
Principios de visualización de datos

PID_00249365

Ignasi Alcalde Perea

Tiempo mínimo de dedicación recomendado: 1 hora



Índice

Introducción.....	5
1. Breve historia de la visualización de datos.....	7
2. La visualización de datos en la inteligencia de negocio.....	8
3. Visualizaciones explicativas.....	9
4. Visualizaciones exploratorias.....	10
5. La memoria y la visión.....	12
6. Variables visuales.....	13
7. Breves conceptos sobre los datos: medidas y dimensiones.....	14
8. Fuentes de datos.....	15
9. Catálogos de gráficos.....	16
10. Herramientas de visualización de datos.....	17
11. Diez consejos a la hora de visualizar datos.....	19
Resumen.....	21
Bibliografía.....	23

Introducción

Con el *big data* la cantidad de datos disponibles en el ámbito empresarial se dispara; sin embargo, todos los datos en sí mismos son inútiles; de hecho, pueden convertirse en un pasivo para la empresa si no podemos entenderlos.

El dato en sí, en crudo, es difícil de consumir, de contextualizar y de darle sentido.

La visualización de datos permite comprender estos rápidamente, aparte de la historia que hay escondida detrás de ellos, yendo más allá de la creación de un *dashboard* o cuadro de mando.

Pero ¿por qué visualizamos los datos? Visualizamos la información para satisfacer una necesidad muy básica: contar una historia o comunicar y facilitar la toma de decisiones.

Los seres humanos somos seres visuales por naturaleza y **comprendemos visualmente**. Diversos estudios han descubierto que las imágenes visuales aumentan la disposición a leer en un 80 % y que percibimos una escena visual en menos de 0,10 segundos. La visualización de datos nos permite reconocer fácilmente patrones y encontrar excepciones mientras interpretamos los datos de una forma más eficaz.

Con el paso del tiempo, hemos encontrado nuevas maneras de visualizar la información. Hoy en día, estamos familiarizados con los tipos de gráficos básicos, como el gráfico de líneas, el gráfico de barras y el gráfico circular o de tarta. Sin embargo, rara vez nos detenemos a pensar por qué son más eficaces que un texto un número o una tabla. Si no lo hacemos, tampoco podremos detectar fácilmente en qué casos los gráficos están mal diseñados y cuáles pueden mejorarse.

El propósito de este material es ofrecer las bases que permitan comprender de una forma básica cómo procesamos la información visual y que este conocimiento permita diseñar cuadros de mando de manera más clara y eficaz.

1. Breve historia de la visualización de datos

La visualización de datos tiene sus raíces en la representación de datos científicos e históricos mediante gráficos que sirvieran para comprender comportamientos, detectar oportunidades y tomar mejores decisiones. Algunos de los ejemplos más relevantes por orden cronológico son estos:

- **Charles Minard: Napoleon's Russian Campaign.** Este gráfico de la campaña rusa de Napoleón de 1812 es considerado el mejor gráfico estadístico de la historia.
- **John Snow: The Cholera Epidemic in London.** Con este mapa de Londres de 1854, John Snow cambió la forma en que se estudia una epidemia.
- **Florence Nightingale: Mortality in the army.** Con este gráfico sobre la mortalidad de la Armada británica en 1958, Florence Nightingale introdujo la estadística en el mundo de la medicina, revolucionando el modo de entender la enfermería.
- **Otto Neurath: ISOTYPE Language.** El filósofo Otto Neurath, durante su cargo de director del Social and Economic Museum of Vienna a principios de los años veinte, creó el lenguaje ISOTYPE con la finalidad de representar datos estadísticos de forma visual.

Más sobre visualización de datos

Se pueden explorar más ejemplos de relevancia histórica en la visualización de datos en los siguientes recursos:

- **Brief History of Data Visualizations.** Breve historia ilustrada de la visualización de datos.
- **Visualization Milestones.** Una cronología ilustrada de las principales innovaciones en la historia de la visualización de datos.

En INFRAGISTICS podéis encontrar una visualización sobre los momentos históricos más relevantes en el desarrollo de esta disciplina.

