

Data Science Basics

Visualización de datos



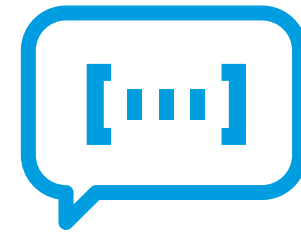
Visualización de datos en Data Science



Source: opendatascience.com

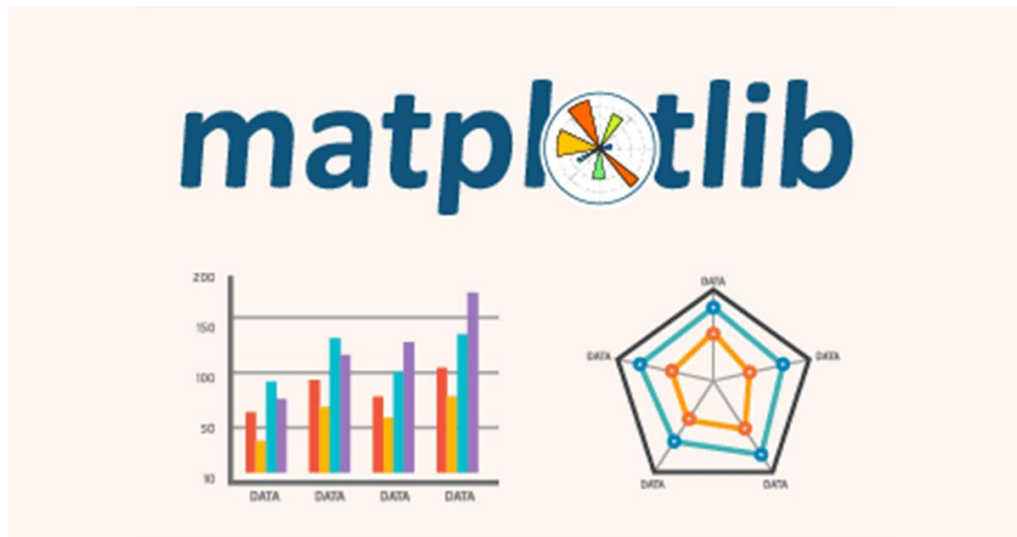
- Convertir datos crudos en **representaciones gráficas** como mapas, gráficas, grafos y dashboards, hace un dataset complejo, **más accesible, entendible y utilizable**.
 - La **Neurociencia** ha demostrado que el cerebro humano le cuesta mucho más entender los datos complejos cuando se codifican en números y texto en comparación a utilizando representaciones visuales y gráficas.
 - Las técnicas de **Dataviz** aprovechan esa calidad, ayudando a presentar cantidades masivas de información en la forma más eficiente y visual.
- En particular, es muy útil en la **etapa de comprensión de los datos**, generar distintas visualizaciones utilizando tablas y gráficos de distinto tipo (puntos, barras, torta), seleccionando atributos y graficándolos para analizar su distribución y relación con otros atributos del dataset.

