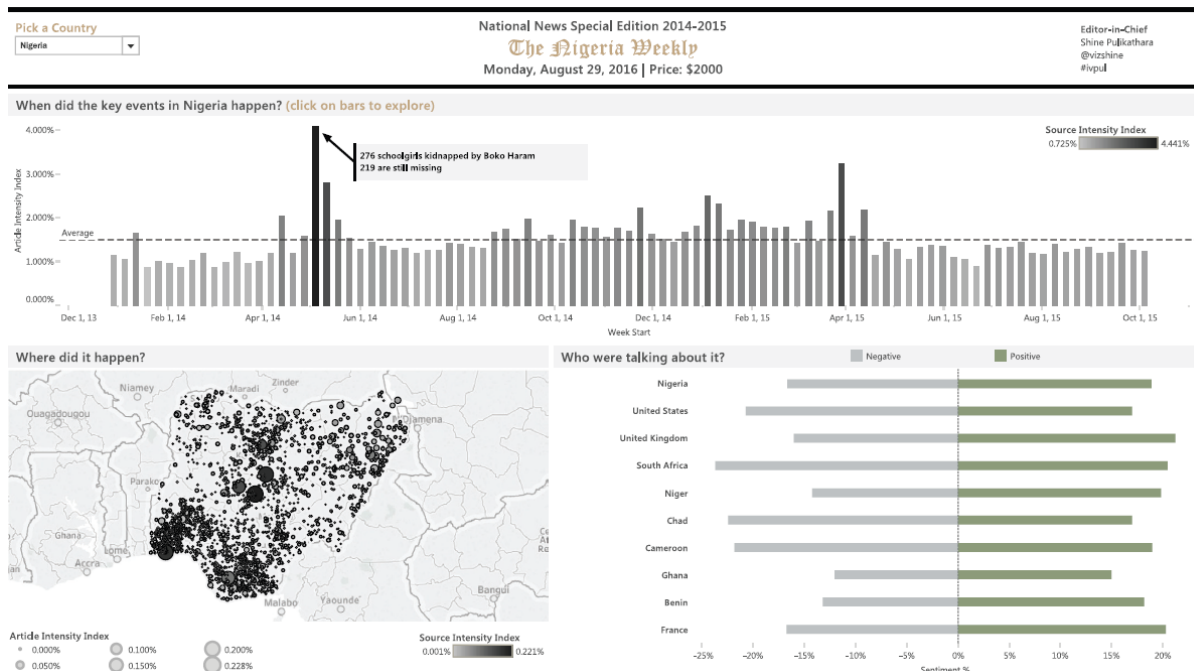


Figure 1.15 Winning visualization by Shine Pulikathara during the 2015 Tableau Iron Viz competition. Source: Used with permission from Shine Pulikathara.

El color debe usarse con un propósito. Por ejemplo, el color se puede utilizar para llamar la atención del lector, resaltar una parte de los datos o distinguir entre diferentes categorías. El color debe usarse en la visualización de datos de tres formas principales: secuencial, divergente y categórico. Además, a menudo existe la necesidad de resaltar datos o alertar al lector de algo importante. La Figura 1.16 ofrece un ejemplo de cada uno de estos esquemas de color.

## USE OF COLOR IN DATA VISUALIZATION

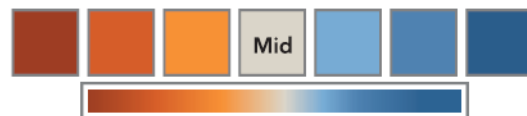
### SEQUENTIAL

color is ordered from low to high



### DIVERGING

two sequential colors with a neutral midpoint



### CATEGORICAL

contrasting colors for individual comparison



### HIGHLIGHT

color used to highlight something



### ALERT

color used to alert or warn reader



FIGURE 1.16 Use of color in data visualization.

