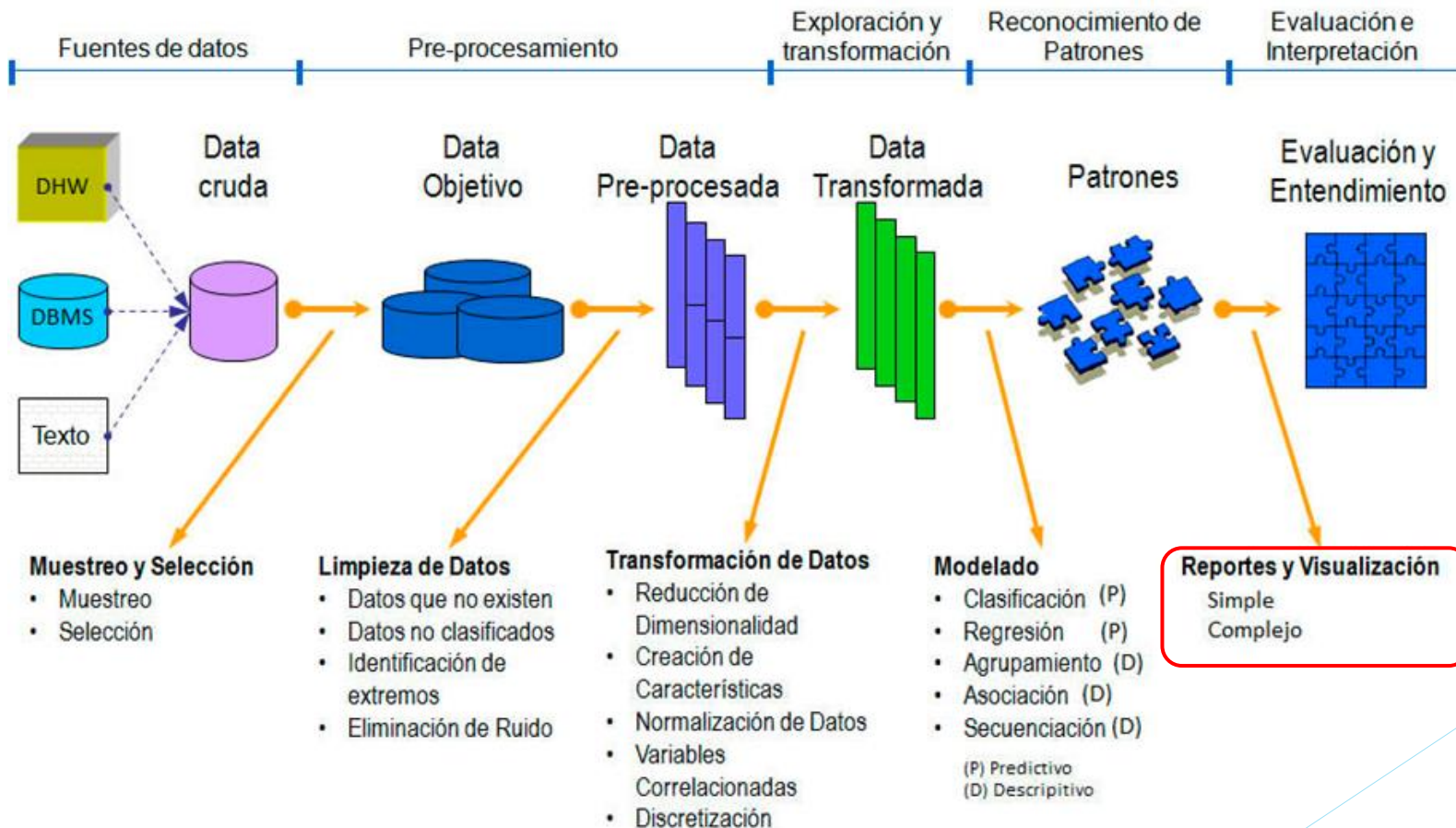


Introduccción al BigData

Ing. Ricardo Velasteguí (M.B.A.)

Visualización de Datos

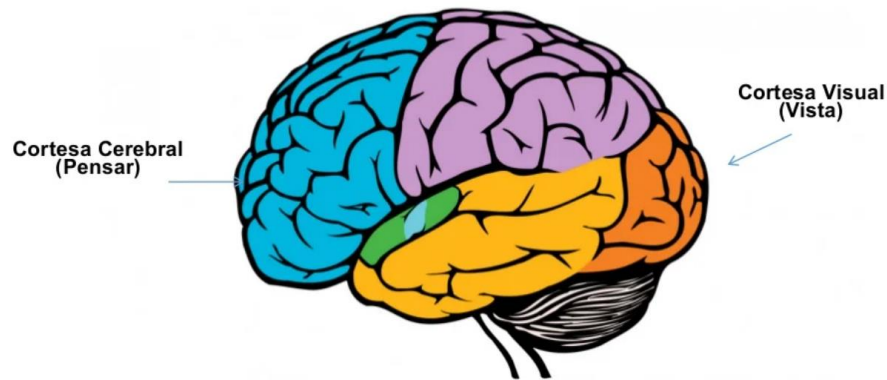
- Data visualization es la representación de datos a través de atributos visuales fáciles de entender, como gráficas, diagramas, mapas, piezas interactivas y otros.
- Permite representar cualquier tipo de información de una forma visual y sencilla, permite difundir el análisis previo de manera precisa y consistente, para ser visualizada, al igual que proporciona la opción de comunicar el significado de los datos de una manera más entendible.



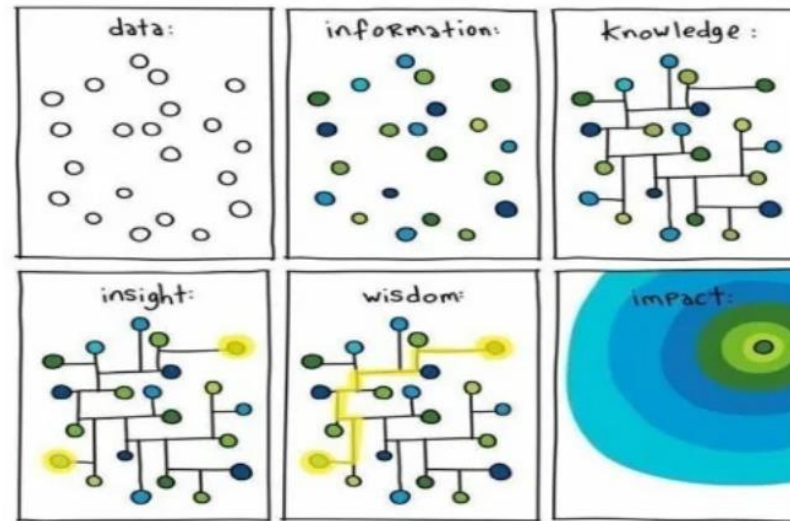
Visualización de Datos

- ▶ Debido a la forma en que el cerebro humano procesa la información, utilizar gráficos para visualizar grandes cantidades de datos complejos es más fácil que examinar las hojas de cálculo o los informes.
- ▶ La visualización de datos es una manera rápida y fácil de transmitir conceptos de una manera universal y puede experimentar con diferentes escenarios haciendo pequeños ajustes.

Pensar es lento vs La vista es más rápida



El cerebro ama las diferencias....

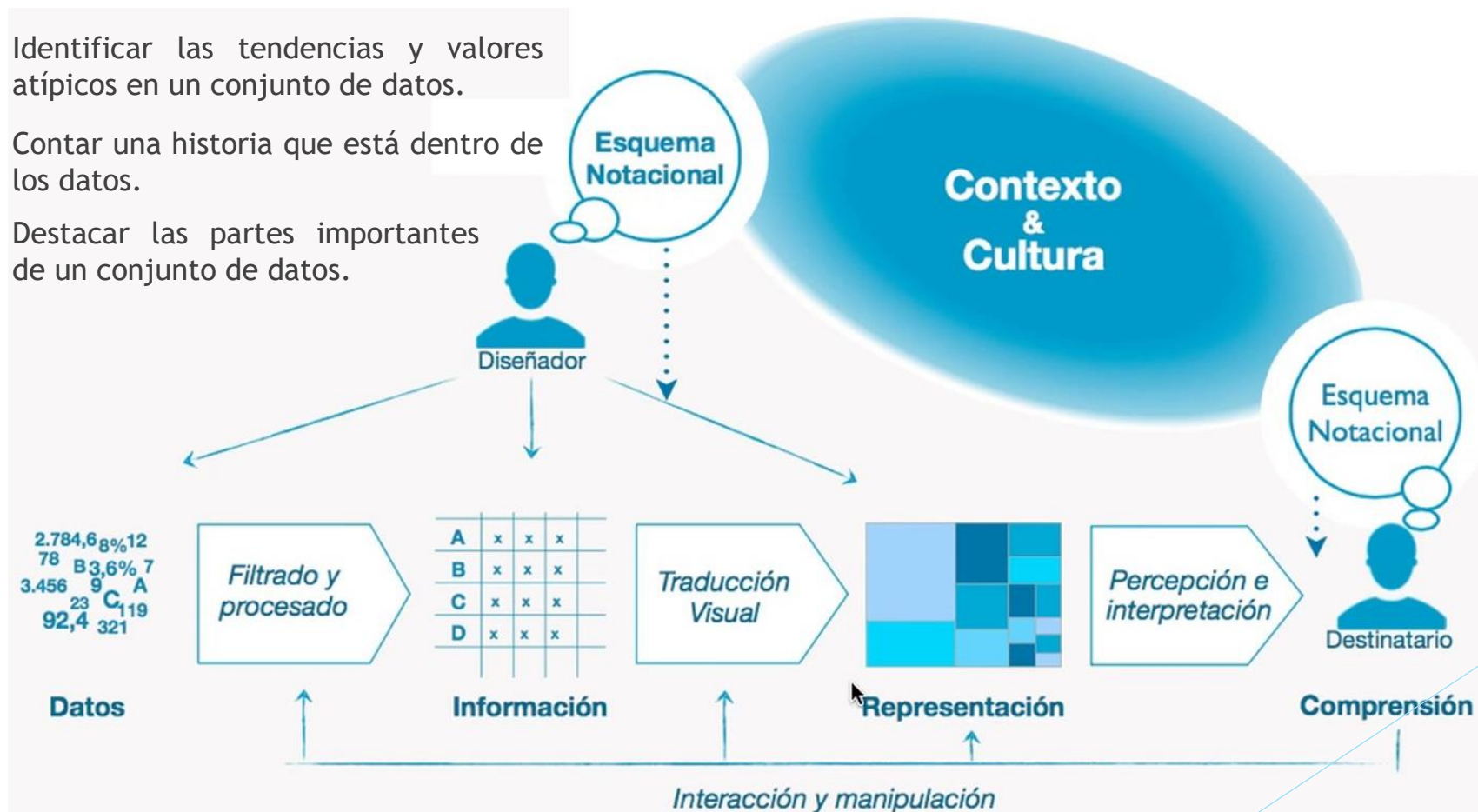


- ▶ Cuando vemos imágenes y formas nuestra mente se estimula y realiza un proceso cognitivo para darles un significado, por lo tanto, el proceso de visualización de datos es muy natural.
- ▶ El cerebro humano es capaz de procesar una imagen completa en **13 milésimas de segundo**.
- ▶ A su vez es capaz de procesar una imagen **60,000.00 veces más rápido** que un texto.
- ▶ El cerebro humano sólo puede **retener y almacenar entre 5 y 8 números**, elementos...