01



Librerías disponibles

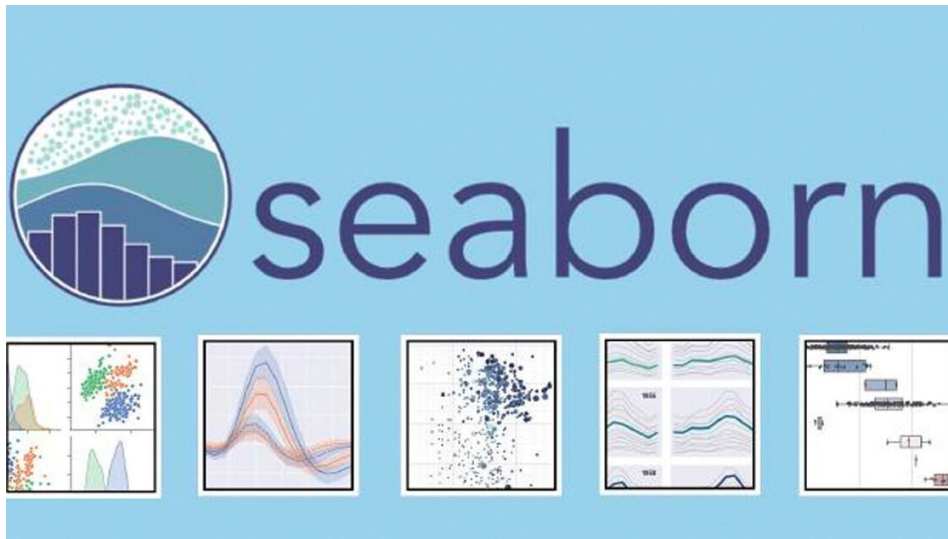
Matplotlib



<https://matplotlib.org/>

- Ofrece varias opciones de gráficos 2D, como gráficos de líneas, gráficos de dispersión, gráficos de barras, histogramas y más.
- Es fácil de usar y ofrece amplias opciones de personalización, lo que facilita la creación de gráficos y figuras de alta calidad adecuados tanto para medios impresos como digitales.
- Adecuado para crear visualizaciones estáticas, que son útiles para presentar datos en una publicación o informe.
- Es altamente personalizable, lo que permite a los usuarios crear gráficos y figuras visualmente atractivos.

Seaborn



<https://seaborn.pydata.org/>

- Es una potente librería de visualización de datos de Python que se basa en matplotlib.
- Proporciona una interfaz de alto nivel para crear gráficos estadísticos informativos y atractivos.
- Es particularmente útil para explorar y visualizar grandes conjuntos de datos, ya que incluye una variedad de funciones de gráficos avanzadas que se pueden personalizar fácilmente.

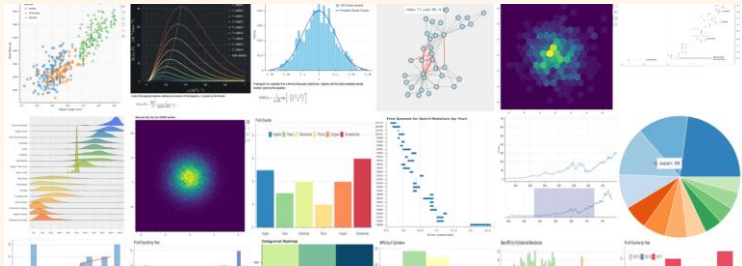
Plotly



<https://plotly.com/>

- Es una librería de código abierto muy popular que se utiliza para crear **visualizaciones de datos interactivas**.
- Plotly ofrece dos API principales para crear gráficos: la API Graph Objects de bajo nivel y la API Figure Factory de alto nivel.
 - **La API Graph Objects** se utiliza para crear objetos de seguimiento individuales, que se pueden combinar para crear un gráfico.
 - **La API Figure Factory**, ofrece funciones predefinidas para crear tipos específicos de gráficos.

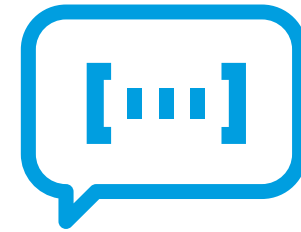
Bokeh



- Es una librería muy popular que se utiliza para crear **visualizaciones interactivas para navegadores web**.
- Es una librería de código abierto y ofrece una variedad de herramientas y funciones para crear visualizaciones de datos interactivas.

<https://bokeh.org/>

02



¿Qué gráficos usar?