## Contraste

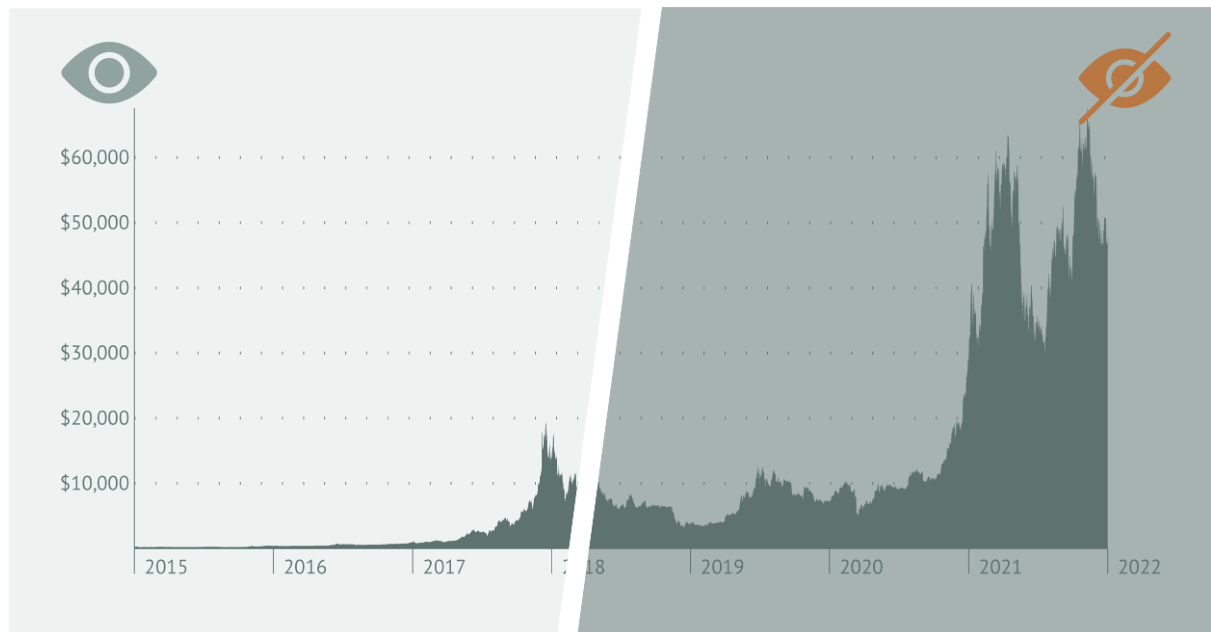
La pauta de contraste de color más utilizada es la WCAG. Hay dos niveles de contraste de color más utilizados: AA (contraste mínimo) y AAA (contraste mejorado):

El nivel AA requiere una relación de contraste de al menos 4.5:1 para el texto normal y 3:1 para el texto grande. WCAG 2.1 requiere una relación de contraste de al menos 3:1 para gráficos y componentes de la interfaz de usuario.

El nivel AAA requiere una relación de contraste de al menos 7:1 para el texto normal y 4.5:1 para el texto grande.

Hay muchos comprobadores de contraste de color que pueden ayudarte a seguir las pautas:

[Colorable](#), [WebAIM](#), [Coolers](#), [ColorContrast](#), y muchos más. Pero recuerda hacer una revisión manual y probarlo con los usuarios, porque las pautas a veces pueden ser defectuosas.



## Tamaño del texto y tipos de letra

No hay directrices oficiales sobre el tamaño mínimo del texto, pero generalmente se recomienda utilizar fuentes de tamaño 12pt (16px) para el texto del cuerpo al utilizar tamaños de monitor tradicionales. El tamaño mínimo de la fuente es de 9pt (12px). Por debajo de este tamaño, la fuente puede volverse técnicamente legible para algunas plataformas. Según las pautas WCAG, debería haber una opción para redimensionar el texto sin pérdida de contenido o funcionalidad.

Los tipos de letra que se utilizan en la visualización también pueden tener un gran impacto en la accesibilidad. Puede hacer una diferencia para las personas con mala visión, discapacidades de aprendizaje, afasia, dislexia o bajo alfabetismo adulto. Gareth Ford Williams seleccionó 8 puntos que deberías considerar al elegir un tipo de letra.