Visualización de Datos

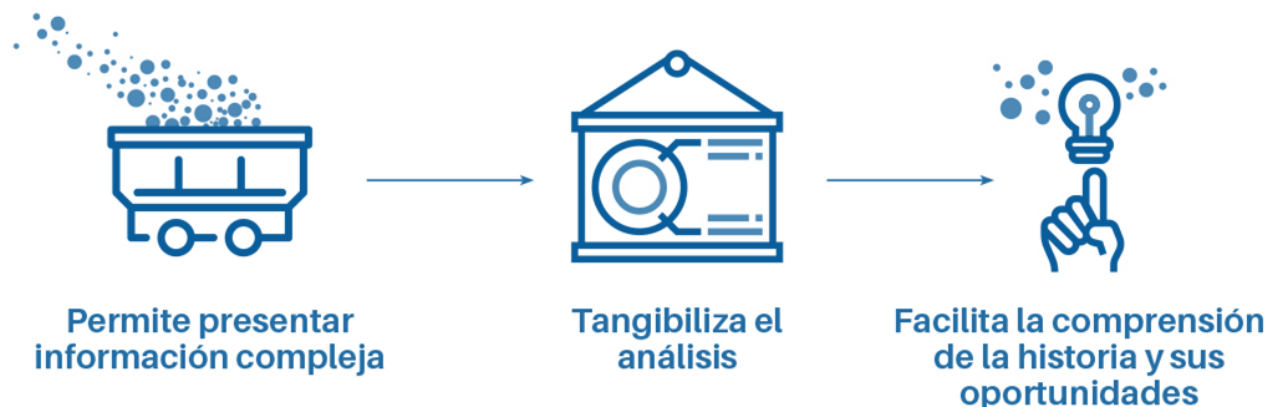
► Con la visualización de datos también se puede:

- Hacer que los datos luzcan llamativos y sean fáciles de entender.
 - Identificar las tendencias y valores atípicos en un conjunto de datos.
 - Contar una historia que está dentro de los datos.
 - Destacar las partes importantes de un conjunto de datos.
- Reforzar una opinión o un argumento.



Visualización de Datos

- ▶ Algunos beneficios de la Visualización de Datos son:



- ▶ Se **reduce el tiempo** a la hora de tomar decisiones y por tanto obtener conclusiones porque los gráficos son más fáciles de interpretar que millones de datos en celdas.
- ▶ **Mejora en la toma de la decisión** acertada, porque, ya que durante el proceso se ha garantizado el dato fiable, la información mostrada es veraz y correcta.
- ▶ **Identificar relaciones entre variables y datos atípicos**, que se han obtenido tras la ejecución de los algoritmos de tratamiento de datos.
- ▶ **Detectar tendencias en el mercado** para anticiparse a las futuras necesidades de los consumidores.
- ▶ **Conocimiento en tiempo real** de la situación de la compañía, ya que las herramientas de visualización se actualizan constantemente.
- ▶ **Facilita la comunicación y la explicación de los datos** al resto de empleados o dirigentes de la empresa.

Visualización de Datos

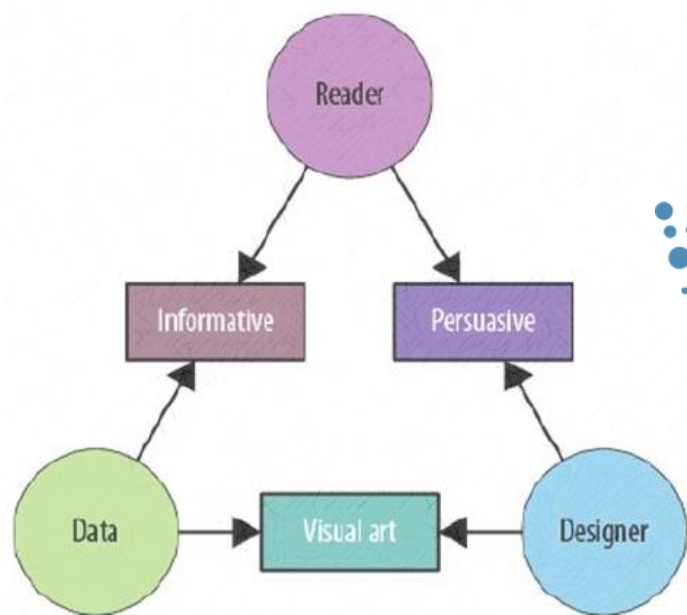
- ▶ Antes de implementar una nueva tecnología de visualización de datos, hay algunos pasos que usted necesita tomar. No sólo necesita tener una sólida comprensión de sus datos, sino que también necesita entender sus metas, necesidades y audiencia.
- ▶ Los pasos son:
 - ▶ Comprender los datos que está intentando visualizar, incluyendo su tamaño y cardinalidad.
 - ▶ Determine qué está intentando visualizar y qué tipo de información desea comunicar.
 - ▶ Conozca a su audiencia y comprenda cómo procesa la información visual.
 - ▶ Utilice un visual que transmita la información en la mejor y más simple forma para su audiencia.



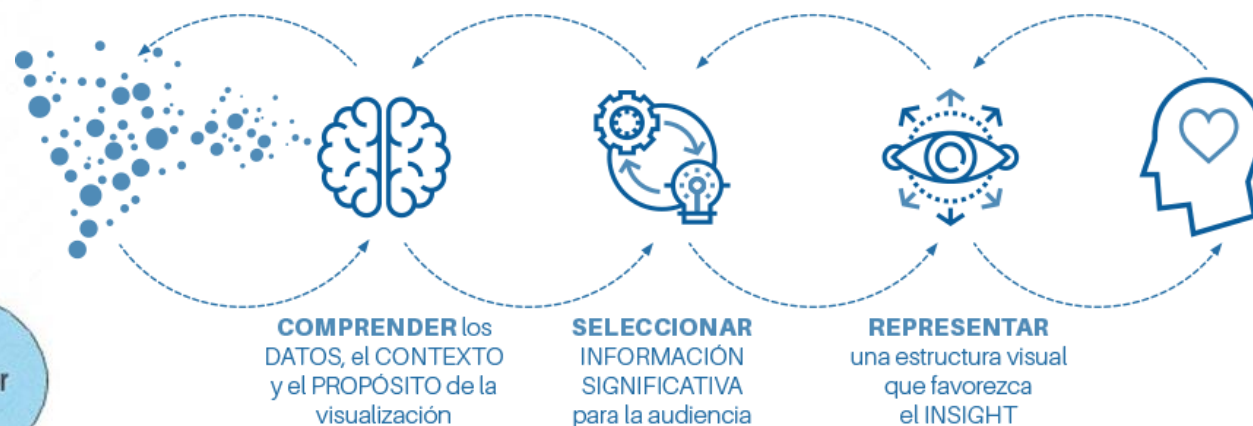
- ▶ Para **poder medir y comparar** el rendimiento de la percepción humana sobre diversos fenómenos se necesita una **métrica**, es decir, una medida para evaluar de forma fiable.

Visualización de Datos

- ▶ Según Iliinsky y Steele, es útil pensar en una eficaz visualización de datos consistente, y se trata de un taburete de tres patas: el diseñador, el lector y los datos. Donde cada una de ellas ejerce una fuerza, o aporta una perspectiva independiente, que debe ser considerada para que una visualización sea estable y exitosa.