Usando el tipo correcto de visualización



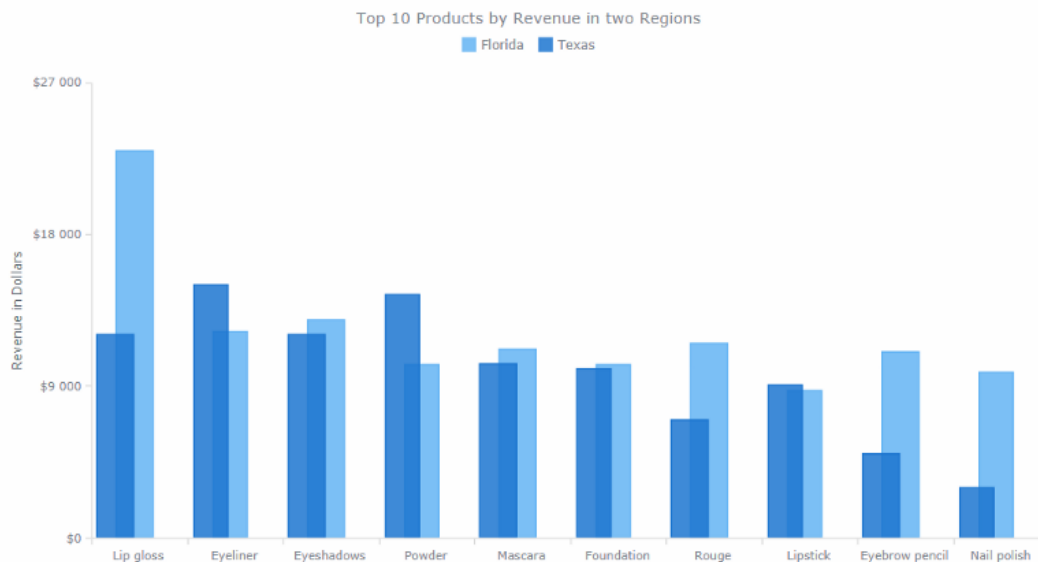
Source: datos.gob.es

- A lo largo del tiempo, han surgido **variedad** de técnicas de visualización, herramientas y tipos de gráficos.
- Por esta razón, dependiendo de los **datos** que se quieren explorar, su **naturaleza** y las **hipótesis** que deseamos validar, la técnica o herramienta que usaremos será distinta.
 - Ejemplo, si el tipo de gráfico elegido no es el correcto, podemos omitir un problema de sesgo en un conjunto de datos o podríamos no transmitir correctamente las conclusiones de un análisis de datos al cliente final.
- Una forma muy fácil de separar las distintas técnicas de visualización de datos disponibles es de acuerdo a la **tarea que tenemos como objetivo**.

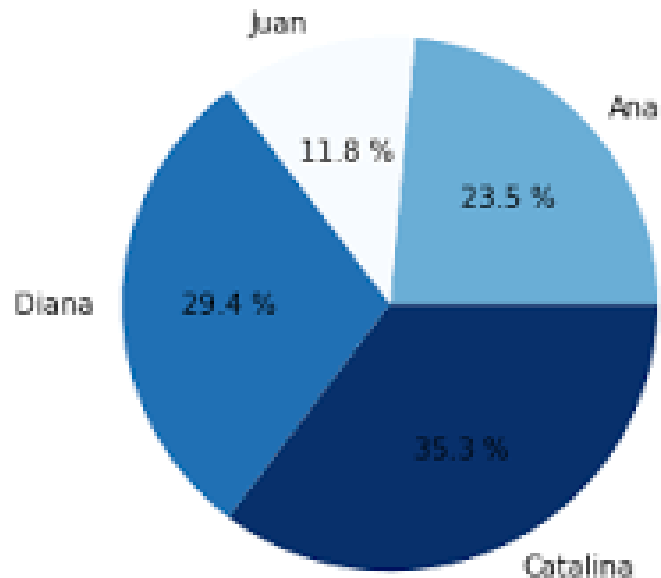
Comparar Datos

Para comparar datos, por ejemplo, de diferentes variables de un mismo dataset, o datos de un dataset contra otro, los siguientes tipos de gráficos son muy útiles:

- **Barras:** Permite comparar de forma directa, valores cuantitativos por variable o categoría
- **Pilas apiladas:** Permiten analizar valores cuantitativos por categoría y a la vez compararlos con diferentes muestras. Por ejemplo, analizar el PBI de diferentes países y teniendo en cuenta el aporte porcentual de cada sector.
- **Gráfico Radar:** Es otra alternativa para comparar los valores cuantitativos de diferentes variables. Por ejemplo, mostrar el puntaje en diferentes habilidades que hace a la destreza de un jugador de fútbol, como Conversión de Goles, Precisión en pases, etc.