1. Usa una fuente apropiada para tu público. Si se trata de niños o personas con baja visión, es mejor elegir una fuente sans serif y si tu público lee rápido, las fuentes serif serán una mejor elección.
2. Evita usar formas similares para diferentes símbolos. La "i" mayúscula, la "L" minúscula y el "1" deben verse diferentes.
3. Evita las letras espejo. Por ejemplo, la "p" y la "q" deben contener diferencias además del espejo.
4. Las letras deben ser fácilmente distinguibles. Un ejemplo es la "o" y la "c". En este caso, la "c" necesita estar notablemente abierta para que sea difícil confundirla con la "o" para las personas con baja visión.
5. Usa tipos de letra Humanistas en lugar de Grotescas, ya que tienden a tener más variación en las formas de las letras y son más legibles en tamaños más pequeños.
6. Usar fuentes con un espaciado más definido entre las letras, ya que el espaciado compacto tiende a parecer pegado, lo que tiene una mala influencia en la legibilidad.
7. La altura de las letras mayúsculas y los ascendentes deben variar. Como ejemplo, la "i" mayúscula y la "l" minúscula deben tener una altura diferente.
8. Probar el tipo de letra en el contexto en el que lo vayas a utilizar.



## Generar información accesible para todo el público

No todo el mundo procesa la información de la misma manera y con ciertas prácticas de visualización podemos llegar a dejar a grupos de personas “afuera” del análisis. Por ejemplo, sabemos que las personas con Daltonismo pueden tener problemas detectando algunos colores o que las personas con Dislexia pueden tener más dificultad para leer ciertas tipografías y tamaños de letras.

A continuación les dejamos unos links con recomendaciones para conseguir visualizaciones más inclusivas:

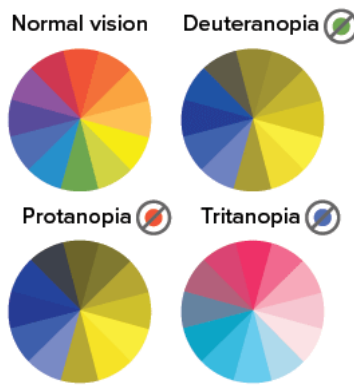
- [Detección de colores](#)
- [Dislexia](#)
- [Consejos para diseñar gráficos accesibles para lectores con discapacidad visual](#)
- [Do No Harm Guide: Underrepresented Groups](#) (Guía de Acción Sin Daño: Grupos subrepresentados)

### Deficiencia de la visión del color (daltonismo)

Según la investigación (Birch 1993), aproximadamente el 8 por ciento de los hombres tienen deficiencia de visión cromática (CVD) en comparación con solo el 0,4 por ciento de las mujeres. Esta deficiencia es causada por la falta de uno de los tres tipos de conos dentro del ojo necesarios para ver todos los colores. La deficiencia comúnmente se conoce como “daltonismo”, pero ese término no es del todo exacto.

De hecho, las personas que padecen CVD pueden ver el color, pero no pueden distinguir los colores de la misma manera que el resto de la población. El término más preciso es “deficiencia de la visión del color”. Dependiendo del cono que falte, puede ser muy difícil para las personas con CVD distinguir entre ciertos colores debido a la forma en que ven el espectro de colores. Hay tres tipos de deficiencia de visión cromática:

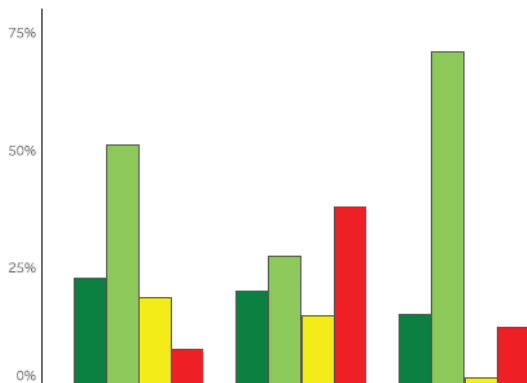
1. Protanopia es la falta de conos de onda larga (rojo débil).
2. La deuteranopia es la falta de conos de onda media (verde débil).
3. Tritanopia es la falta de conos de onda corta (azul).  
(Esto es muy raro y afecta a menos del 0,5 por ciento de la población).



La CVD es principalmente hereditaria y, como puede ver en los números, afecta principalmente a los hombres. En empresas más grandes o cuando se presenta una visualización de datos al público en general, los diseñadores deben comprender el CVD y diseñar con eso en mente.

El principal problema entre las personas con CVD es con los colores rojo y verde. Por eso es mejor evitar el uso conjunto de rojo y verde y, en general, evitar los colores de semáforo de uso común.