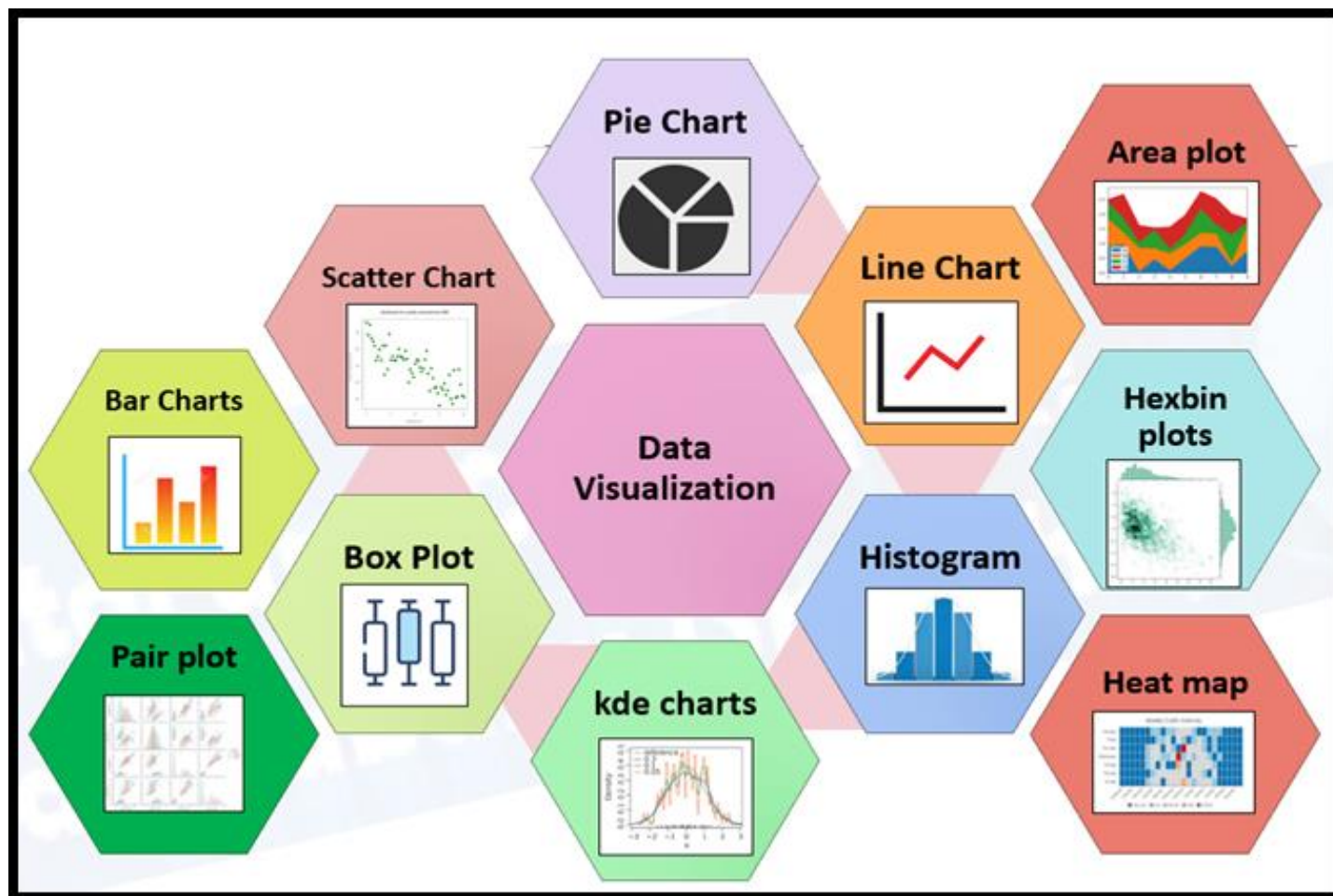


Estrategia para visualizar información



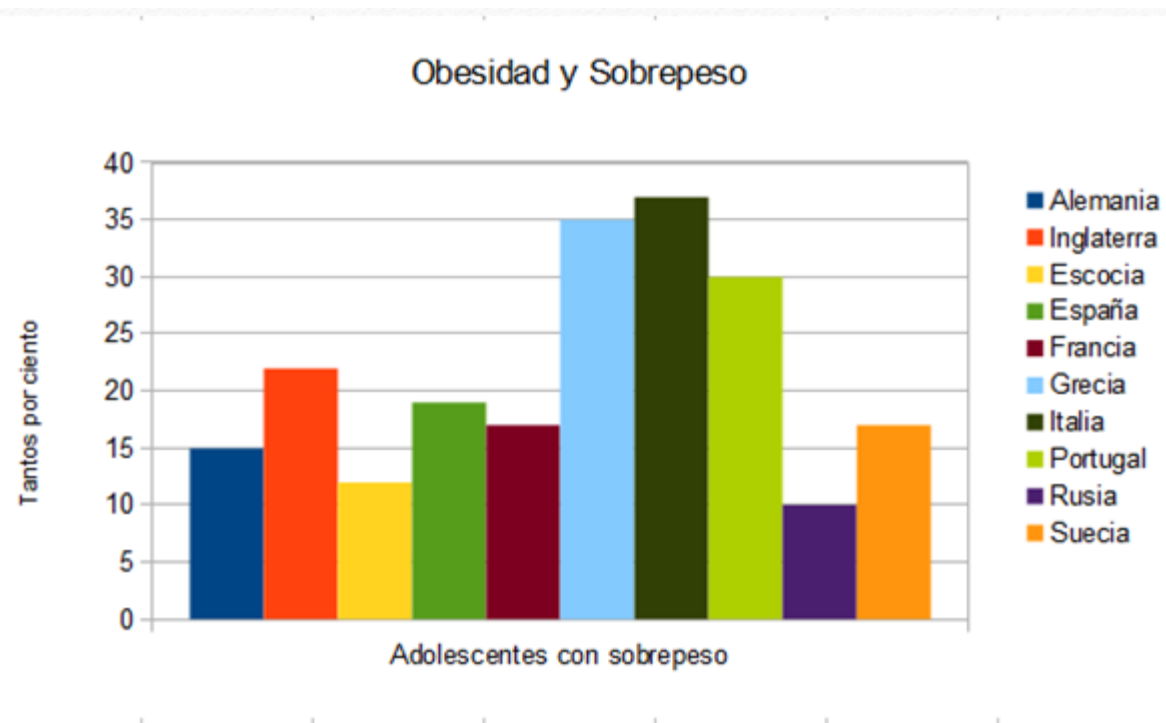
Visualización de Datos

- ▶ Actualmente, existen muchos tipos de visualización o representación de datos:



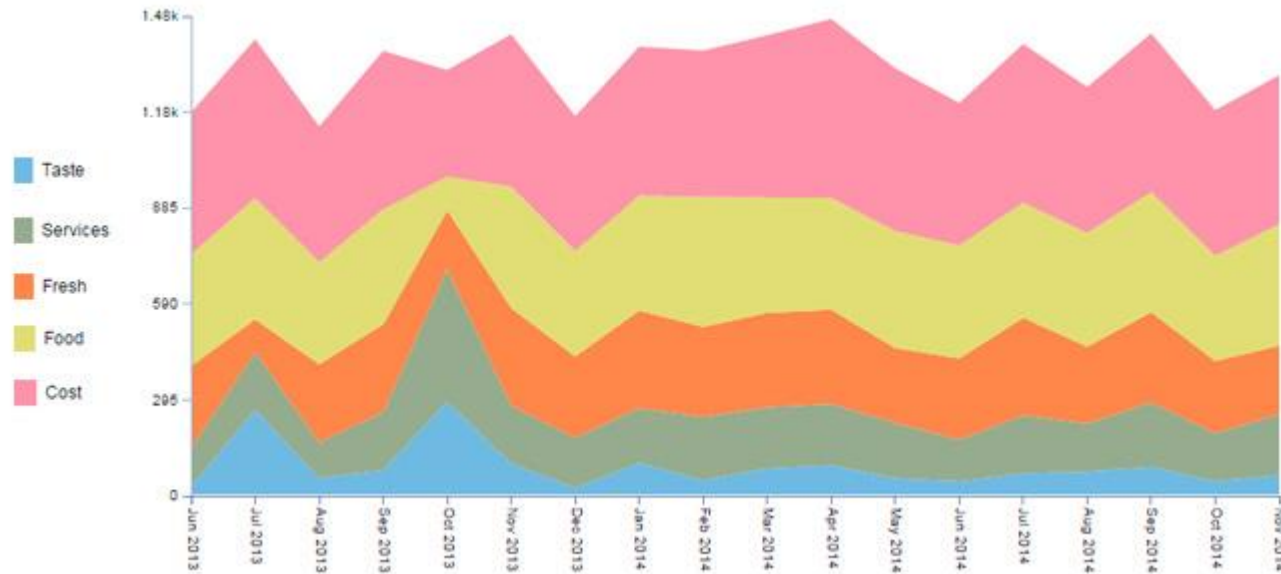
Visualización de Datos

- ▶ Algunos tipos comunes de visualización de datos, los más usados habitualmente:
- ▶ **Gráficos de barras.**- Consiste en representar los datos en un conjunto de barras (verticales u horizontales) para categorizar diferentes puntos de datos (por ejemplo, un gráfico de barras que recoja la cantidad de precipitaciones a lo largo del año en un área geográfica determinada).



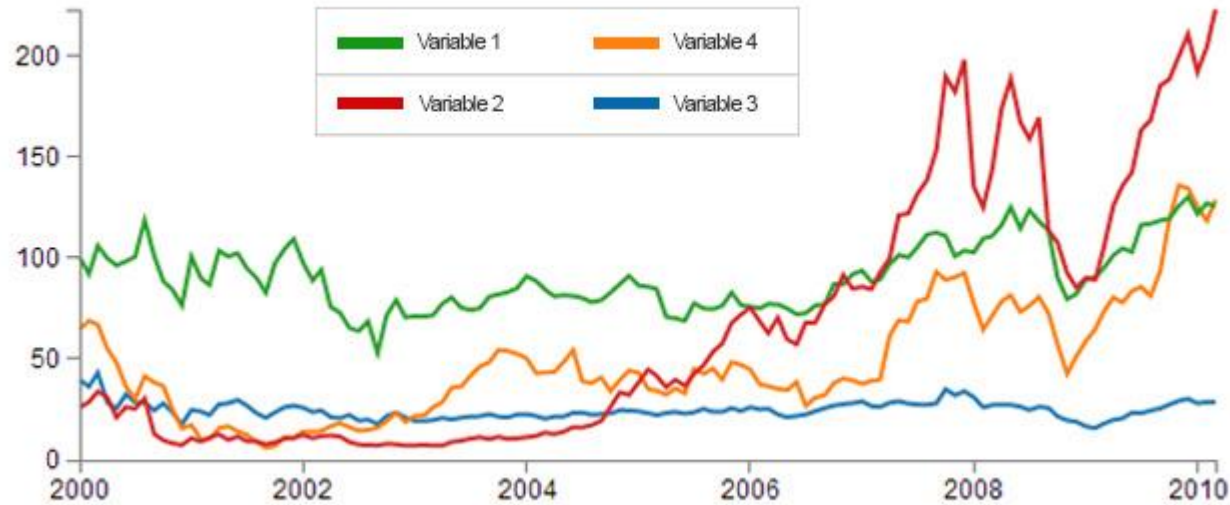
Visualización de Datos

- ▶ Algunos tipos comunes de visualización de datos, los más usados habitualmente:
- ▶ **Gráficos de área.**- Se emplean especialmente para mostrar tendencias y se realizan trazando puntos de datos, unidos por una línea y llenando el espacio bajo dicha línea.



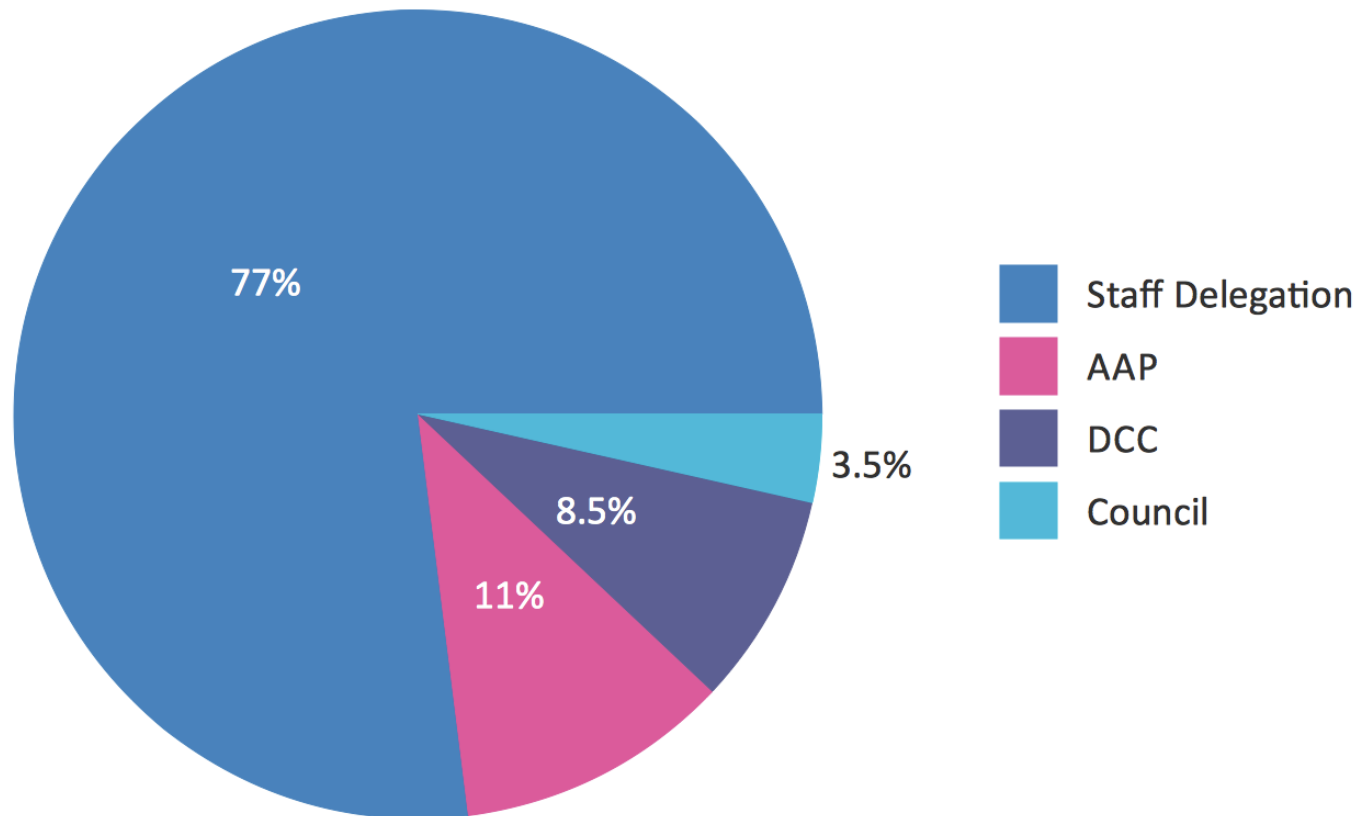
Visualización de Datos

- ▶ Algunos tipos comunes de visualización de datos, los más usados habitualmente:
- ▶ **Gráficos de líneas.** -Se caracteriza por reflejar las cosas a medida que cambian con el tiempo o en categorías ordenadas. Cabe señalar que el número de registros de datos del gráfico de líneas debe ser superior a 2, es mejor no superar las 5 polilíneas en el mismo gráfico.



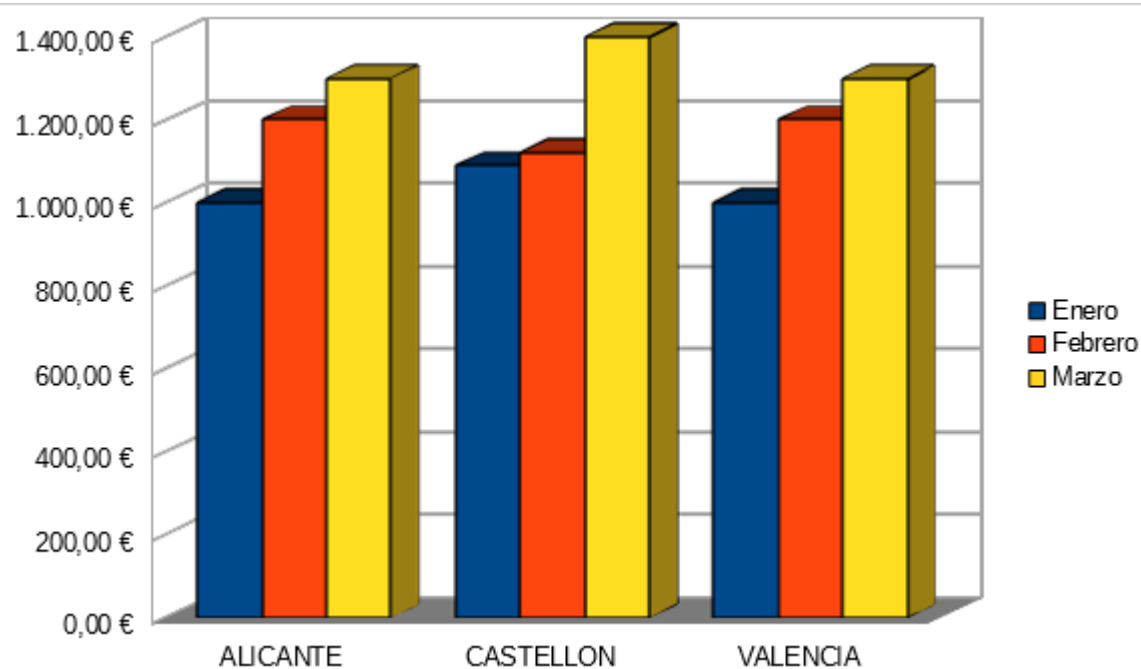
Visualización de Datos

- ▶ Algunos tipos comunes de visualización de datos, los más usados habitualmente:
- ▶ **Gráfico circular o tarta.**- Con este tipo de gráfico se representan cantidades porcentuales de un todo (un ejemplo serían los gráficos de intención de voto de unas elecciones).



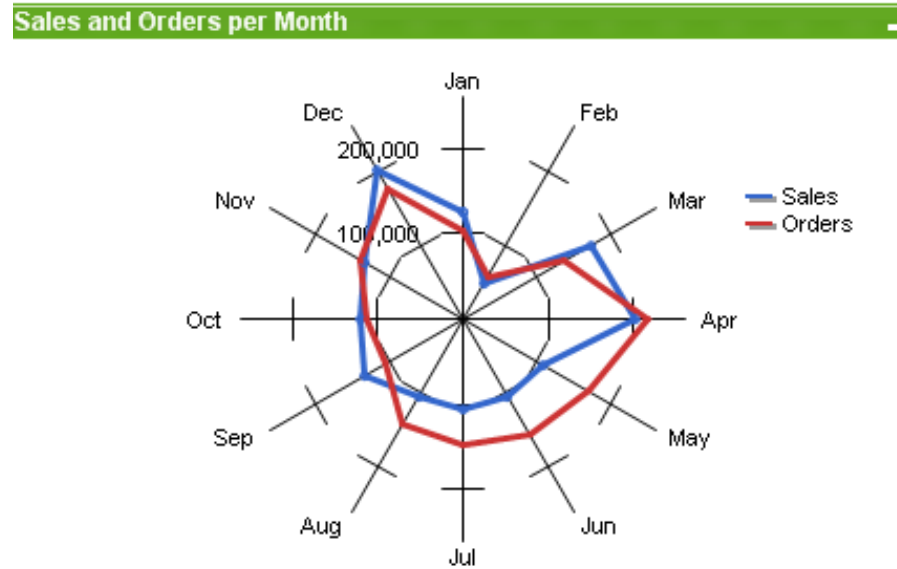
Visualización de Datos

- ▶ Algunos tipos comunes de visualización de datos, los más usados habitualmente:
- ▶ **Gráficos de columnas.**- Los gráficos de columnas, valga la redundancia, utilizan columnas verticales para mostrar comparaciones numéricas entre categorías, y el número de las mismas no debe ser demasiado grande. Este aprovecha la altura de la columna para reflejar la diferencia en los datos, y el ojo humano es sensible a las diferencias de altura. La limitación es que solo es adecuado para conjuntos de datos pequeños y medianos.



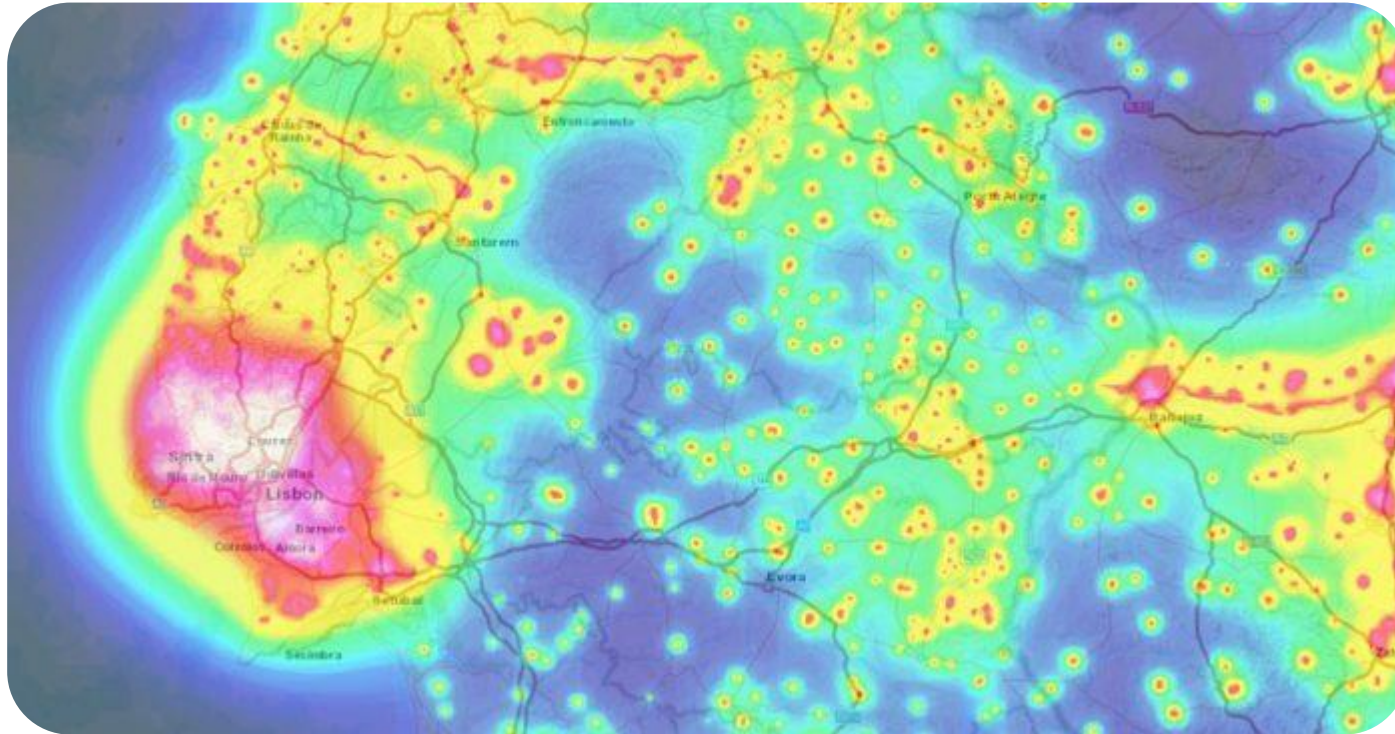
Visualización de Datos

- ▶ Algunos tipos comunes de visualización de datos, los más usados habitualmente:
- ▶ **Gráficos de Radar.**- Los gráficos de radar se utilizan para comparar múltiples variables cuantificadas. En ellos se pueden observar ver qué variables tienen valores similares o si hay valores extremos.
- ▶ También ayudan a ver qué variables del conjunto de datos tienen valores más altos o más bajos.



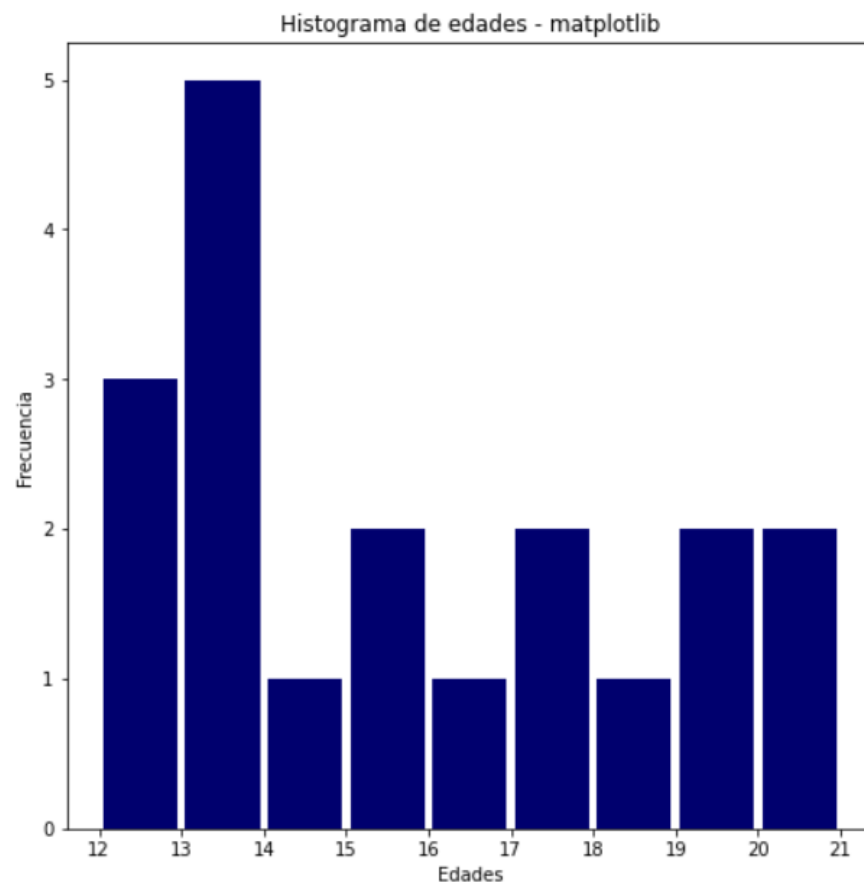
Visualización de Datos

- ▶ Algunos tipos comunes de visualización de datos, los más usados habitualmente:
- ▶ **Mapa de Calor.** - El mapa de calor se utiliza para indicar el peso de cada punto en el área geográfica. La variación del color en un mapa de calor generalmente se refiere a la densidad.



Visualización de Datos

- ▶ Los tipos de visualización de datos más comunes, no son los únicos, existen otros métodos:
- ▶ **Histograma**. - Es la representación gráfica en forma de barras, que simboliza la distribución de un conjunto de datos. Sirven para obtener una "primera vista" general, o panorama, de la distribución de la población, o de la muestra, respecto a una característica, cuantitativa y continua.



Visualización de Datos

- ▶ Los tipos de visualización de datos más comunes, no son los únicos, existen otros métodos:
 - ▶ **Gráficos de calibre.**- Un calibre en la visualización de datos es una especie de gráfico materializado. La escala representa la métrica, el puntero representa la dimensión y el ángulo representa el valor.
 - ▶ Puede representar visualmente el progreso o la situación real de un indicador.

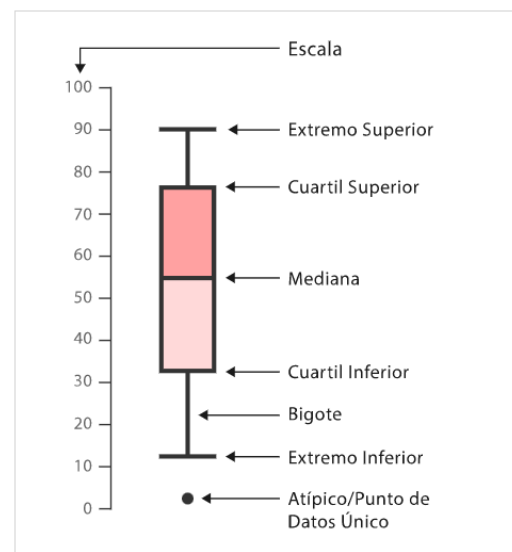
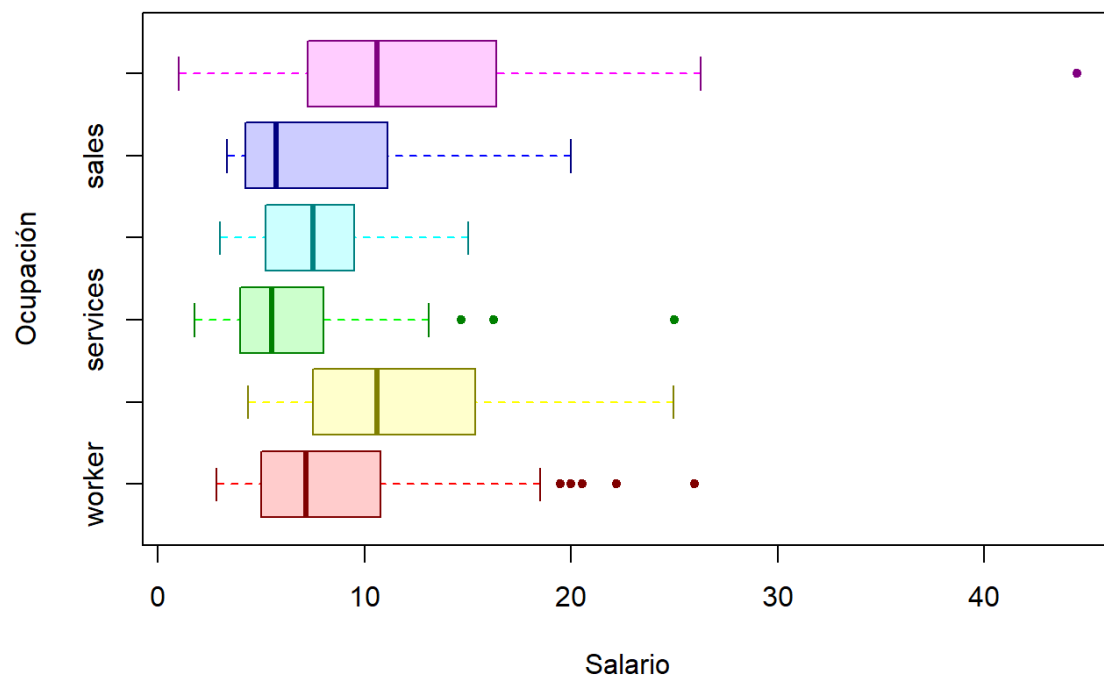
Ventas por Zona - Avance



Visualización de Datos

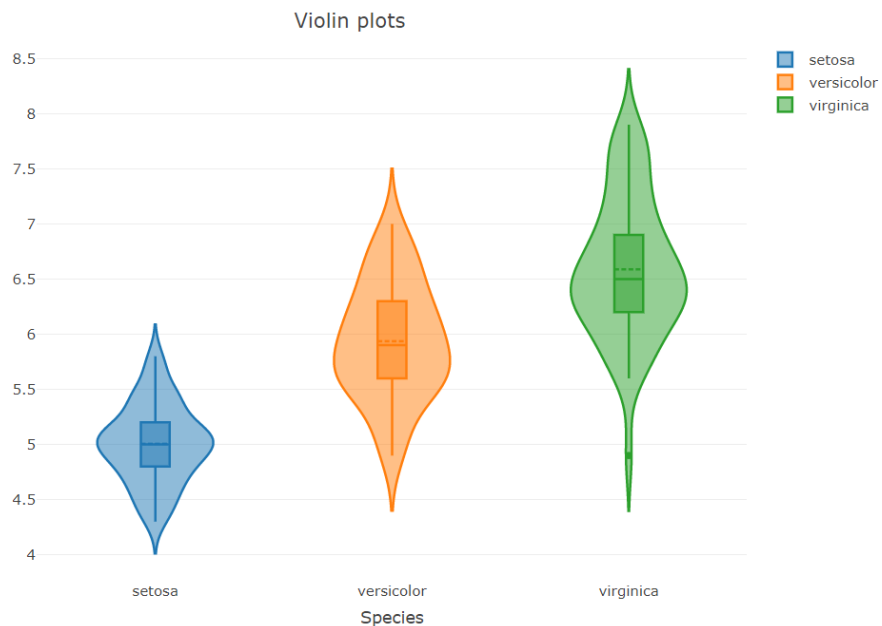
- ▶ Los tipos de visualización de datos más comunes, no son los únicos, existen otros métodos:
- ▶ **Diagrama de caja.**- Un diagrama de caja, del inglés, boxplot, es una representación de una variable cuantitativa o categórica con el propósito de identificar rápidamente los cuartiles del conjunto de datos.

Salario (dólares hora) según ocupación

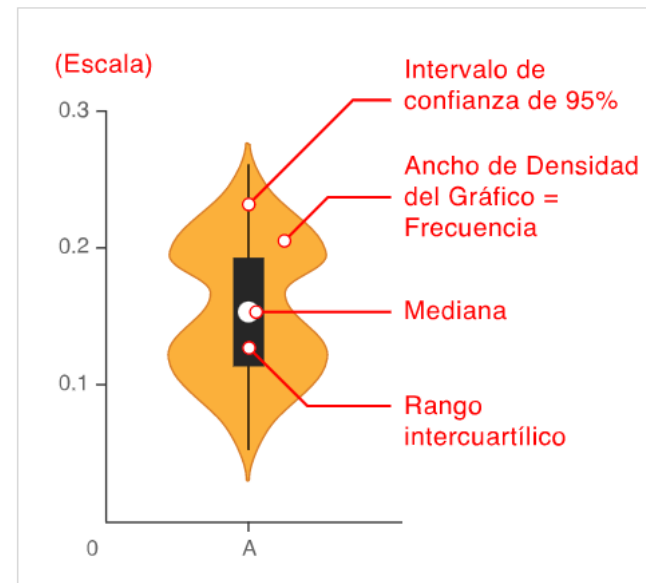


Visualización de Datos

- ▶ Los tipos de visualización de datos más comunes, no son los únicos, existen otros métodos:
- ▶ **Gráficos de violín.**- Estos gráficos nos dicen básicamente lo mismo que uno de caja, pero la diferencia radica en que nos muestran de mejor manera la distribución que tienen los datos, por lo que si encontramos una variable con datos muy homogéneos el gráfico de violín tenderá a ser más delgado, mientras que si encontramos otra con una forma más leptocúrtica el gráfico será más gordo.
- ▶ Se utiliza para visualizar la distribución de los datos y su densidad de probabilidad.

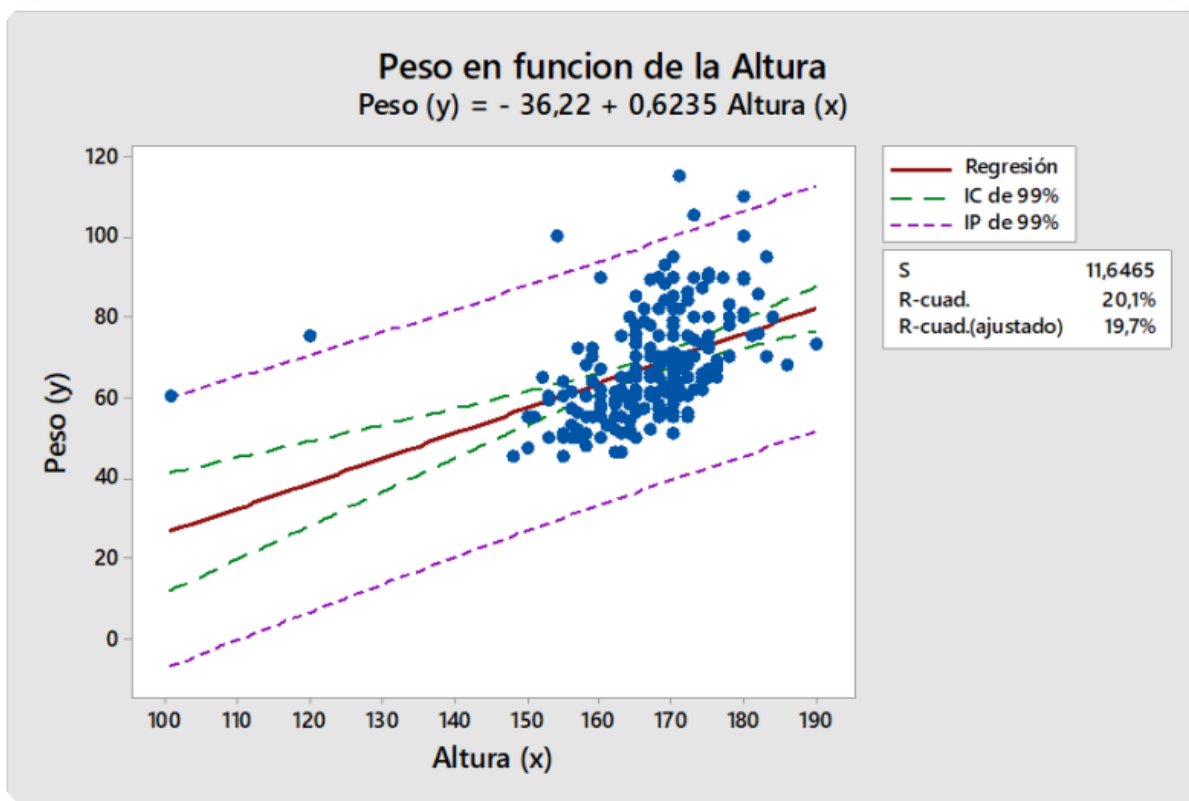


Anatomía



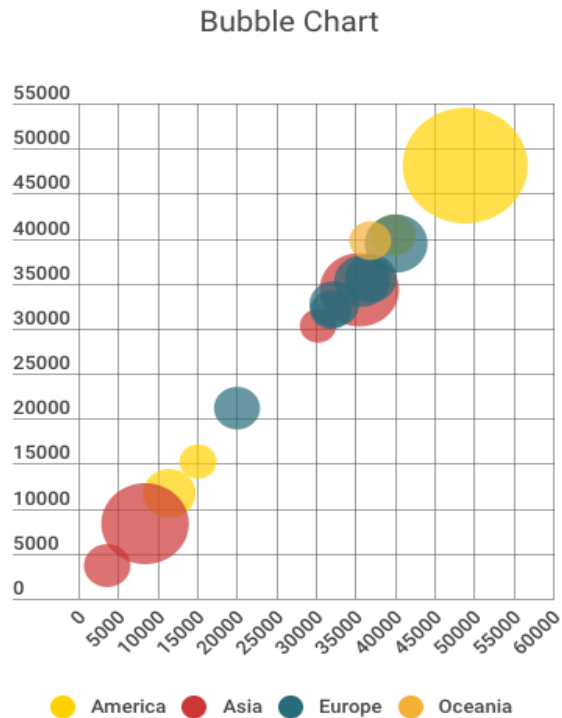
Visualización de Datos

- ▶ Los tipos de visualización de datos más comunes, no son los únicos, existen otros métodos:
 - ▶ **Diagrama de dispersión.**- Permite estudiar las relaciones entre dos conjuntos asociados de datos que aparecen en pares (por ejemplo, (x,y), uno de cada conjunto).
 - ▶ El diagrama muestra estos pares como una nube de puntos correlacionadas.
 - ▶ La correlación no es más que cómo se relacionan ambas variables entre sí.



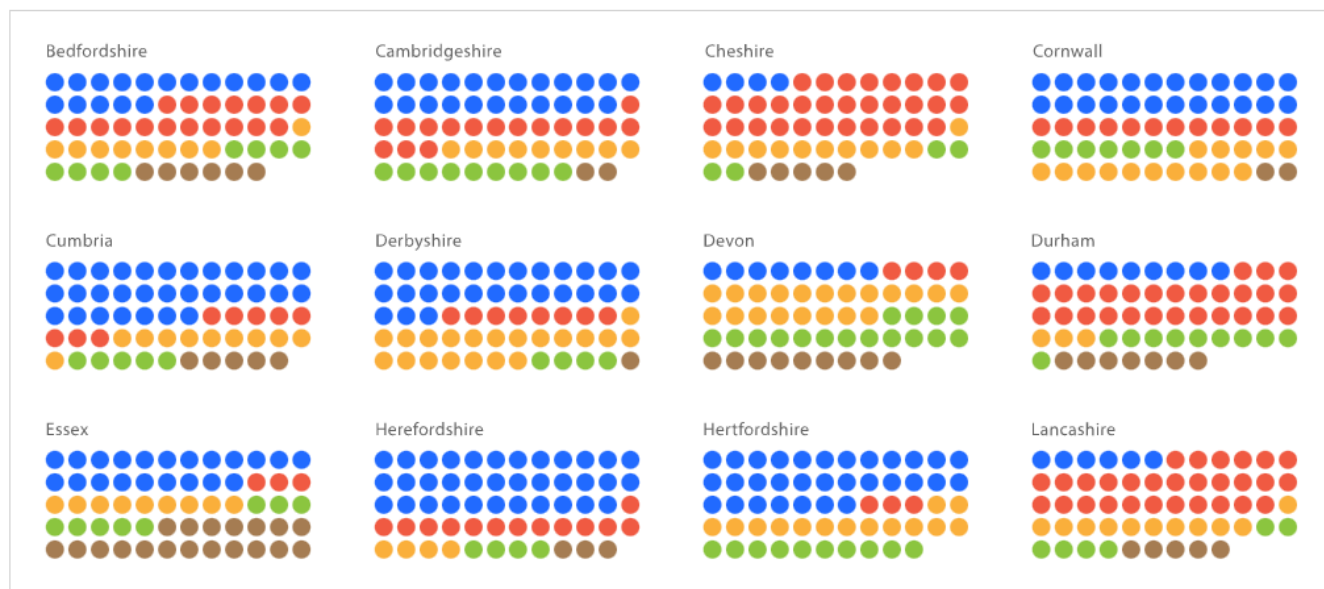
Visualización de Datos

- ▶ Los tipos de visualización de datos más comunes, no son los únicos, existen otros métodos:
- ▶ **Diagrama de Burbujas.**- Es una variación de un gráfico de dispersión en el que los puntos de datos se reemplazan por burbujas y una dimensión adicional de los datos se representa en el tamaño de las burbujas.



Visualización de Datos

- ▶ Los tipos de visualización de datos más comunes, no son los únicos, existen otros métodos:
 - ▶ **Diagrama de Matriz.**- El gráfico de matriz de puntos muestra datos discretos en unidades de puntos, cada uno coloreado para representar una categoría particular y agrupado en una matriz. Se utiliza para dar una visión general rápida de la distribución y las proporciones de cada categoría en un conjunto de datos y también para comparar la distribución y la proporción entre otros conjuntos de datos, con el fin de descubrir patrones.

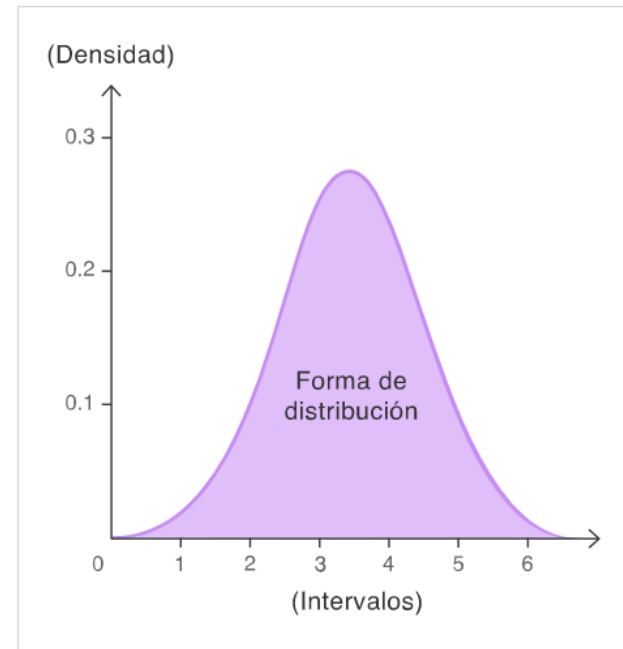
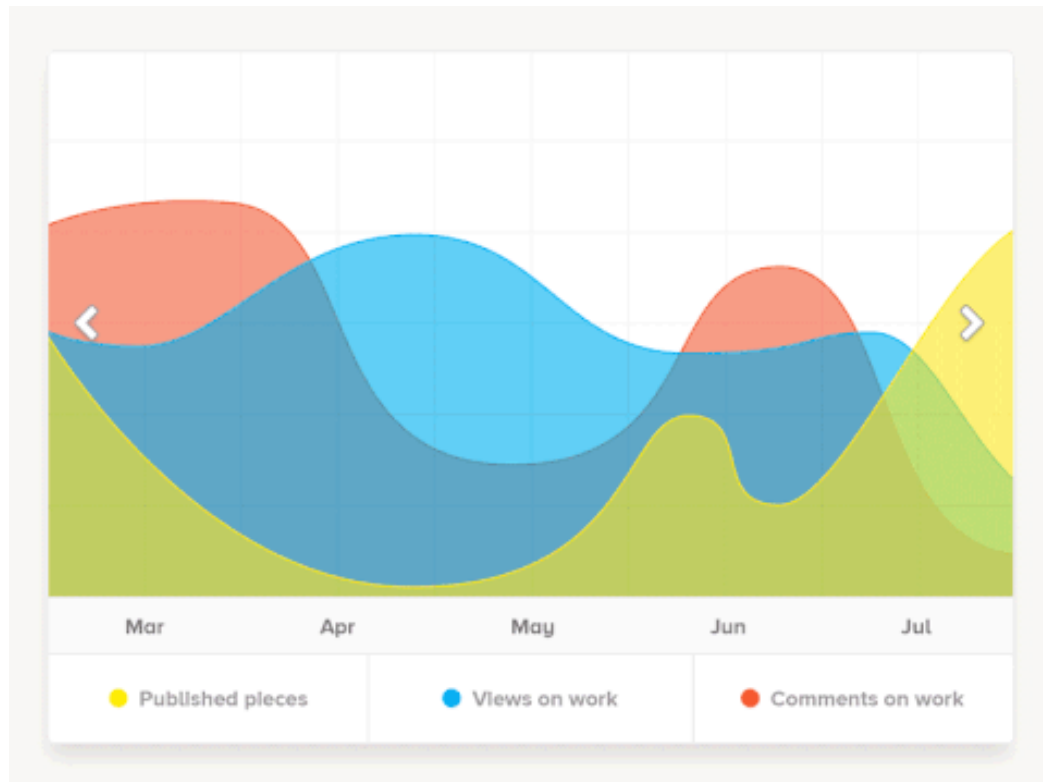


Anatomía



Visualización de Datos

- ▶ Los tipos de visualización de datos más comunes, no son los únicos, existen otros métodos:
- ▶ **Diagrama de Densidad.**- Un diagrama de densidad o curva de densidad es básicamente lo mismo que un histograma, solo que en una versión mucho más suavizada, ya que no se utilizan barras, si no una simple línea que de igual manera nos muestra la distribución de los datos observados, pero esto en porcentajes que deben dar un total de 1.



Visualización de Datos

- ▶ Los tipos de visualización de datos más comunes, no son los únicos, existen otros métodos:
- ▶ **Gráfico de Flujo.**- Este tipo de visualización es una variación del gráfico de área apilada, muestran los cambios en los datos a lo largo del tiempo de diferentes categorías a través del uso de formas fluidas y orgánicas que se asemejan un poco a un arroyo. Son ideales para mostrar conjuntos de datos de alto volumen con el fin de descubrir tendencias y patrones a lo largo del tiempo en una amplia categoría.

Anatomía

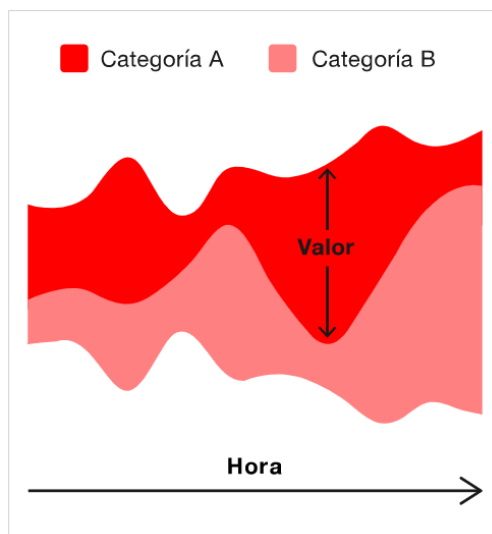
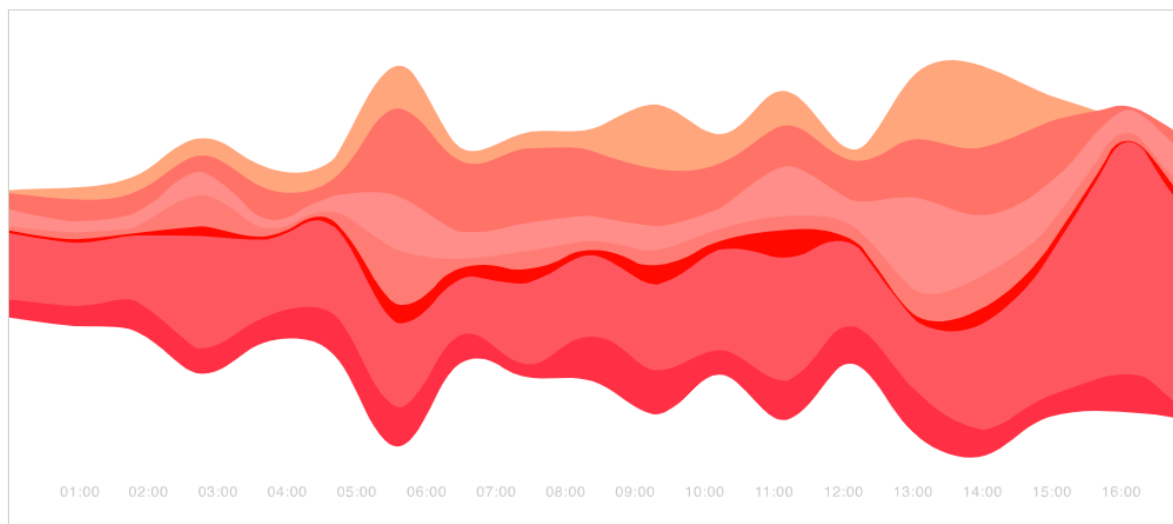


Gráfico de Flujo



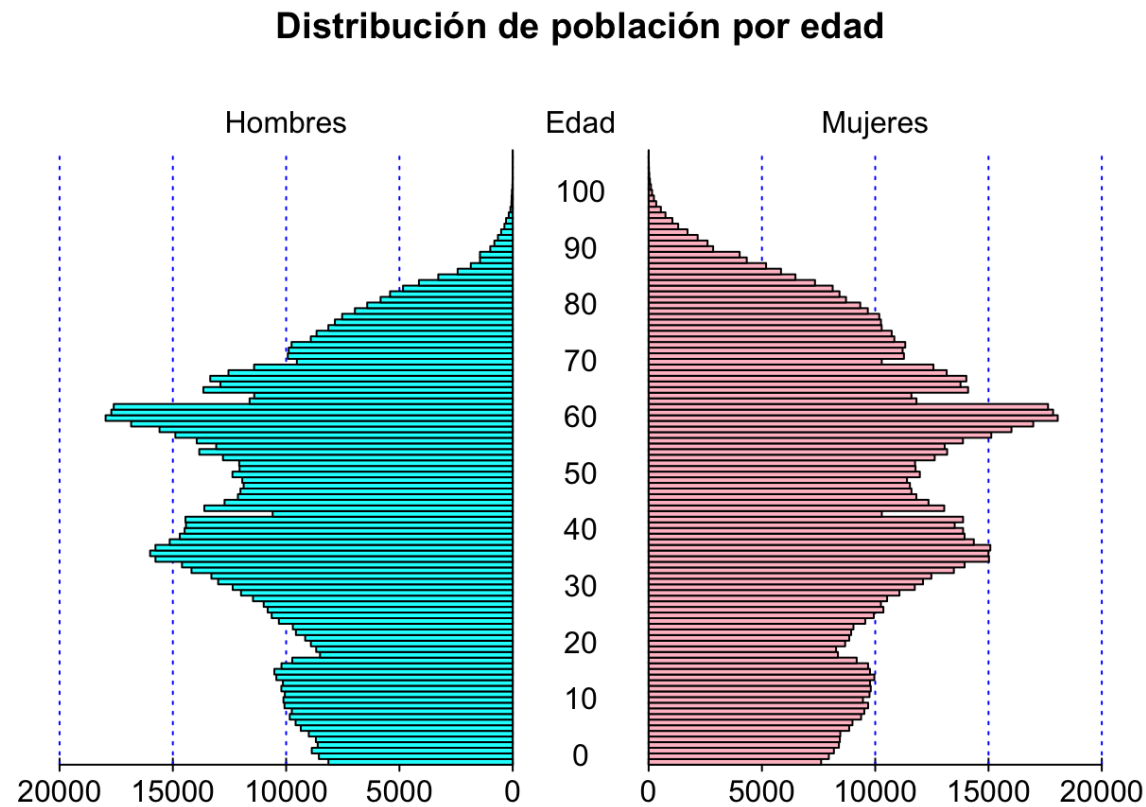
Visualización de Datos

- ▶ Los tipos de visualización de datos más comunes, no son los únicos, existen otros métodos:
- ▶ **Gráfico de Árbol.**- Un mapa de árbol o treemap es un tipo de visualización que dispone los datos de manera jerárquica. Tiene la estructura de un árbol (de ahí su nombre) en la que los datos se organizan en rectángulos anidados (uno dentro de otro). El tamaño del rectángulo corresponde al valor de la categoría o subcategoría.



Visualización de Datos

- ▶ Los tipos de visualización de datos más comunes, no son los únicos, existen otros métodos:
- ▶ **Gráfico de Pirámide.**- Estos gráficos son usados casi exclusivamente para ver como ha variado la distribución de la población para ambos sexos.



Visualización de Datos

- ▶ Para seleccionar un gráfico en función de nuestro análisis o necesidades de comunicación, se recomienda considerar la conexión entre las variables, la tendencia a lo largo del tiempo y el orden relativo de las variables, comparar variables y su distribución. Se sugiere agrupar por las categorías siguientes:

- ▶ **Relación y correlación**

- ▶ Diagramas de dispersión.
- ▶ Histograma.
- ▶ Mapa de calor (matriz).

- ▶ **Evolución en el tiempo**

- ▶ Gráfico de Líneas
- ▶ Gráfico de áreas
- ▶ Gráfico de áreas apiladas
- ▶ Gráfico de flujo

- ▶ **Clasificación**

- ▶ Gráfico de barras verticales
- ▶ Gráfico de barras horizontales
- ▶ Gráfico de barras apiladas
- ▶ Diagrama Matriz de puntos

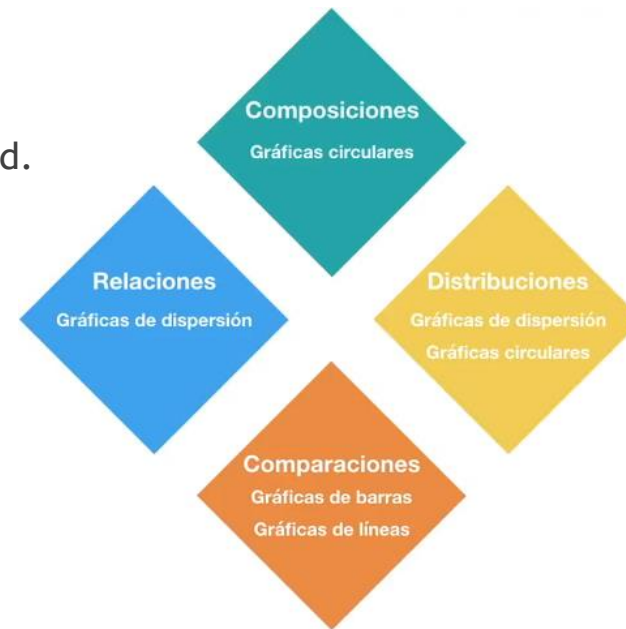
- ▶ **Distribución**

- ▶ Histograma.
- ▶ Diagrama de Densidad.
- ▶ Gráfico de caja
- ▶ Gráfico de violín
- ▶ Gráfico de pirámide

- ▶ **Comparaciones**

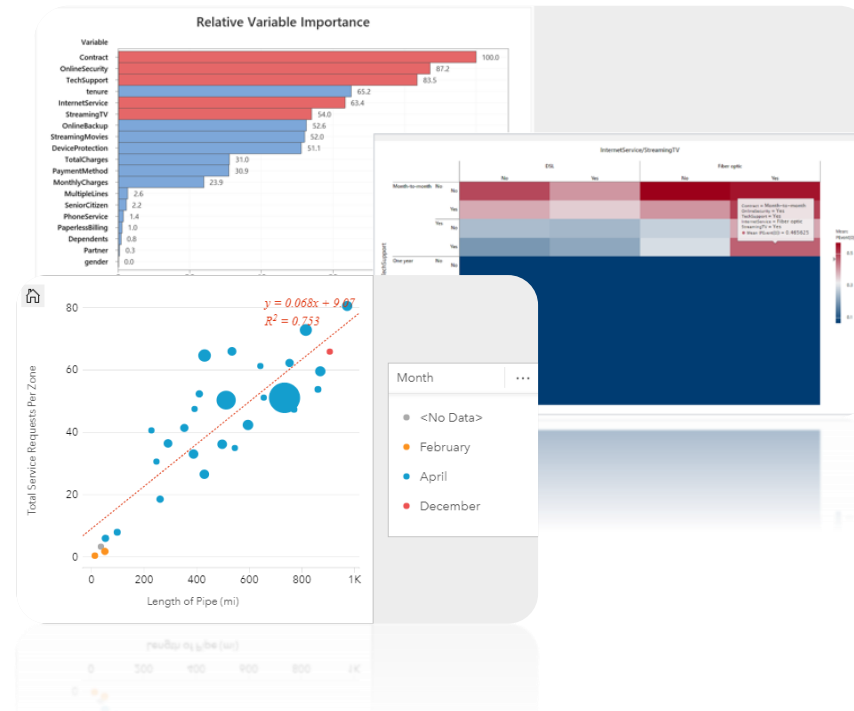
- ▶ Gráfico de Burbujas
- ▶ Gráfico Circular
- ▶ Gráfico de Árbol
- ▶ Gráfico de Calibre

- ▶ **En resumen**, los gráficos, mapas y diagramas ayudan a las personas a comprender datos complejos, encontrar patrones e identificar tendencias.



Visualización de Datos

- ▶ Los diagramas de dispersión, los gráficos de líneas, los mapas de calor y otros tipos de gráficos constituyen **formas eficaces de presentar los resultados visualmente**.
- ▶ Comunicar los conocimientos adquiridos a partir de nuestro análisis, resulta muy importante para justificar las acciones que procedan.
- ▶ Examinar los resultados del análisis y decidir que presentar o reportar como valor o conjunto de valores más importante, considerando indicadores relevantes, y si se cumplieron exitosamente los objetivos del análisis.
- ▶ Debemos tener en cuenta que no todos los resultados pueden ser optimistas, **todos los resultados** deben presentarse de manera que se puedan tomar decisiones informadas
- ▶ El análisis puede mostrar resultados que contradicen lo que se esperaba encontrar, o resultados que sean inconcluyentes o confusos, también se debe mostrar estos resultados.
- ▶ Algunos de estos resultados pueden ser confusos, y los resultados inconcluyentes pueden conducir a nuevos análisis, el objetivo de reportar los resultados es decidir cuál debe ser el siguiente paso.
- ▶ Se recomienda mantener **copias de seguridad** en tablas con los datos de detalle del análisis, si por algún motivo se necesita **profundizar en los resultados**.



Aplicaciones para la Visualización de Datos

- ▶ Existen numerosas opciones de software para la visualización de datos, aunque por lo general, todas tienen un funcionamiento similar:

- ▶ Importan los datos, ofrecen múltiples maneras de visualizar la información y se publican los cuadros de mando desarrollados de manera fácil y sencilla.

- ▶ A continuación, enumeramos algunas herramientas/aplicaciones:

- ▶ R es un paquete de software para análisis general de datos. También cuenta con potentes capacidades de visualización.



- ▶ Python es un lenguaje de programación de propósito general que también tiene una serie de paquetes para soportar el análisis de datos y los gráficos.



- ▶ Plotly es una herramienta online o software colaborativo que permite el desarrollo de visualizaciones complejas y avanzadas de datos. Lo principal de este software es que está completamente integrado con los entornos y lenguajes de programación Python, JavaScript o Matlab, con lo que las posibilidades de personalización y desarrollo son enormes. Puede importar casi cualquier formato de datos y de cualquier fuente, entre ellas de repositorios en la nube como por ejemplo Dropbox.



- ▶ D3 es una biblioteca JavaScript para producir visualizaciones web interactivas y documentos basados en datos.



Aplicaciones para la Visualización de Datos

- ▶ Leaflet Map es una biblioteca JavaScript de peso ligero optimizada para móviles para crear mapas interactivos.
- ▶ Tableau es un software con una simplicidad de uso y capacidad para producir visualizaciones interactivas más allá de las proporcionadas por soluciones generales. es especialmente adecuado para manejar los grandes y cambiantes conjuntos de datos que se utilizan en las operaciones de Big Data, incluidas las aplicaciones de inteligencia artificial y aprendizaje automático, gracias a la integración con una gran cantidad de soluciones de bases de datos avanzadas, incluidas hadoop, amazon aws, my sql , sap y teradata.
- ▶ Tableau Public nos permite crear visualizaciones en nuestro perfil público, y compartirlas o ponerlas en un sitio o blog.
- ▶ Google Data Studio es un visualizador de datos del gigante de la Internet que permite conectar fuentes de información y extraer conclusiones rápidamente. Tiene la posibilidad de administrar paneles propios y compartirlos con otros usuarios, así como se hace en Google Drive.
- ▶ Google Charts proporciona compatibilidad entre navegadores y portabilidad de su plataforma cerrada para iPhone y Android.
- ▶ Timeline es una biblioteca JavaScript que nos permite crear líneas de tiempo.



Aplicaciones para la Visualización de Datos

- ▶ Jupyter es un proyecto de código abierto que permite el análisis de big data, la visualización y la colaboración en tiempo real en el desarrollo de software en más de una docena de lenguajes de programación. la interfaz contiene el campo para la entrada de código, y la herramienta ejecuta el código para entregar la imagen visualmente legible en función de la técnica de visualización elegida.
- ▶ Qlik es una herramienta que se ha convertido en el mayor competidor de Tableau en el último tiempo. También es un software descargable de fácil integración con las fuentes de los datos. Su diferencia está en la capacidad de albergar gran cantidad de KPIs que apoyan los gráficos creados.
- ▶ PowerBI desarrollada por Microsoft, es más compleja de utilizar que Tableau y las herramientas de Qlik, pero para personas que tengan conocimientos sólidos de Excel. Se trata de una herramienta muy potente y dispone de una interfaz intuitiva y sencilla para la creación de cuadros de mando. Además, permite el almacenamiento, tratamiento y descubrimiento de datos. Está disponible en modo On Premise y en la nube (SaaS), y cuenta con aplicaciones para su utilización en móviles y tablets.
- ▶ Jamovi es un software estadístico gratuito y de código abierto para Gnu/Linux, Windows y MacOS, que se puede usar como una alternativa a SPSS. El programa jamovi tiene la apariencia de una hoja de cálculo, pero es una interfaz gráfica de R (en inglés, una GUI, Graphical User Interface).



Aplicaciones para la Visualización de Datos

GRÁFICOS

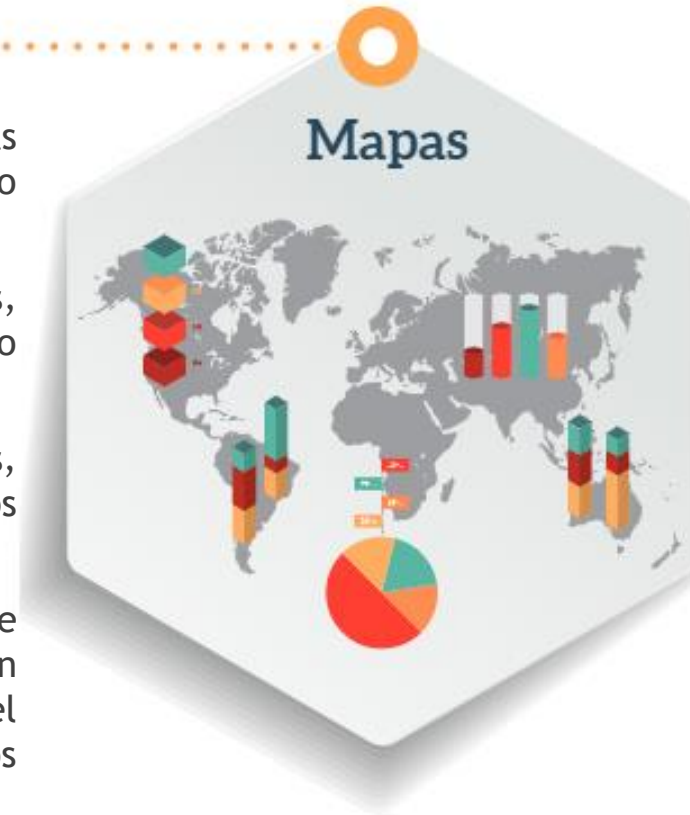
- ▶ Para la elección de la herramienta de visualización a utilizar, lo principal a tener en cuenta es:
 - ▶ **Fuentes de datos** para evaluar si la herramienta es fácilmente integrable con ellas.
 - ▶ **Capacidades y conocimientos de las personas** que la van a utilizar para realizar los cuadros de mando o visualizaciones.
 - ▶ **Volumen de datos**, este punto es vital para manejar la información de forma ágil y asegurarnos de que los informes se procesan adecuadamente.
 - ▶ **Coste de la herramienta/software**. Cada tipo de empresa y entorno tiene unos requisitos y capacidad económica. Precios en dólares - USD, en euros - EUR según fuentes y previa evaluación del producto con los distribuidores del software al momento de adquisición del producto y sus actualizaciones con el tiempo.
 - ▶ **Funcionalidad**. Se refiere a los servicios ofrecidos por la herramienta. Características propias de cada aplicación que hacen que sea práctico y utilitario para el análisis.



Aplicaciones para la Visualización de Datos

MAPAS

- ▶ **Plataforma.** Incluye los tipos de arquitectura, sistemas operativos, ambientes, lenguaje de programación o interfaces de usuarios compatibles.
- ▶ **Integración con otros productos.** Representa módulos, ambientes los cuales son permitidos conexión o combinación con las aplicaciones.
- ▶ **Prestigio de la Herramienta.** Calificación de los usuarios, personas, empresas respecto al uso de la herramienta o los años que lleva en el mercado.
- ▶ **Soporte y Garantía.** Líneas de contacto, foros, sitios de soporte respuesta mediante errores o inconvenientes con la aplicación. Servicio que proporcionan asistencia con el hardware y software para resolver problemas determinados con el producto.
- ▶ **Licenciamiento.** Enfocado al manejo de usuarios, si hay para empresa o se puede manejar en desktop.
- ▶ **Rendimiento.** Capacidad de respuesta ante la solicitud de los usuarios.
- ▶ **Eficiencia.** Se refiere al uso de recursos de cómputo para su desempeño, como por ejemplo memoria RAM, HD, procesador, etc.
- ▶ La elección de una herramienta también depende de ello de todos estos tipos de criterios.



The background features abstract, overlapping geometric shapes in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are primarily located on the right side of the frame, creating a modern, dynamic feel.

Gracias

Ing. Ricardo Velasteguí (M.B.A.)