Explorar relaciones Part-To-Whole (de la parte al todo)

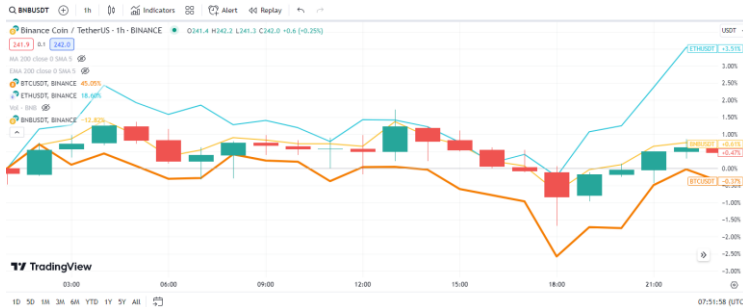


Cuando lo que se quiere es explorar la composición por ejemplo de una **variable categórica** y analizar relaciones del tipo part-to-whole, los siguientes tipos de visualizaciones son muy útiles:

- **Gráfico de tarta (Pie chart/Donut chart):** Permite analizar de forma básica de la composición porcentual de una variable categórica.
- **Gráfico Piramidal (Pyramid chart):** Cuando los datos representan o responden a una "jerarquía", puede ser interesante utilizar este tipo de gráficos.
- **Mapa de Árbol (Treemap chart):** Permite representar relaciones jerárquicas complejas. Por ejemplo, la cantidad de ingresos anuales de una compañía agregados por unidades de negocio y para cada unidad de negocio por proyecto.
- **Gráfico de Embudo (Funnel chart):** Este tipo de gráfico se utiliza mucho para comparar variables cuantitativas, que de alguna forma están relacionadas entre sí como parte de un proceso. Por ejemplo, se puede utilizar para analizar ventas de un sitio de eCommerce, ayudando a identificar los cuellos de botella del proceso.

Análisis en el tiempo

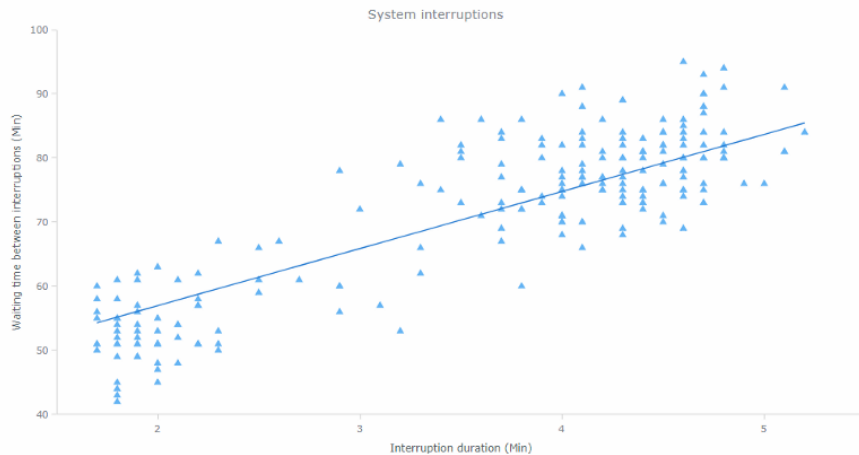
Cuando se busca analizar la relación entre una o más variables con una componente temporal, como por ejemplo puede ser analizar la evolución de salarios, respecto del tiempo, los siguientes tipos de visualizaciones son los indicados:



- **Gráfico de línea/curva (Line/Spline chart):** Es la representación gráfica más simple y utilizada para mostrar la evolución de una variable cuantitativa en función de otra variable.
- **Gráfico de área (Área chart):** A diferencia de la anterior, utiliza áreas para representar una variable cuantitativa.
- **Gráfico vela Japonesa (Candlestick chart):** Es muy útil cuando se quiere analizar una variable que puede tomar múltiples valores para un mismo momento, por ejemplo, analizar la evolución del precio de las acciones de una compañía frente a otras, tomando como referencia el precio de apertura y cierre de las acciones en la bolsa por día. Es una variante de las gráficas OHLC.
- **Gráfico Sparkline (Sparkline chart):** Consiste en construir versiones simplificadas y pequeñas de otros gráficos temporales como Line Charts o Área Charts, eliminando por ejemplo los ejes del gráfico. Son muy útiles para utilizar en tablas, en donde para cada fila queremos hacer más visual una variable determinada. Por ejemplo, listar los montos totales de exportaciones de granos en Uruguay en un año entero, utilizando Sparklines para mostrar para cada grano la variación mensual del precio de este.

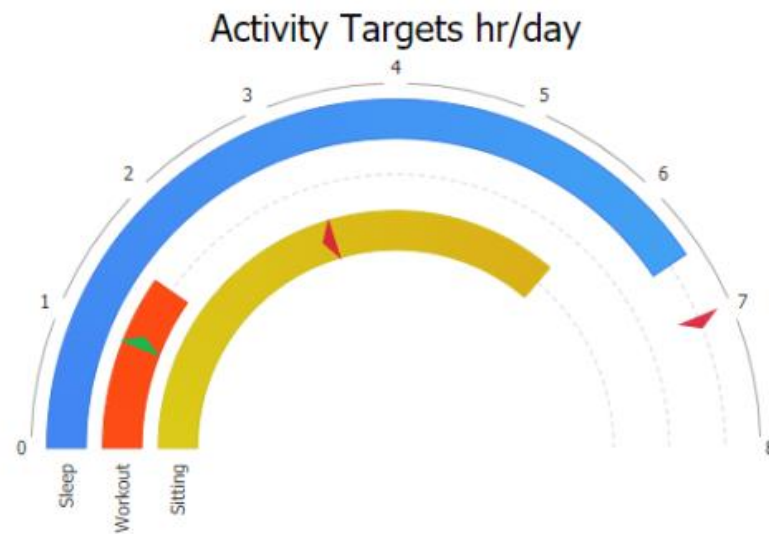
Analizar distribución de un dataset

Para analizar la distribución de una variable en un determinado dataset y comparar por ejemplo con otros datasets, encontrar relaciones y correlaciones, los siguientes tipos de gráficos son muy útiles:



- **Histogramas:** Un histograma es una representación de la estimación de la distribución de valores de un dataset numérico.
- **Gráfico de dispersión de puntos (Dot (scatter) chart):** Permite examinar las tendencias en distribución y correlación entre dos variables.
- **Gráfico de burbujas (Bubble charts):** Cada punto se grafica como un disco, expresando dos de los atributos/dimensiones del dato a través de la ubicación del disco en el sistema de ejes cartesianos y el tercero a través de su tamaño.
- **Gráfico Box-and-whisker**
- **Gráfico de errores**
- **Mapa de calor (Heatmap chart):** Es una representación visual bidimensional de los datos, donde los valores se codifican en colores.
- **Gráfico de rangos (Range chart)**
- **Gráfico Polar (Polar chart)**