Traffic Light Colors



Protanopia Simulation

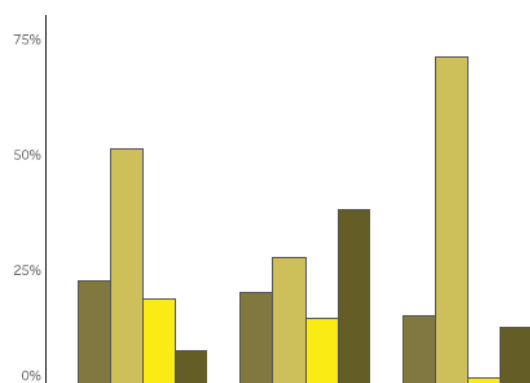
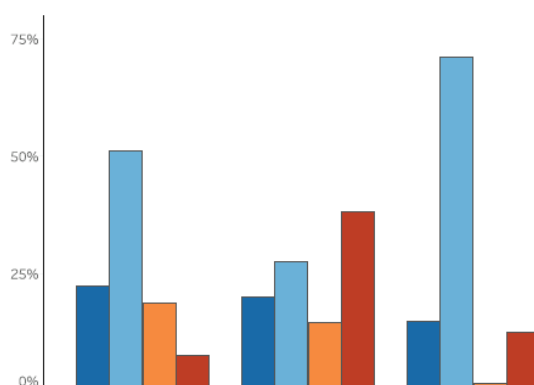


Figura 1.24 Gráfico de barras usando los colores del semáforo y una simulación de protanopia. Observe que las barras rojas y verdes en el panel de la derecha son muy difíciles de diferenciar entre sí para una persona con protanopia.

Color-blind-Friendly Blue and Orange



Protanopia Simulation

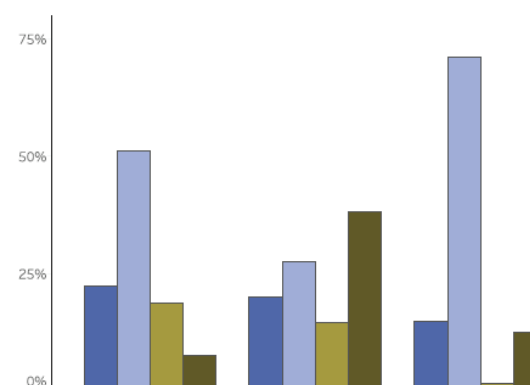


Figura 1.25 Gráfico de barras que usa una paleta azul y naranja para daltónicos y una simulación de protanopia.

### Web Content Accessibility Guidelines

Las Web Content Accessibility Guidelines (WCAG) son un conjunto de recomendaciones para hacer el contenido web más accesible, principalmente para las personas con discapacidades. Estas pautas son publicadas por el World Wide Web

Consortium (W3C), una organización internacional que establece estándares para la web. Las WCAG se desarrollan a través del proceso W3C en cooperación con individuos y organizaciones de todo el mundo, y tienen el objetivo de proporcionar un estándar compartido para la accesibilidad del contenido web que satisfaga las necesidades de individuos, organizaciones y gobiernos a nivel internacional.

Las WCAG se aplican a una amplia gama de tecnologías web y se pueden aplicar a cualquier información o servicio que utilice tecnologías web. Cubren cosas como texto, imágenes, sonido, código (como HTML y CSS), etc.

Las pautas WCAG sirven para varios propósitos:

1. Ayudar a hacer la web más inclusiva: Las WCAG permiten que las personas con discapacidades, incluyendo discapacidades visuales, auditivas, físicas, del habla, cognitivas, del lenguaje, de aprendizaje y neurológicas, puedan usar y contribuir en la web.
2. Proporcionar una referencia estándar: Las WCAG son utilizadas por desarrolladores, diseñadores, administradores de sitios web y otros profesionales para asegurarse de que sus sitios web y aplicaciones sean accesibles.
3. Cumplimiento legal: En muchos países, tener un sitio web accesible es un requisito legal, y las WCAG a menudo se utilizan como base para estas leyes.

## **Bibliografía ampliatoria**

- Data Visualization: A practical introduction (Kieran Healy)
- Do No Harm Guide Centering Accessibility in Data Visualization (Jonathan Schwabish, Sue Popkin, Alice Feng)
- Fundamentals of Data Visualization (Claus O. Wilke)