2. La visualización de datos en la inteligencia de negocio

Un gráfico puede comunicar más información que una tabla en un espacio mucho más pequeño y de una forma más eficaz. Por ejemplo, en la tabla inferior, si intentamos detectar el mes con las ventas más altas gastaremos unos cuantos segundos:

Enero	Febrero	Marzo	Abril	Mayo	Junio
45	56	36	58	75	62

Pero si realizamos una visualización de los mismos datos con una gráfica adecuada, en este caso una gráfica de líneas, lo detectaremos más rápidamente:



Como vemos en el ejemplo, es importante para un negocio visualizar los datos de una forma rápida y eficaz que facilite la comprensión.

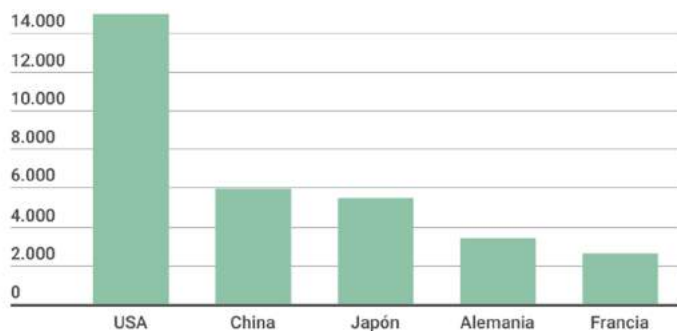
Edward Tufte, uno de los expertos más reputados en la visualización de datos, indica:

«El gráfico excelente es el que da al espectador el mayor número de ideas en el menor tiempo con la menor tinta en el espacio más pequeño.»

En un entorno empresarial, las visualizaciones pueden tener dos objetivos generales, que a veces se superponen: el explicativo y el exploratorio.

3. Visualizaciones explicativas

Las visualizaciones de datos explicativas están destinadas a dirigir al espectador a lo largo de un camino definido. La mayor parte de los cuadros de mando de negocios que encontramos en nuestra vida cotidiana pertenecen a esta categoría. Explicamos datos para resolver problemas específicos o preguntas que tenemos en mente, como por ejemplo: ¿qué país tiene el valor más alto de ventas?



Este tipo de visualización explicativa se utiliza normalmente en los siguientes casos:

- Responder una pregunta. Por ejemplo, ¿cuántas ventas tuvimos en el último trimestre?
- Apoyar una decisión. Por ejemplo, tenemos que aumentar el *stock* del producto «x» ya que se vendieron todos los días la semana pasada.
- Comunicar información. Por ejemplo, los ingresos previstos para el trimestre.

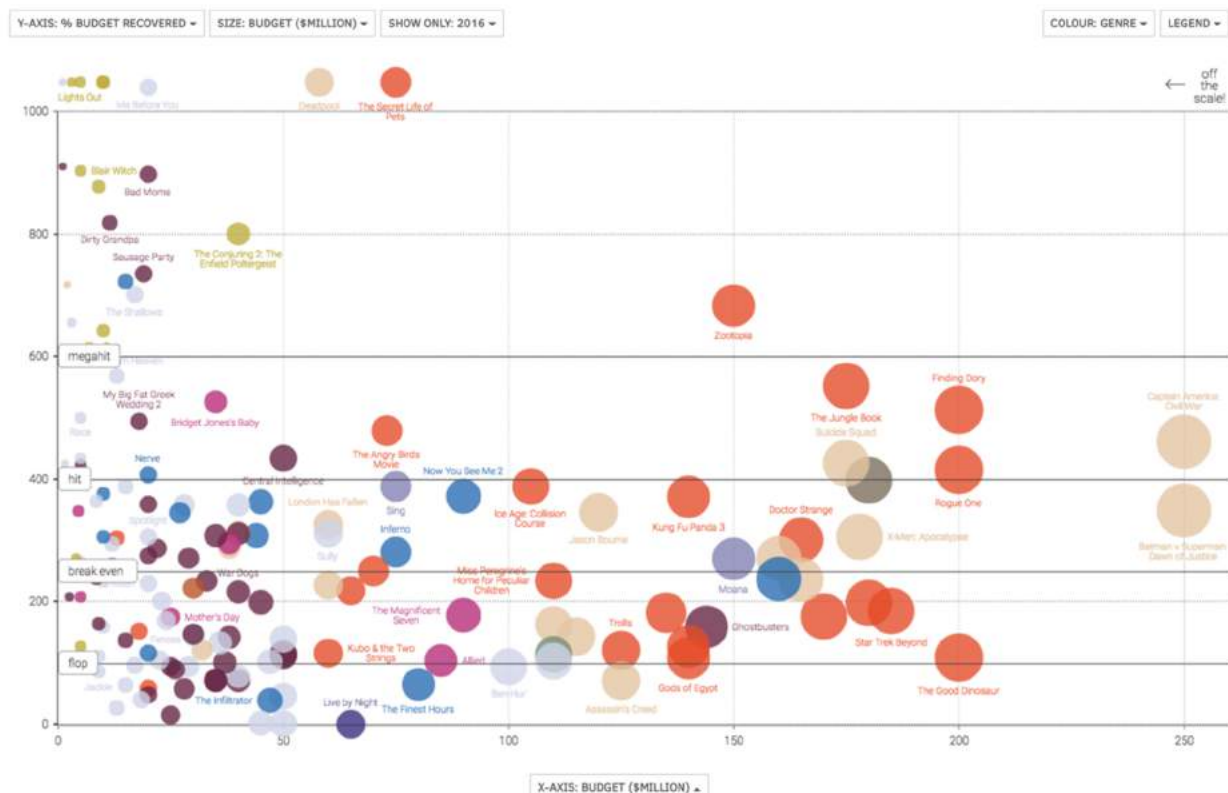
4. Visualizaciones exploratorias

Las visualizaciones de datos exploratorias son ideales para conjuntos de datos grandes o *big data*, ya que permiten al espectador explorar varias dimensiones de un conjunto de datos. Son dinámicas y permiten al lector comparar diversos datos según las preguntas que quiera hacerse sobre los datos.

Por ejemplo, en la visualización de Hollywood In\$ider que presentamos a continuación, se permite explorar diversos parámetros en la gráfica de dispersión, como el presupuesto de una película, el margen de beneficio, la audiencia, etc. Tras familiarizarnos con la visualización, podemos hacernos preguntas según nuestra área de interés.

THE HOLLYWOOD IN\$IDER

Visualization explorer for every major film 2008-2016



Una pregunta que podríamos formular a modo de ejemplo sería: ¿qué película generó mayor beneficio en 2016 en el género de aventuras?

El análisis exploratorio puede ser cíclico, sin un punto final específico. El usuario de la visualización puede explorar varias veces, de forma anidada, la visualización hasta que le permita comprender las diversas preguntas que se formule, y esto le ayude a tomar una decisión concreta.

Este tipo de visualización explicativa se utiliza normalmente en los siguientes casos:

- Plantearse nuevas preguntas.
- Explorar o descubrir.

Las visualizaciones exploratorias son las más adecuadas para el *big data*, donde el análisis exploratorio es necesario.

Es muy importante entender cómo procesamos la información visual y el papel de la memoria en la percepción de la información visual, y cómo aplicarla para facilitar la comprensión.

5. La memoria y la visión

Es muy importante analizar el papel de la memoria en nuestra visión. Hay dos tipos de memoria que entran en juego cuando procesamos información visual:

- **La memoria a largo plazo.** Este tipo de memoria nos hace esperar siempre que las unidades se marquen en los ejes X e Y de un gráfico, o el selector de rango de fechas esté localizado en la parte superior. La memoria a largo plazo está formada por las interacciones y experiencias con las visualizaciones que hemos hecho en el pasado. Debemos tenerla en cuenta siempre que vamos a diseñar la estructura de un *dashboard* o visualización.
- **La memoria de trabajo.** Se trata de la memoria que activamos cuando miramos un gráfico de líneas o un número en un tablero; utilizamos nuestra memoria de trabajo para almacenar solo la información que necesitamos en este momento. Este tipo de memoria rompe toda la visualización en pequeños trozos de información, en un proceso llamado fragmentación. Nuestra mente, como máximo, puede contener tres fragmentos de información a la vez.

Es muy importante, cuando diseñemos una visualización o *dashboard*, limitar el número de elementos que queremos que el usuario observe en los «trozos de información». Es clave sintetizar la mayor información posible en cada trozo, para evitar la sobrecarga de información y mantener una buena comprensión de cada uno.

6. Variables visuales

Para que los objetos gráficos que hemos elegido puedan producir significado, solemos aplicarles variables visuales. Nos referimos a aquellas diferencias que podemos aplicar a los objetos gráficos que permitirán que se establezcan relaciones entre ellos. Jaques Bertin, en su libro *Semiology of Graphics* (1967), establece siete variables básicas:

- Color
- Textura
- Tonalidad
- Tamaño
- Forma
- Orientación
- Posición

De hecho, estamos hablando de **psicología de la percepción**. Aunque no entraremos a fondo en ello, en la figura siguiente podemos ver las variables visuales más comunes y su grado de precisión:



Como podemos comprobar, estas variables visuales entran en juego cuando analizamos cualquier visualización.

Ejemplo

De los elementos de esta ilustración solo «posición» y «longitud» pueden usarse si queremos percibir datos cuantitativos con precisión. Los demás son útiles para percibir otros tipos de datos tales como datos categóricos o relacionales.

7. Breves conceptos sobre los datos: medidas y dimensiones

Como ya hemos comentado, antes de visualizar es importante modelar los datos apropiadamente. Normalmente, podemos trabajar ciertas características de los datos que podemos agrupar en dos bloques:

- **Medidas.** Las medidas las constituyen datos numéricos que pueden calcularse o agregarse, como la suma de ingresos, el coste promedio, el beneficio por persona..., o cualquier dato numérico que podamos imaginar. Con las medidas podemos visualizar observaciones sobre los datos o valores calculados.
- **Dimensiones.** Las dimensiones constituyen datos categóricos como año, producto, país o rango de salarios. Las medidas más comunes son:
 - Categoría, o también llamada nominal para valores discretos.
 - Ordinal, cuando los componentes de la dimensión tienen un orden por defecto.
 - Intervalo, cuando cada valor de la dimensión representa un rango de valores, por ejemplo, rango de notas de una PEC.

8. Fuentes de datos

En este apartado comentaremos brevemente algunos de los bancos de datos abiertos más comunes.

- **World Data Bank.** El Banco Mundial de Datos es una organización sin ánimo de lucro cuyo objetivo es reducir la pobreza y apoyar el desarrollo. Mantiene un banco de datos sobre desarrollo de consulta pública y abierta.
- **Gapminder.** Gapminder es una fundación privada sin ánimo de lucro que promueve el desarrollo sostenible a partir de la comprensión del mundo a través de los datos estadísticos. Como ellos mismos dicen: «Somos un ‘museo’ moderno que hace que el mundo sea más comprensible». Sus datos son abiertos.
- **OECD.** Banco de Datos de la Organización para la Cooperación y el Desarrollo Económico.
- **Base de Datos de las Naciones Unidas.** Banco de datos de acceso libre que reúne los datos de los últimos sesenta años recogidos por el Departamento de Estadística de las Naciones Unidas.
- **Open Data Barcelona.** Catálogo abierto de datos estadísticos sobre la ciudad de Barcelona.

9. Catálogos de gráficos

Elegir el gráfico adecuado para cada visualización no es trivial, pues depende de lo que queramos hacer con él: comparar, distribuir, analizar dimensiones, relaciones, etc.

Existen tres catálogos de gráficos en línea que nos pueden facilitar esta tarea. En ellos se nos dan una descripción de cada gráfico, su anatomía y sus funciones principales, así como variaciones de los mismos:

- [DataViz Catalogue](#)
- [DataViz Project](#)
- [The Python Graph Gallery](#)

10. Herramientas de visualización de datos

En la actualidad, disponemos de múltiples herramientas de visualización de datos, tanto para una visualización explicativa como para una explorativa. En esta sección exponemos un breve resumen de la más populares o más innovadoras. La mayoría de ellas requiere licencia o suscripción, pero disponen de versiones gratuitas de uso personal.

- **Tableau.** Tableau es, a menudo, considerado como «la herramienta» de visualización de datos. Dispone de una base de más de 57.000 usuarios, en diversos sectores, debido a su simplicidad de uso y capacidad de producir visualizaciones interactivas mucho más allá de las proporcionadas por las soluciones de BI más generales. Es adecuada para manejar los enormes conjuntos de datos que se usan en las operaciones de *big data*, incluyendo inteligencia artificial y aplicaciones de *machine learning*, gracias a la integración con un gran número de soluciones de bases de datos avanzadas como Hadoop, Amazon AWS, My SQL, SAP o Teradata.
- **Qlikview.** Qlik, con su herramienta Qlikview, es el software de visualización de datos competidor de Tableau. Tiene más de 40.000 usuarios en más de 100 países y los que lo utilizan frecuentemente comentan como una ventaja clave su configuración altamente personalizable. Cabe remarcar que su curva de aprendizaje es alta si queremos sacar su máximo potencial. Además de la visualización de datos, Qlikview ofrece funcionalidades de inteligencia empresarial, analítica y generación de informes empresariales, y la interfaz de usuario es limpia y ordenada. Qlikview se utiliza junto con su paquete hermano, QlikSense, que maneja la exploración de datos. Dispone de una comunidad en línea muy amplia, con muchos recursos y tutoriales para ayudar a los nuevos usuarios a entender cómo integrarlo en sus proyectos.
- **Power BI.** Power BI es un software de visualización de datos de Microsoft orientado a usuarios con un perfil no técnico que permite agregar, analizar, visualizar y compartir datos. La interfaz de usuario de Power BI es bastante intuitiva para aquellos usuarios familiarizados con Excel, y su profunda integración con otros productos de Microsoft lo convierte en una herramienta muy versátil que requiere poco tiempo de aprendizaje si estás familiarizado con los productos de Microsoft.
- **Infogram.** Infogram es un software de visualización de datos en la nube que permite crear gráficos interactivos con un estilo infográfico. Tiene una versión gratuita restringida y dos opciones de pago que incluyen características como más de 200 mapas o archivos compartidos. Dispone de una interfaz fácil de usar, su conjunto de gráficos está bien diseñado y permite conectar a fuentes de datos en tiempo real, como Google Sheets o JSON.
- **FusionCharts.** Este paquete de visualización de gráficos está basado en JavaScript y se ha establecido como uno de los líderes en el mercado de

pago. Puede producir 90 tipos de gráficos diferentes y se integra con un gran número de plataformas y marcos que dan una gran flexibilidad. Una característica que le ha hecho muy popular es que, en lugar de tener que iniciar cada nueva visualización desde cero, los usuarios pueden elegir entre una serie de plantillas de ejemplo, simplemente conectando sus propias fuentes de datos según sea necesario.

- **Highcharts.** Al igual que FusionCharts, requiere una licencia para uso comercial, aunque puede ser utilizado gratuitamente para uso personal. Su página web afirma que lo utilizan 72 de las 100 compañías más grandes del mundo y que, a menudo, es la herramienta elegida cuando se desea implementar una solución rápida y flexible, con una mínima necesidad de formación especializada en visualización de datos. Una de las claves de su éxito ha sido su enfoque hacia el soporte multinavegador; es decir, que cualquiera puede ver y ejecutar sus visualizaciones interactivas, lo que no siempre ocurre con las plataformas más nuevas.
- **Datawrapper.** Datawrapper es cada vez más popular, particularmente entre las organizaciones de medios que lo utilizan para crear gráficos y presentar estadísticas. Tiene una interfaz simple y clara que hace que sea muy fácil cargar datos csv y crear gráficos sencillos, y también mapas, que pueden ser incorporados rápidamente a los informes.

11. Diez consejos a la hora de visualizar datos

A continuación, presentamos diez consejos que deberíamos tener en mente antes de crear una visualización de datos para convertir los datos en conocimiento y, así, permitir la comprensión por parte de nuestros usuarios:

- 1) **Pensad siempre en el público objetivo.** Es vital personalizar cualquier visualización de datos para satisfacer las necesidades de información de nuestra audiencia. Debemos tener siempre en mente a nuestro público objetivo. Piensa siempre quién es y cómo es ese público, y a qué pregunta les va ayudar a responder la visualización de datos.
- 2) **Elegid el gráfico adecuado.** No todos los gráficos sirven para todo. Cada uno de los gráficos está pensado para mostrar un tipo de información de forma eficaz, en función de lo que queramos hacer. Nos podemos realizar las siguientes preguntas: ¿está bien representado?, ¿he elegido el gráfico adecuado?, ¿las escalas son adecuadas para lo que queremos representar/comparar?
- 3) **La forma sigue la función.** La forma de los gráficos debe pensarse en función de lo que se quiere hacer con los datos. Por ejemplo, para comparar una categoría con categorías más pequeñas en las que hay una relación de cada parte con el todo, la gráfica ideal sería la de barras apiladas, ya que muestran el porcentaje del total de cada grupo y se representan por el porcentaje de cada valor en la cantidad total.
- 4) **Resaltad la información más importante.** Al diseñar una visualización debemos facilitar la lectura con principios como el color, el tamaño o la fuente que permitan dirigir la atención a los puntos más importantes de información.
- 5) **Usad las tablas lo menos posible.** Las tablas solo deben usarse cuando necesitemos mostrar valores precisos. Los gráficos deben utilizarse para mostrar información sobre relaciones entre los datos, patrones, tendencias, cómo cambia la información con la variable tiempo, etc. Siempre que podamos, debemos reducir el uso de tablas y aumentar el de gráficos.
- 6) **Proporcionad contexto a los datos.** Una buena visualización siempre requerirá algún tipo de interacción con los datos que se presentan, por lo que es muy importante proporcionar contexto para que esta se pueda llevar a cabo con éxito; por ejemplo, jugando con el color o el tamaño e incluyendo pequeñas descripciones cortas que resalten las ideas clave.
- 7) **Alinead los gráficos de datos correctamente.** Aseguraos de que los diferentes gráficos y pantallas de la visualización estén alineados horizontal y verticalmente para que puedan compararse con precisión y no crear ninguna ilusión óptica engañosa.
- 8) **Usad el color de forma inteligente.** El color es poderoso. Podemos usarlo para llamar la atención sobre las zonas clave de los datos, pero debemos tener cuidado al seleccionar combinaciones de colores. Por ejemplo,

si usamos el rojo y verde en el mismo gráfico, las personas daltónicas no los diferenciarán.

- 9) **El título del gráfico debe proporcionar información.** Debemos elegir un buen titular para ayudar a contextualizar, enfocar y atraer la atención del lector. Pero, en paralelo, debemos tener en cuenta detalles como las etiquetas de los ejes, las unidades numéricas, los rangos, etc.
- 10) **Aportad interactividad siempre que sea posible.** La mayoría de las herramientas de visualización de datos permite crear un cierto nivel de interactividad con los datos. Si proporcionamos interactividad, permitiremos obtener más detalle en algunas informaciones, lo que redundará en facilitar la comprensión de los datos representados.

Si seguís estas pautas aumentaréis la eficacia de las visualizaciones de datos, permitiendo a la audiencia tomar decisiones más precisas sobre los datos recopilados y presentados.

Resumen

A lo largo de este material se ha explorado la visualización de la información como un proceso de destilación de conocimiento desde los datos originales. Una visualización explica historias, simplifica, mide, compara, explora, descubre, etc., proporcionando un acceso más sencillo a los datos subyacentes, al tiempo que permite obtener conocimiento de dichos datos mediante la presentación o exploración de esos datos de manera visual.

Bibliografía

Alcalde, I. (2015). *Visualización de la información de los datos al conocimiento*. Barcelona: Editorial UOC.

Cairo, A. (2012). *The Functional Art: An introduction to information graphics and visualization*. New Riders.

Evergreen, S. D. (2013). *Presenting data effectively: Communicating your findings for maximum impact*. SAGE Publications.

Few, Stephen (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press.

Krum, R. (2013). *Cool infographics: effective communication with data visualization and design*. John Wiley & Sons.

McCandless, D. (2009). *Information is beautiful*. Londres: Collins.

Meirelles, I. (2013). *Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations*. Rockport publishers.

Tufte, Edward (2001). *The Visual Display of Quantitative Information* (2.^a ed.). Graphics Press.

Ware, Colin (2008). *Visual Thinking*. Morgan Kaufmann.

Ware, Colin (2012). *Information Visualization* (3.^a ed.). Morgan Kaufmann.

Wickham, H. (2013). *ggplot2: elegant graphics for data analysis*. Springer Science & Business Media.

Yau, N. (2013). *Data points: visualization that means something*. John Wiley & Sons.