Evaluar performance de un dato

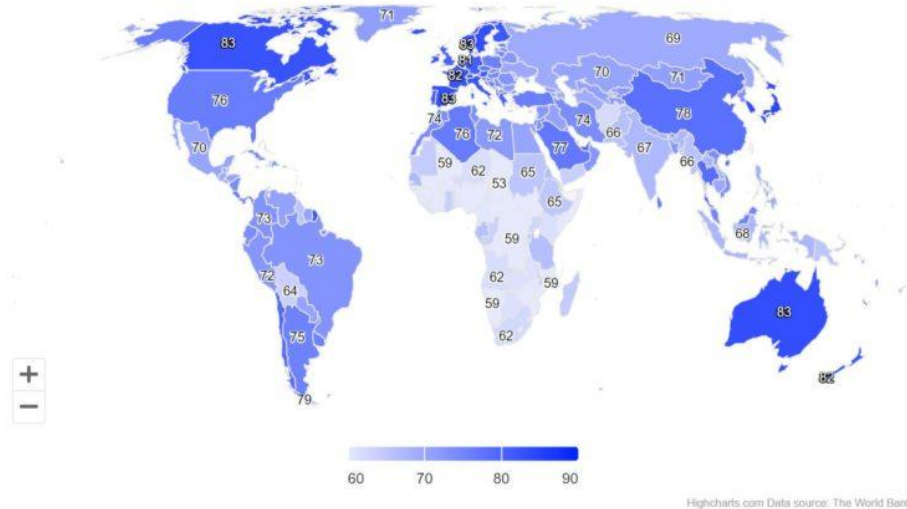


Cuando se quiere evaluar el rendimiento de una variable y comparar el valor real con valores KPI se suele recurrir a los siguientes tipos de gráficos:

- **Gauge Circular gauge-bullet**
- **Gauge Lineal**
- **Bullet Graph gauge-bullet**

Datos geográficos

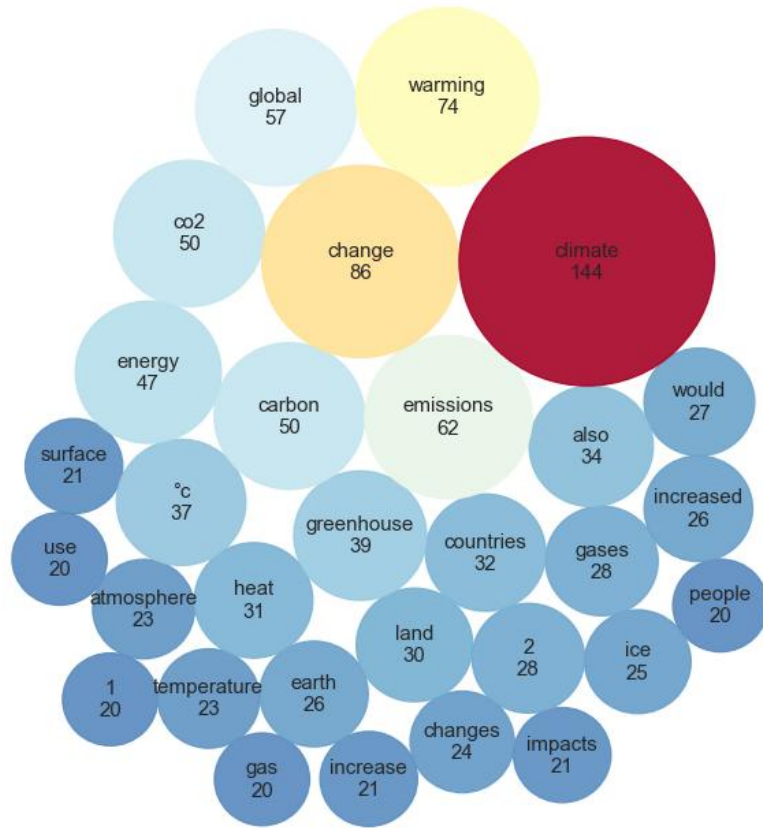
Life expectancy by country (2021)



Muchas veces, la mejor forma de darle sentido a datos geográficos (geo-referenciados) es utilizando visualizaciones en un mapa.

- **Mapa de Colores (Choropleth maps):** Permiten identificar diferencias entre áreas geográficas.
- **Mapa de Puntos (Dot maps):** Proporcionan una visualización accesible de la distribución geográfica de los fenómenos en las áreas geográficas en la que se encuentran.
- **Mapa de Burbujas (Bubble maps):** Similar a los gráficos de burbujas, permiten agregar una tercera dimensión a la posición de un fenómeno, utilizando el tamaño de una burbuja. Son muy útiles para agregar la información de un mapa de puntos cuando contamos con tantos puntos que se hace difícil entender la visualización.
- **Mapa de Conexiones (Connector map):** Permite representar relaciones entre los distintos puntos del mapa mediante líneas que unen los puntos.
- **Mapa de Flujos (Flow map):** Útiles para explorar cómo los objetos se mueven entre ubicaciones, en especial cuando interesa mostrar la dirección del flujo.

Análisis de Textos

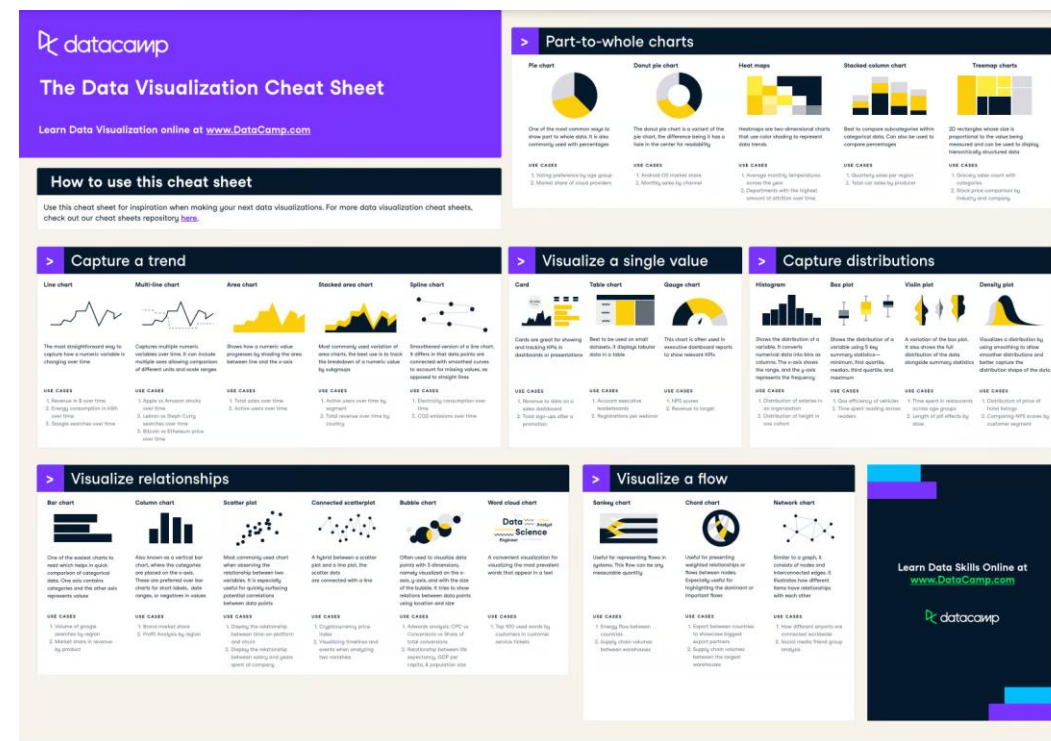


De la mano de la explosión de las redes sociales, ha explotado todo lo relacionado al análisis de datos en este tipo de redes, apareciendo nuevas áreas enfocadas específicamente en el análisis de este tipo de datos.

- Por ejemplo, Social Listening o Social Media Measurement y Social Business Intelligence (SBI), entre otros.

Uno de los requerimientos más simples en este tipo de análisis es cuantificar de que se está hablando en las redes y lograr comunicar esto con visualizaciones intuitivas, por ejemplo, utilizando **nubes de palabras**.

Cheatsheets



<https://datos.gob.es/es/blog/como-elegir-el-grafico-correcto-para-visualizar-datos-abiertos>

<https://www.datacamp.com/cheat-sheet/data-viz-cheat-sheet>



Next steps



We would like to know your opinion!

Please, let us know what you think about the content.
From Netmind we want to say thank you, we appreciate time
and effort you have taking in answering all of that is
important in order to improve our training plans so that you
will always be satisfied with having chosen us
quality@netmind.es

Thanks!

Follow us:

