

Taller visualización

Master Big Data

2016

PROFESOR/A
Carlos Garcia



Esta publicación está bajo licencia Creative Commons Reconocimiento, Nocomercial, Compartirigual, (by-nc-sa). Usted puede usar, copiar y difundir este documento o parte del mismo siempre y cuando se mencione su origen, no se use de forma comercial y no se modifique su licencia. Más información:

<http://creativecommons.org/licenses/by-nc-sa/3.0/>

1. Por qué usamos visualizaciones

Lo primero que tenemos que entender es que es un medio extraordinariamente potente para examinar, entender y transmitir información. Me gustaría llamar vuestra atención sobre la última parte de la frase, “transmitir información”, ya que esta es la clave y el gran objetivo, tener un mensaje y transmitirlo de manera clara y concisa a la audiencia objetivo. Pero no nos adelantemos.

Las visualizaciones nos permiten obtener toda la potencia de nuestro sistema de visión, lo que nos permite transmitir gran cantidad de información al cerebro.

Nos permiten aprovechar la capacidad de nuestro cerebro para construir e identificar patrones, relaciones y asociarlas con un significado.

Las visualizaciones, son en un gran número de ocasiones, el origen de nuevas preguntas que conllevan a realizar nuevos análisis o exploraciones en mayor profundidad. Dirigidas por un nuevo objetivo (en muchos casos nos permiten identificar nuevas áreas de exploración)

Las visualizaciones son excelentes para mostrar tendencias, cantidades, realizar comparaciones, ver o descubrir datos atípicos o de especial interés.

1.1. Objetivo de una visualización

El objetivo de toda visualización es transmitir información, es decir transformar los datos en información entendible para los lectores (audiencia). Para ello el creador de la visualización debe conocer cuál es el objetivo y las características de la audiencia, ya que la formación, si pertenecen al sector, nacionalización...etc., condicionan y afectan a la capacidad de decodificar el mensaje que se quiere transmitir.

El diseñador, debe elegir que datos debe mostrar, identificar sus características y encontrar las propiedades de visualización que mejor se ajustan al objetivo, transmitir un mensaje claro, fácil de decodificar y entendible sin unívocos.

2. Clasificación de las visualizaciones

2.1. Complejidad

Podemos clasificar la visualización en función del número de dimensiones que queremos representar en la visualización. De una forma más clara, el número diferente de tipos discretos de información que van a ser codificados de manera visual. Añadimos más claridad a este concepto, por ejemplo si tenemos un gráfico de líneas que muestra la evolución del precio de un producto a lo largo del tiempo, estaríamos hablando de dos dimensiones. Si mostramos diferentes productos hablaríamos de tres dimensiones, si añadimos el volumen de ventas diario de cada producto, tendríamos entonces 4 dimensiones (productos, precio, volumen ventas y tiempo).

Debemos tener siempre en mente que cuanto mayor sea la complejidad de nuestra visualización mayor es su dificultad, es más difícil encontrar la manera de transmitir la información de manera clara. Otro tema que quiero clarificar es que tener más datos, por ejemplo mostrar un periodo mayor de tiempo, no añade complejidad a la visualización, sino volumen, esto añade dificultad ya que nos impone el reto de desacoplar el volumen con la densidad de la información y la claridad en transmitir la información.

A la hora de afrontar una visualización, la complejidad nos plantea varios retos. El primero es sin duda la relación entre las dimensiones y las propiedades de visualización. Cuantas más dimensiones, más propiedades de visualización hay que utilizar, un punto a tener en cuenta es que cuando mostramos dimensiones, no las mostramos de forma aislada, esto no tendría mucho sentido. Mostramos de manera explícita o implícita relaciones entre ellas. El otro reto que se afronta cuando tenemos visualizaciones complejas, es la falta de convenciones, políticas o mejores prácticas que hagan más sencilla o guíen el proceso de generación de la visualización.

Tengamos en mente que las visualizaciones más efectivas son las que tienen 3 o 4 dimensiones, por encima comienza a ser un reto importante mantener la claridad y el mensaje.

2.2. Infografías o Visualizaciones de datos

Podemos hacer una primera diferenciación de infografía y visualización, las infografías son un tipo de visualización en la que la carga visual, creativa y manual es muy alta, mientras que la visualización puede verse más basada en método, análisis, suelen manejar más dimensiones son fácil de iterar y rehacer.

Infografías

Las podemos clasificar por las siguientes características:

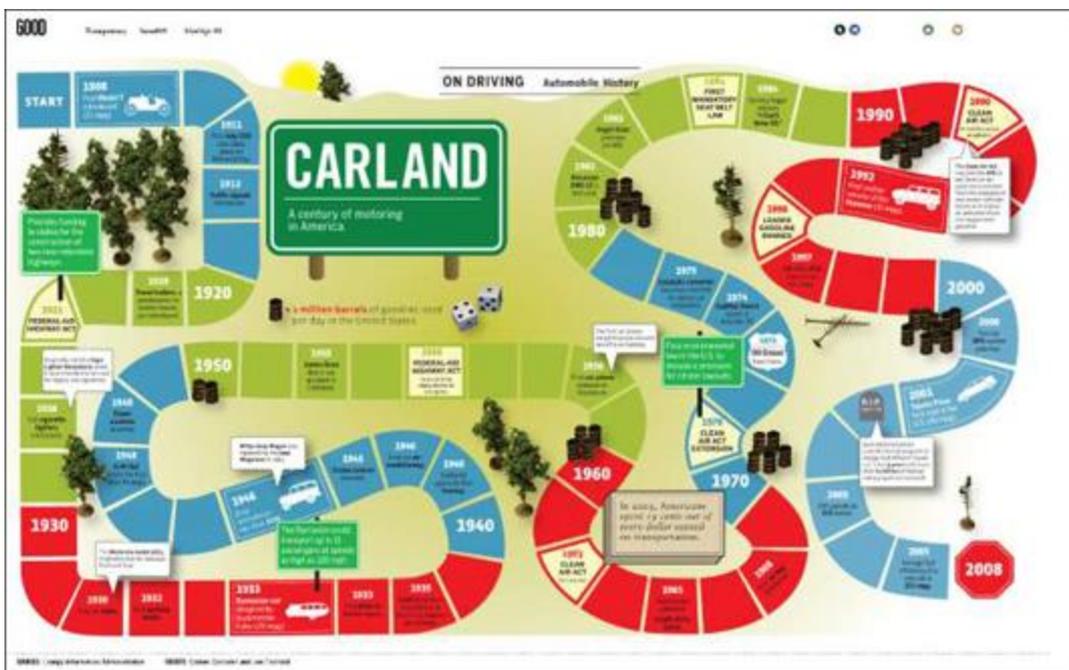
Suelen representar pocas dimensiones de datos. El motivo es que cada dimensión debe generarse la representación gráfica de manera manual.

Son específicas, no solo a los datos sino también a la audiencia y canal. Lo que hace difícil ser reutilizadas en otros escenarios.

Son generadas usando diferentes programas de diseño, ‘a mano’ o mejor de manera artesanal. Esto limita la cantidad de información que pueden transmitir (no es su propósito) y es complejo el proceso de actualización (manual)

Estéticamente muy potentes

La siguiente infografía fue creada por Coleen Corcoran y Joe Prichard, está accesible en el link



Visualizaciones.

Las podemos clasificar por las siguientes características:

Son generadas de manera algorítmica, aunque siempre o casi siempre hay retoques manuales para reforzar el mensaje.

Son fáciles de re-generar con diferentes datos o añadir nuevas dimensiones (nos basamos en algoritmos).

Estéticamente limitadas, las infografías son mucho más ricas

Nos permiten transmitir y tratar un mayor número de dimensiones.

2.3 Visualizaciones: Exploración Vs Explicación

Tenemos dos tipos de visualizaciones: Visualizaciones de exploración y visualizaciones de explicación. Es importante identificar el tipo de visualización que estamos haciendo ya que cada una de ellas tiene un propósito diferente y la forma de afrontar cada una es diferente.

Exploración

Las visualizaciones de exploración son adecuadas cuando tenemos un conjunto de datos y no sabemos qué son o qué dicen. Necesitamos analizar los datos para poder tener una idea de qué es lo que hay en los datos. Por eso realizamos una investigación usando herramientas visuales para identificar las características relevantes en los datos, distribuciones, tendencias, líneas, puntos anómalos, puntos de interés.

Exploración es típicamente parte de la fase de análisis, cuando tenemos datos en los que existe posiblemente una gran parte de ruido, y se hace a un nivel de granularidad alto.

Es la fase en la queremos saber que nos pueden decir los datos.

.

Explicación

Las visualizaciones explicativas, son adecuadas cuando conocemos lo que nos pueden decir los datos, los conocemos y si queremos verlo de esta manera, tenemos el dataSet normalizado y entendemos sus relaciones. En este punto de partida nuestro objetivo es comunicar que lo que los datos nos dicen a la audiencia. Podemos decir que las visualizaciones explicativas son la parte de comunicación y explicación de los datos frente a la de exploración que es conocer los datos.

Este tipo de visualización se enfoca en identificar los datos que van a soportar el mensaje y como codificar visualmente la información para:

Que comunique claramente lo que queremos decir

Sea fácilmente entendible para la audiencia.

3. Claves en el proceso

3.1 Objetivo

Toda actividad de visualización tiene un objetivo, que debe ser claro, el objetivo va a definir muchas de las decisiones de diseño y es el motivo por que el decidamos cual es la mejor forma de mostrar los datos, gráficos a utilizar etc.

Es importante comenzar por el principio, ¿estamos ante una visualización exploratoria o explicativa?, esto nos va a facilitar poner el foco en lo importante, estamos en la fase de exploración o estamos en la fase de explicación. Si estamos en la fase de explicación, identificamos qué tenemos que comunicar para cumplir el objetivo, es importante comentar el tipo de mensaje que podemos o queremos dar, queremos presentar la información o queremos mandar un mensaje persuasivo.

3.2 Audiencia

Una vez tenemos claro el objetivo, el tipo y el mensaje que queremos dar es el momento de pensar en la audiencia, en aquellos que van a ver e interpretar la información.

¿Qué información van a poder aprender de esta visualización?

¿Qué mensaje o mensajes quieres transmitir?, es fundamental que los mensajes estén alineados con los objetivos. Debes tener el objetivo mente, es las visualizaciones es muy sencillo ‘sobreactuar’ es decir mostrar más datos de los necesarios o usar formatos complejos, esto lo único que va a traer como consecuencia es que el mensaje que quieres comunicar sea más difícil de ser entendido por la audiencia.

Ten en mente el principio de simplicidad, hacer las cosas de una manera lo más simple posible (esto no implica no hacer bien las cosas, sino todo lo contrario), el transmitir el mensaje de la manera más clara posible. El aplicar el conocimiento y la experiencia que tienes como experto de visualización para transmitir el mensaje adecuado. Es muy típico que las personas con menos experiencia tiendan a realizar visualizaciones más complejas, es debido en parte a la inseguridad y a las ganas de demostrar capacidades. Esto es un error en el que es muy sencillo caer.

Un ejercicio muy recomendable es ponerte en la situación de ser el lector de la visualización, dentro de este ejercicio es muy importante que intentemos contestar a la siguiente pregunta: ¿Qué información debo transmitir para que el lector obtenga la información que necesita (objetivo)?, ¿Qué condicionantes o características tiene (experiencia, posición, sector)?

Debes tener en cuenta el contexto, es decir, cuando está codificando la información tienes un gran conocimiento de los datos y manejas con fluidez las propiedades visuales. Esto hace que tiendas a asumir algunas cosas que no son transmitidas en la información. Sin embargo el contexto y condicionantes del lector no son las mismas que las tuyas, y por supuesto va a introducir una cierta distorsión en la interpretación. Una manera de evitar en parte este problema es mediante el uso de buenas prácticas y si la hubiese de la política corporativa de reglas de visualización (bien sea formal o informal, instaurada por los años). Debes intentar que tu contexto no condicione la visualización ya que este no se transmite y no debes darlo por implícito.

Otro tema importante a tener en cuenta es la cantidad de información, el detalle que necesita, esto condiciona la visualización, si necesitamos un nivel de detalle importante debemos pensar en incluir tablas en las partes claves.

3.3 Los datos

Otro tema importante a tener en cuenta es la cantidad de información, el detalle que necesita, esto condiciona la visualización, si necesitamos un nivel de detalle importante debemos pensar en incluir tablas en las partes claves.

El tercer condicionante de una visualización son los datos. Una buena visualización es aquella que cumple con el objetivo y revela aquello que encierran los datos.

Debemos entender los datos para definir la mejor visualización. Esto implica que debemos iterar conforme nuestro conocimiento de los datos aumenta. Comenzamos siempre por las técnicas básicas de representación y la vamos evolucionando conforme descubrimos la naturaleza de los datos que tenemos.

Debemos llegar a conocer las siguientes características, que nos guiarán en el proceso:

Estamos manejando series temporales o una jerarquía de datos

Que dimensiones existen, ¿cuáles son las importantes y relevantes para nuestro objetivo?

Como se relacionan los datos, y más concretamente las dimensiones relevantes

Que tipos manejamos, con categóricos, discretos, continuos, son lineales no lineales, cuáles son sus límites superiores e inferiores

4. Datos y sus tipos. Visualizaciones

Una vez que comprendemos los datos, pasamos a prepararnos para definir las visualizaciones adecuadas eligiendo las propiedades de visualización correctas. Hay dos factores de las propiedades de visualización que son claves, si la propiedad de visualización tiene orden natural y cuantos valores diferentes de esta propiedad puede interpretar correctamente el lector.

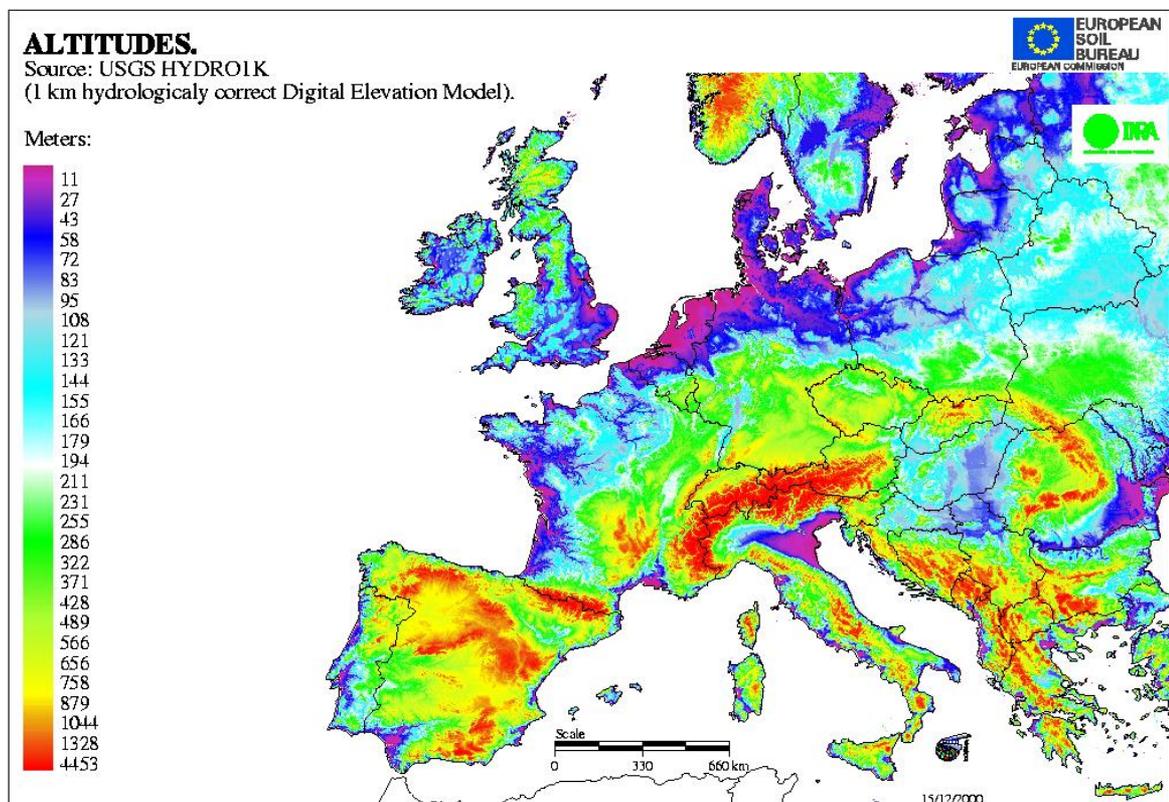
4.1 Orden Natural

Lo que determina si una propiedad visual tiene orden natural es nuestro cerebro, ejemplo de propiedades visuales que tienen esta propiedad; posición (ejes), longitud, el tamaño y grosor de las líneas. Propiedades que no tienen orden natural; color, formas, el tipo de las líneas

El color no tiene orden natural, no es identificado de esta forma por el cerebro, brillo y saturación sí (esto es lo que nos lleva a pensar que los colores tienen orden natural). Hay una fuerte convención social sobre el significado de los colores, no caigas en el error de utilizar colores para indicar orden. Si quieres usar dentro de un mapa o un 'heatmap' los colores para indicar orden usa

las propiedades de brillo y saturación (puedes jugar con estas propiedades que varíen a lo largo de los ejes o dentro del mapa) pero no intentes indicar orden (por ejemplo si el violeta es el siguiente color al amarillo, o el verde al rojo no es algo establecido y dependerá de la interpretación de cada lector).

Pongamos el siguiente ejemplo en el que se utiliza una leyenda de arcoíris para identificar alturas. Es realmente difícil de interpretar, primero por la complejidad y los rangos de los colores. Segundo porque al no tener orden natural muchos esperaríamos que la escala fuese de rojo a verde por convención social y cultural

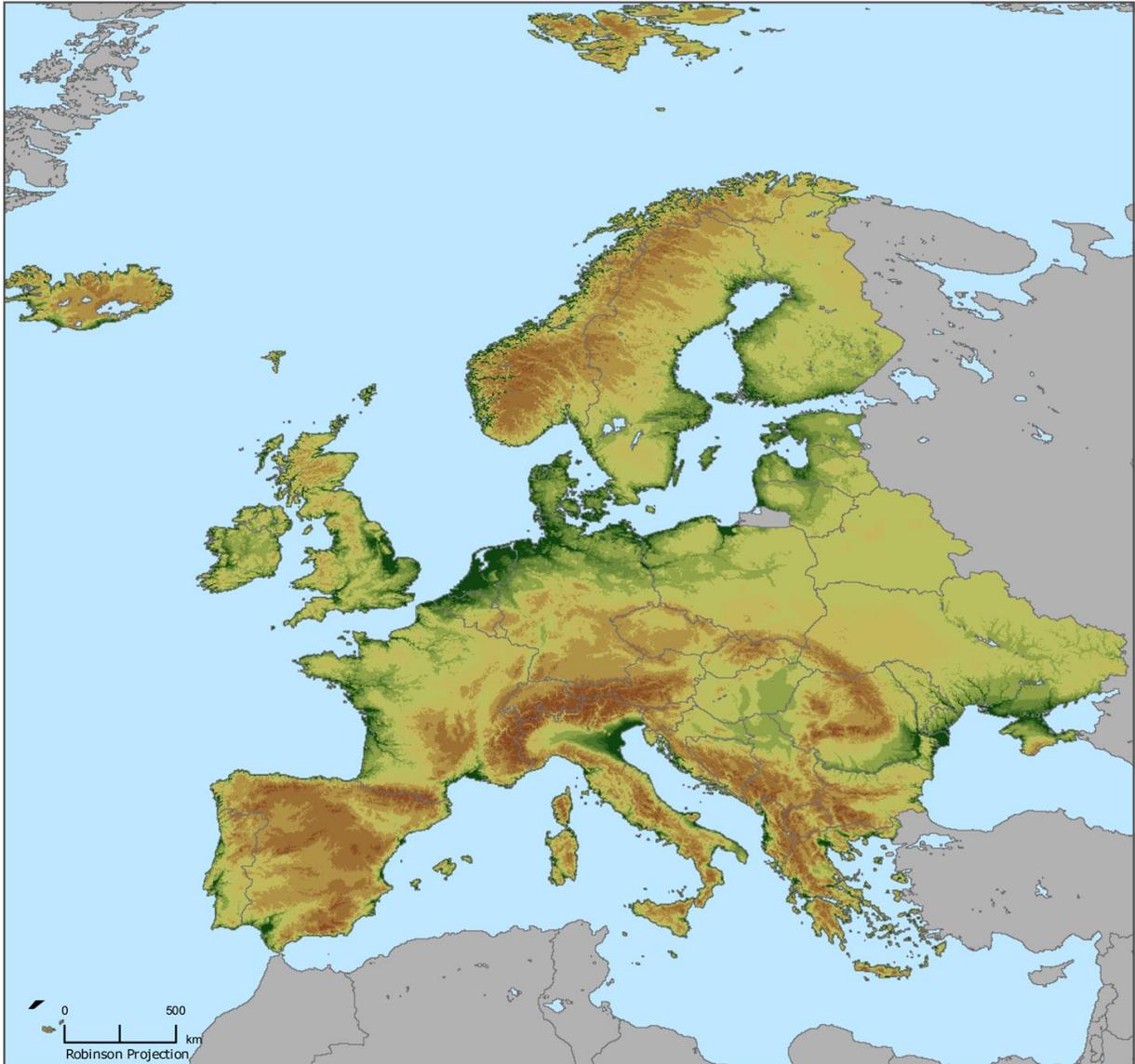


European Soil Portal. European Union. <http://eusoils.jrc.ec.europa.eu/>

Si queremos usar el color para identificar orden, debemos tener colores que se diferencien claramente e un numero de intervalos que no superen los 15 rangos, esto nos ayudará a decodificar la información.

Elevation Zones

Europe



Digital elevation data (meters above mean sea level) were obtained as a 1 kilometer resolution elevation\bathymetry raster product from ISciences, L.L.C. Elevation zones were created by aggregating ranges of land elevation values into 12 thematic elevation classes. The 2004 ISciences data were resampled from their native 30 arc-second resolution to match GPW's population and land area 2.5 arc-minute spatial footprint. Source: ISciences, L.L.C. 300 N. Fifth Ave. Suite 120., Ann Arbor, MI 48104 <http://www.isciences.com/>.

Elevation

- | | |
|---------------|-----------------|
| < 5 m | 200 - 399.9 m |
| 5 - 9.9 m | 400 - 799.9 m |
| 10 - 24.9 m | 800 - 1499.9 m |
| 25 - 49.9 m | 1500 - 2999.9 m |
| 50 - 99.9 m | 3000 - 4999.9 m |
| 100 - 199.9 m | >= 5000 m |



Copyright 2007. The Trustees of Columbia University in the City of New York. Source: Center for International Earth Science Information Network (CIESIN), Columbia University, Population, Landscape, and Climate Estimates (PLACE). Further information available at: <http://sedac.ciesin.columbia.edu/place/>

Publish Date: 03/13/07



This document is licensed under a Creative Commons 2.5 Attribution License <http://creativecommons.org/licenses/by/2.5/>

4.2 Número de valores distintos

Lo primero que vamos a identificar son los tipos de datos que podemos manejar.

Variable cualitativa Las variables cualitativas se refieren a características o cualidades que no pueden ser medidas con números.

Categorica o cualitativa nominal presenta modalidades no numéricas que no admiten un criterio de orden. Por ejemplo: El estado civil, con las siguientes modalidades: soltero, casado, separado, divorciado y viudo.

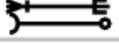
Variable cualitativa ordinal presenta modalidades no numéricas, en las que existe un orden. Por ejemplo: La nota en un examen: suspenso, aprobado, notable, sobresaliente. Puesto conseguido en una prueba deportiva: 1º, 2º, 3º, ...

Variable cuantitativa Una variable cuantitativa es la que se expresa mediante un número, por tanto se pueden realizar operaciones aritméticas con ella

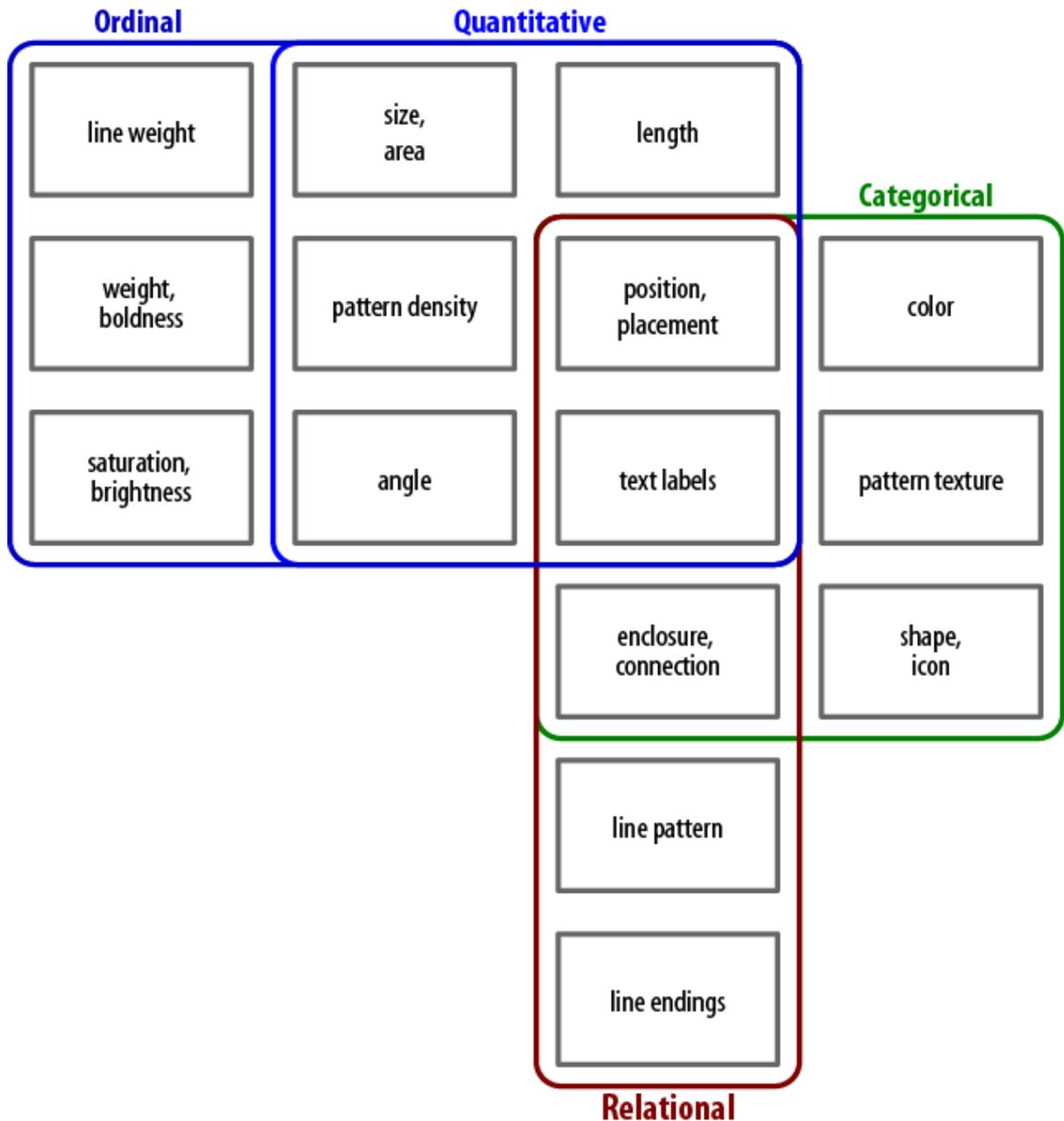
Variable discreta Una variable discreta es aquella que toma valores aislados, es decir no admite valores intermedios entre dos valores específicos. Por ejemplo: El número de hermanos de 5 amigos: 2, 1, 0, 1, 3.

Variable continua Una variable continua es aquella que puede tomar valores comprendidos entre dos números. Por ejemplo: La altura de los 5 amigos: 1.73, 1.82, 1.77, 1.69, 1.75.

La siguiente tabla sirve para poder seleccionar la propiedad visual más apropiada a utilizar en función del tipo de dato.

Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	yes	infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium	Good	Good		
	pattern density	yes	few	Good	Good		
	weight, boldness	yes	few		Good		
	saturation, brightness	yes	few		Good		
	color	no	few (<20)			Good	
	shape, icon	no	medium			Good	
	pattern texture	no	medium			Good	
	enclosure, connection	no	infinite			Good	Good
	line pattern	no	few				Good
	line endings	no	few				Good
	line weight	yes	few		Good		

La siguiente tabla sirve para poder seleccionar la propiedad visual más apropiada tomando como partida el tipo de dato que tenemos, es otra forma más resumida de la misma tabla, 'indexada' por tipo de dato.



4.3 Condicionamientos adicionales

Títulos y etiquetas

Una vez identificada la audiencia definimos el vocabulario a utilizar; Están familiarizados con la jerga del negocio, pertenecen al sector, en qué áreas trabajan. Una vez aleccionado el lenguaje adecuado, podemos elegir si mayúsculas o minúsculas. Esta decisión puede parecer que no afecta a la legibilidad de la visualización, pero sí que afecta.

Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn't mttar in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe.

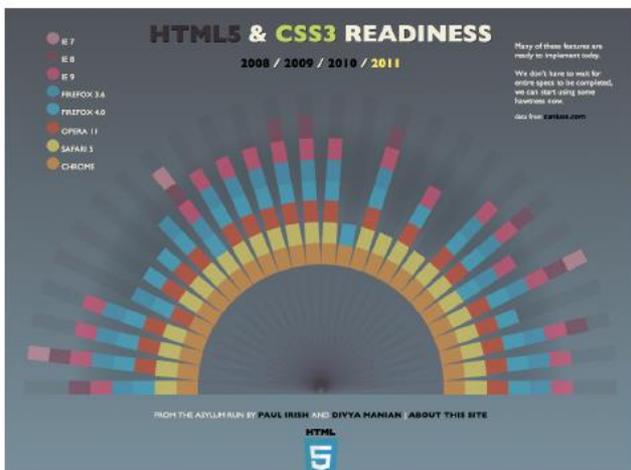
Puedes leer el documento entero en el siguiente link

Esta propiedad no se mantiene cuando todo el texto es en mayúsculas (te reto a que lo compruebes).

Evita que el vocabulario sea una barrera para comprender la visualización

Colores

Primera advertencia, aproximadamente el 7% de la población tiene problemas para ver correctamente los colores. Identifica que hay alguna connotación cultural o de contexto, hay culturas en los que los colores se asocian con suerte, connotaciones negativas etc. También debes tener en cuenta que los colores que los colores asociados a formas tienen un fuerte significado (stop por ejemplo), también debes considerar si hay alguna relación de la información que estas presentando con determinados colores. Miremos el siguiente ejemplo, el no utilizar los colores que identifican cada producto hace que el grafico sea realmente difícil de interpretar



Browser	Existing color	Better color

La regla de oro

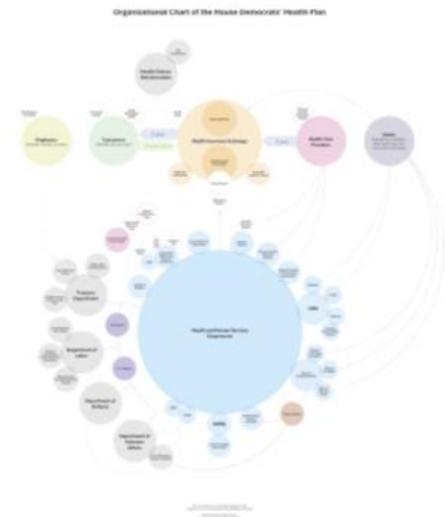
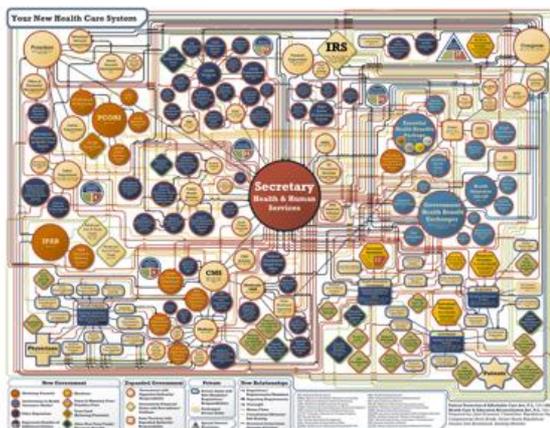
Principio KISS (Keep it simple stupid) , cuanto más sencilla sea una visualización que transmita el mensaje que se quiere transmitir y cumpla con el objetivo, más dice a favor de su autor. Hacer las cosas simples es mucho más complejo que no hacerlas. Pensemos un poco sobre lo que implica hacer las cosas simples:

Tenemos que tener un conocimiento exhaustivo del problema y los datos

Requiere capacidad de sintaxis

Requiere como representar la información sintetizada y no de la manera que muestra el problema.

Comprueba las siguientes dos visualizaciones, aunque parezca mentira y sin entrar en detalle ambas muestran el plan de salud de estados unidos en el 2010. Sin entrar a valorar la intención de ambas visualizaciones (la primera fue realizada por el partido que no gobernaba) podemos apreciar la diferencia, la primera es imposible de seguir o entender (salvo que es muy complejo), mientras la segunda sí que nos permite entender el sistema.

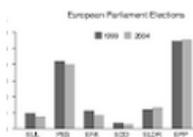


5. Visualizaciones y su uso

En esta sección vamos a explorar las diferentes alternativas de visualización en función de los tipos de datos y el objetivo de la visualización, para poder asegurar una visualización eficaz.

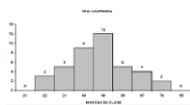
5.1 Formatos cuantitativos y de comparación

Gráficos de barras.

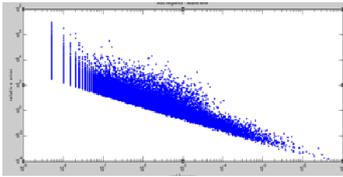


Excelentes para comparar valores dentro de una categoría o entre categorías. Recomendados para datos discretos, para datos continuos evalúa el uso de gráficos de líneas. Suelen representar valores acumulados entre diferentes categorías, lo que puede dificultar su interpretación, por lo que debes mostrar en la comparación las características más importantes, siempre en el mismo orden.

Histogramas.



Son un tipo especializado de gráfico de barras, su propósito es mostrar la distribución de los valores entre los posibles valores, para entender la distribución de los valores. El área representa la suma total de las instancias de cada valor. Nos sirve para entender la distribución



Son útiles para ver la correlación entre dos dimensiones cualitativas, o para representar datos que varían en función de dos dimensiones. Si el número de dimensiones es mayor a dos, consideremos los gráficos de barras o tarta.

Tablas.

K	L	M	N	O	P	Q	R
Country	Postal Cod	Market	Region	Product ID	Category	Sub-Categ	Product Ná
United States	10024	US	East	TEC-AC-1000	Technology	Accessories	Plantronics C
Australia		APAC	Oceania	FUR-CH-1000	Furniture	Chairs	Novimex Exe
Australia		APAC	Oceania	TEC-PH-1000	Technology	Phones	Nokia Smart
Germany		EU	Central	TEC-PH-1000	Technology	Phones	Motorola Sm
Senegal		Africa	Africa	TEC-SHA-100	Technoloev	Coolers	Sharo Wirele

Tablas de texto si bien no son una representación gráfica como tal las utilizamos cuando se requiere precisión en los datos.

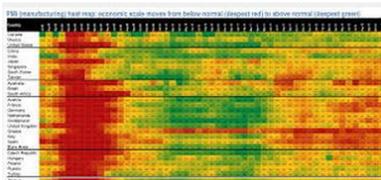
TreeMaps.



Son excelentes para mostrar jerarquías y valores, cada elemento de la visualización se representa con un color, celda tiene un tamaño y un color que representan atributos.

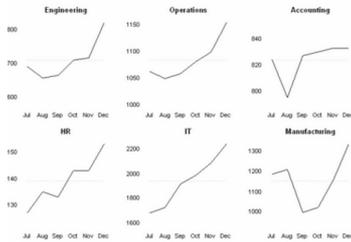
En el ejemplo se muestra el S&P; el tamaño de cada celda es relativo a la capitalización bursátil de la empresa y el color indica el % de variación del precio de la acción, agrupados por sectores.

HeatMaps



Son gráficos con dos ejes en la que los colores representan variaciones en los valores o rangos de valores. Se suelen usar como código de colores rojo - verde o rojo-amarillo-verde.

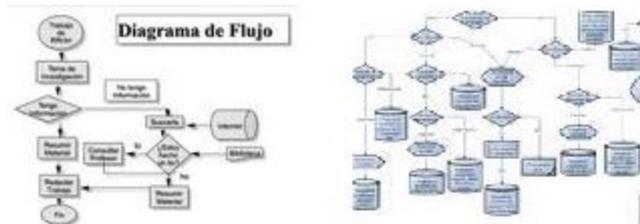
Array de gráficos



Son un conjunto de gráficos, en el que cada uno representa una parte pequeña de los datos. Se suelen utilizar como resumen o para poder hacer una comparación sencilla y visual de tendencias, captando así el foco en una de las gráficas para seguir con la visualización.

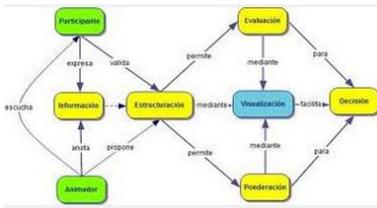
5.2 Formatos relacionales

Diagramas de flujos de datos, Entidad relación.



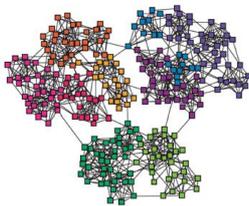
Existen un gran número de formatos disponibles para representar datos relacionados, los más comunes son los diagramas de flujos y entidad relación. Los usamos cuando tenemos datos con jerarquías relaciones ente ellos. Se pueden usar también cuando los datos tienen una fuerte relación entre ellos y queremos destacarla o mostrarla.

Mapas de decisión y gráficos de flujos.



Son un tipo específico de diagramas relacionales que muestran el proceso de decisión con las diferentes alternativas. Los nodos representan acciones o estados y se suelen codificar usando colores y formas.

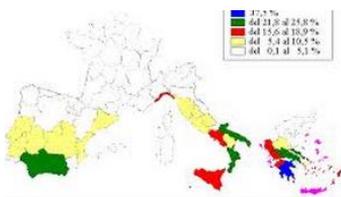
Gráficos de redes sociales.



Son una especialización de los gráficos de ‘grafos’ se utilizan para visualizar conexiones entre personas, compañías etc. Las aristas muestran la relación que estamos analizando. Podemos tener diferentes tipos de nodos y diferentes tipos de relaciones en la misma visualización.

5.3 Formatos geográficos

Mapas geográficos.



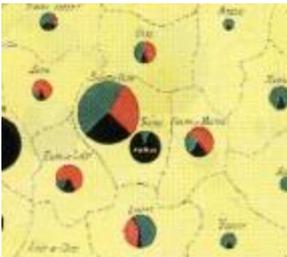
Son una herramienta muy potente visualmente, usamos la propiedad de color para identificar regiones (seleccionarlas o resaltarlas), podemos incluir líneas que indiquen relaciones entre las diferentes áreas o conexiones. Podemos añadir más dimensiones mediante el uso de ‘burbujas o círculos’. Podemos usar el área de las regiones para identificar medias o ponderar resultado. En este caso es importante que el lector identifique la metáfora o técnica que estamos usando.

Mapas geográficos burbujas.

Cuando queremos añadir una tercera dimensión a los datos representados en un mapa, la primera opción que debemos evaluar es la de añadir círculos (o burbujas). También podemos usar esta técnica como opción de redundancia, cuando queremos enfatizar los datos representados mediante colores para añadir precisión, facilitar su comparación o resaltar su importancia.



Cuando querremos añadir una cuarta dimensión, reemplazamos los círculos por gráficos de tarta.



Cuando incluyamos una cuarta dimensión, es importante que tengamos en mente la legibilidad del gráfico. Si intentamos representar mucha información en un único gráfico es más que probable que la complejidad de decodificación sea tan alta que el lector se pierda o lo encuentre demasiado complejo.

6. Propiedades visuales

En este apartado vamos a analizar las propiedades de visualización para su uso óptimo en las visualizaciones, una vez conocemos los datos y las diferentes formas de representación, vamos a comentar las propiedades de visualización más comunes, ventajas, como maximizar el beneficio de su uso y los errores más comunes.

6.1 Color

El color no tiene orden natural, y este es el error más común en su uso. Tendemos a pensar que es así por las fuertes connotaciones sociales que tiene y su amplio uso, pero es un error. Tenemos que evitar caer en este error y usar brillo y saturación para denotar orden o rankings.

Color, es la propiedad adecuada para mostrar categorías, o categorías no ordenadas (genero, regiones, sistemas operativos, aplicaciones...).

Una regla que debemos tener en mente es el número diferente de colores que debemos usar en una visualización, un máximo razonable es 12 y no exceder nunca de los 20 (20 categorías diferentes en una única visualización es probablemente demasiado). Para una comunicación efectiva, una buena recomendación son 6 colores.

La siguiente tabla, muestra el uso preferente de colores:

1	Red	
2	Green	
3	Yellow	
4	Blue	
5	Black	
6	White	
7	Pink	
8	Cyan	
9	Gray	
10	Orange	
11	Brown	
12	Purple	

Cuando utilices los colores en una visualización, debes tener en mente el objetivo y que el lector va a buscar patrones, si la visualización se compone de varias visualizaciones debes ser coherente y consistente en los patrones que utilizas en todo el ejercicio.

Los colores, si bien no tienen orden, sí que tienen asociaciones de significado debido a convenciones sociales, emociones y estética, enumeramos algunas:

Rojo: Se asocia con peligro y precaución. También se asocia con pasión, sangre. En Asia se asocia con prosperidad y buena suerte.

Verde: Se asocia con permiso (especialmente combinado con rojo), naturaleza, ecologismo y religión.

Amarillo: Se asocia con precaución, también con felicidad. Su uso extensivo puede resultar irritante.

Azul: Se asocia con calma, agua dependiendo de la forma se puede asociar con policía, ejercito.

Negro: Se asocia con muerte o malo, pero también con los conceptos de lujo y sofisticación

Rosa: Se asocia con afecto, imaginación, niños. El rosa claro tiene una fuerte asociación con niñas y el azul claro con niños pequeños.

Gris: Se asocia con madurez, modestia, conservador y neutralidad.

Naranja: Se asocia con emergencia, fuego, energía, en Asia con espiritualidad. En contexto de nutrición se asocia con salud y vigor.

Marón: Se asocia con tierra, elementos como cuero o excrementos.

Violeta: Se asocia con nobleza, realeza, magia pero también con falsedad y artificialidad.

6.2 Tamaño

Es una propiedad muy buena para indicar la importancia relativa dentro de las entidades. Si quieres captar la atención sobre una entidad puedes usar el tamaño como forma de capturar la atención sobre ella. De esta manera aunque el tamaño no sea indicativo del tamaño real (puedes por ejemplo representar esta entidad igual en tamaño que la mayor) es una forma muy práctica de capturar la atención del lector y ayuda a que ponga foco.

Un error común a la hora de usar el tamaño es hacer el valor como diámetro del círculo, esto nos lleva que las áreas no sean comparables en realidad, porque no guardan la proporcionalidad de las medidas. No quiero decir que no lo usemos si queremos resaltar las diferencias, solo que seamos conscientes.

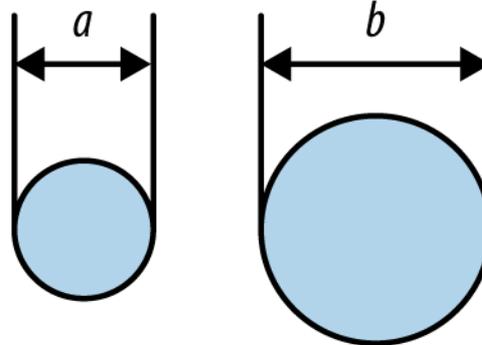
Values to encode:

a

b

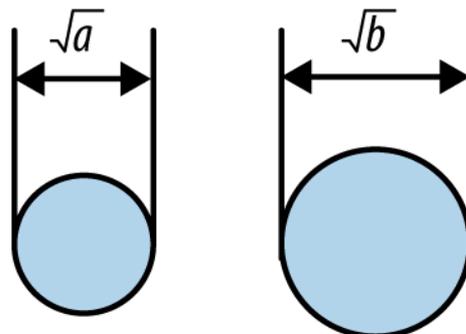
Wrong:
diameters proportional
to values

areas proportional
to values²



Right:
diameters proportional
to square root of values

areas proportional
to values

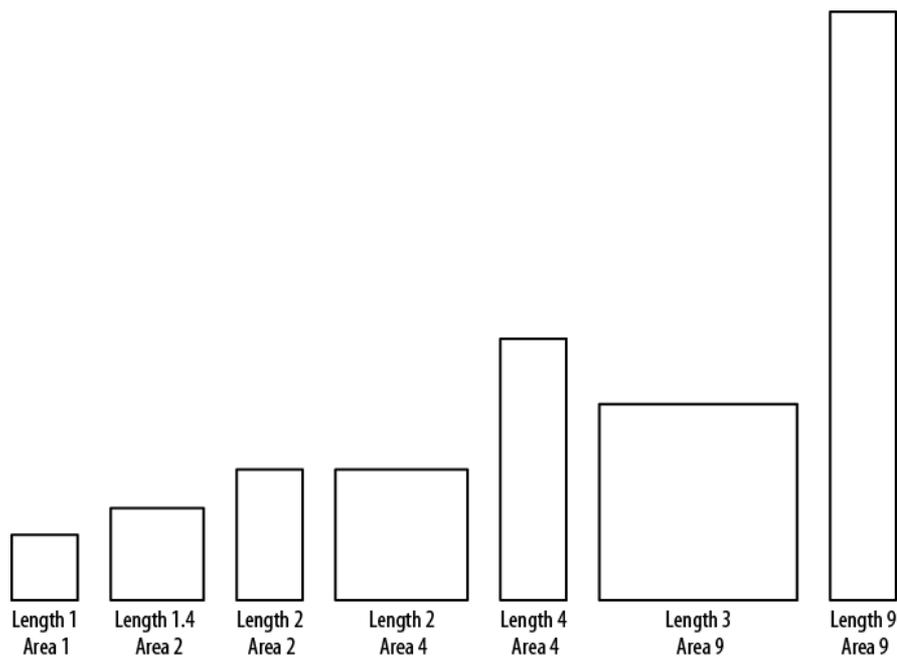


Un error común a la hora de usar el tamaño es hacer el valor como diámetro del círculo, esto nos lleva que las áreas no sean comparables en realidad, porque no guardan la proporcionalidad de las medidas. No quiero decir que no lo usemos si queremos resaltar las diferencias, solo que seamos conscientes.

6.3 Comparaciones y tamaño

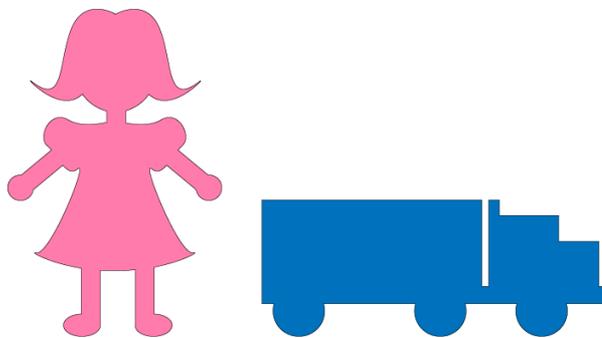
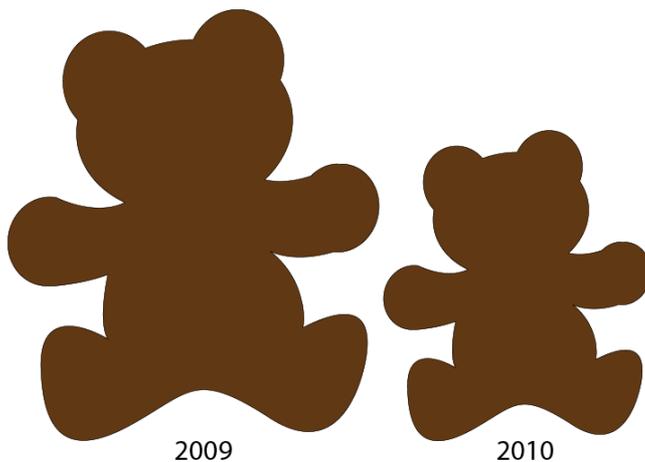
Nuestro cerebro es muy bueno realizando comparaciones de áreas en rectángulos, cuando la única dimensión que cambia es la longitud.

Esta propiedad de nuestro cerebro no se mantiene cuando varían más dimensiones, por ejemplo longitud y ancho, en este caso tendemos a subestimar las diferencias en las áreas.



La recomendación para evitar malas interpretaciones, es usar rectángulos variando la longitud. Si necesitamos añadir información adicional, es importante evaluar el uso de etiquetas para mantener el patrón de comportamiento en la visualización y no mezclar diferentes figuras.

Otro aspecto importante es el uso de figuras no geométricas, en este caso nuestro cerebro no es bueno analizando y comparando áreas. Por supuesto es capaz de identificar qué figura es mayor, pero no es capaz de tener una visión buena del número de veces que una figura es mayor que la otra, quedando a interpretación, casi libre.



Otro aspecto importante es el uso de figuras no geométricas, en este caso nuestro cerebro no es bueno analizando y comparando áreas. Por supuesto es capaz de identificar qué figura es mayor, pero no es capaz de tener una visión buena del número de veces que una figura es mayor que la otra, quedando a interpretación, casi libre.

6.4 Texto

Recuerda la capacidad del cerebro para identificar palabras basándose en la primera y última letra, reconocimiento de patrones. Esto no ocurre cuando usamos mayúsculas, tenlo en cuenta.

Si quieres llamar la atención, puedes usar mayúsculas en las etiquetas, debes valorar que será más difícil de leer, pero por otro lado tendrás el foco debido a esta complejidad añadida.

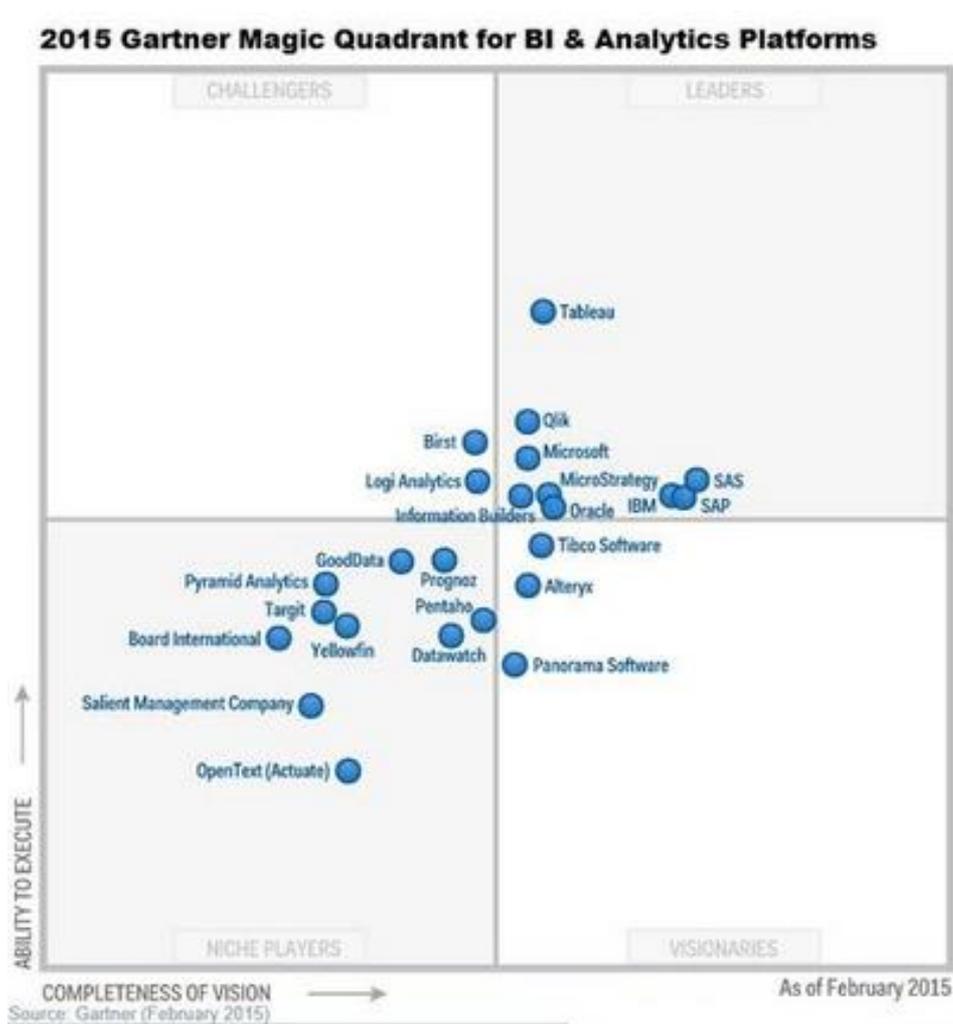
No uses sombras en el texto, esta propiedad no ayuda sino todo lo contrario, dificulta y es percibida como un obstáculo a la comunicación.

7. Tableau

En este apartado vamos a introducir Tableau como una de las herramientas más potentes y sencillas de utilizar en el contexto de Business Analytics.

Tiene su origen en Stanford, en el departamento de CS, área de defensa proyecto para explorar datos en bases de datos relacionales y cubos. Chris Stolte & Pat Hanrahan desarrollaron VizQL, que unía las disciplinas de manejo de datos y técnicas de visualización mediante “drag & Drop”

En 2003, nace como compañía y en los últimos años es señalado por Gartner como el líder en el área de BI & Analytics



7.1 Productos

Tiene varios productos, que en esencia son diferentes versiones del software que añaden más o menos funcionalidad en función del esquema de licencias.

Tableau Desktop

Aplicación visual basada en ventanas, tiene dos ediciones pública y profesional. La versión pública permite conectarse a archivos y grabar localmente las visualizaciones. Mientras que la versión profesional permite conectarse a una gran variedad de fuentes de datos, permite guardar las visualizaciones en un servidor, o publicarlas mediante Tableau online (servidores públicos o servidores Tableau). Es la versión recomendada para realizar analítica de negocios.

Tableau Server

Son para empresa y con un uso extensivo de la herramienta, permiten control de permisos y accesos a una gran granularidad a nivel de usuarios, visualizaciones. Tiene un interfaz web que permite centralizar accesos, proyectos e implementaciones.

Tableau online

En un servicio de hosting para compartir y almacenar visualizaciones, requiere que tengas la versión desktop profesional.

Tableau public

La mejor opción para todos aquellos que hagan un uso ocasional o usen como fuente de datos archivos, es gratuita y la puedes utilizar desde la nube o bajarla como aplicación. Permite compartir visualizaciones a través del servicio de hosting, en la versión 9 también tienes control de acceso

Tableau Reader

Es una aplicación gratuita que permite abrir archivos de trabajo de Tableau (*.twbx) puedes iterar con la visualización y realizar cambios. No permite realizar nuevas visualizaciones.

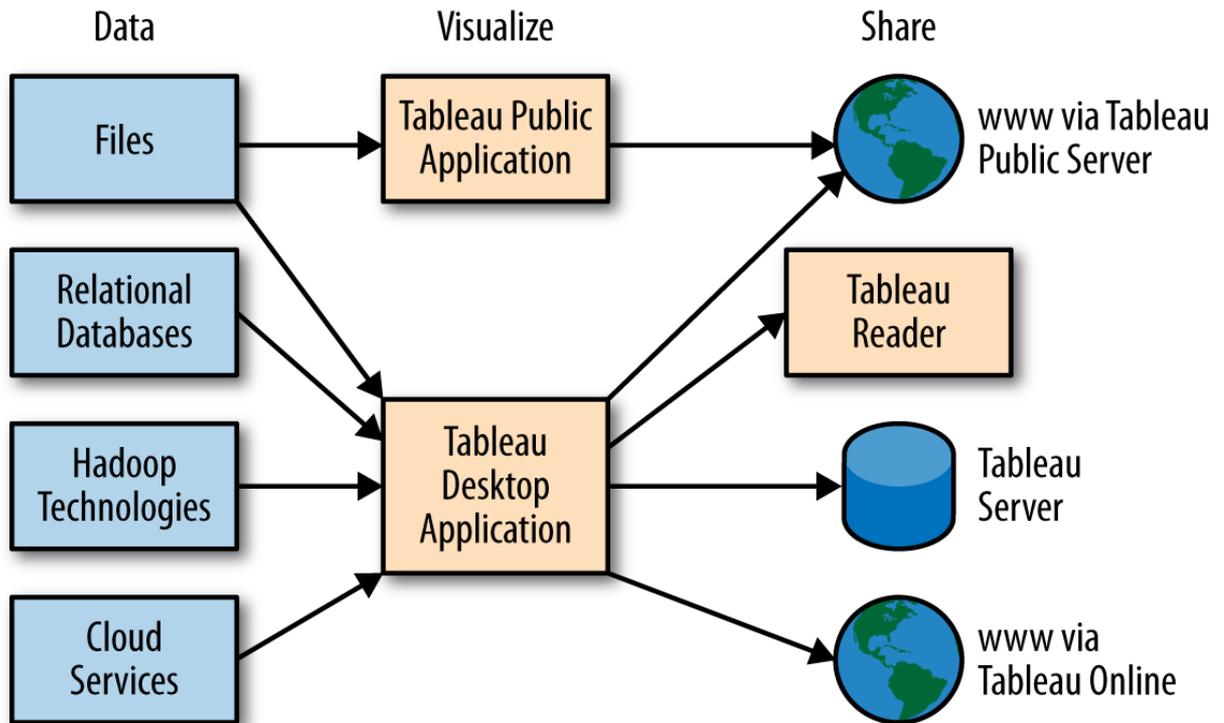


Tableau Reader

Es una aplicación gratuita que permite abrir archivos de trabajo de Tableau (*.twbx) puedes iterar con la visualización y realizar cambios. No permite realizar nuevas visualizaciones.

7.2 Interfaz de usuario

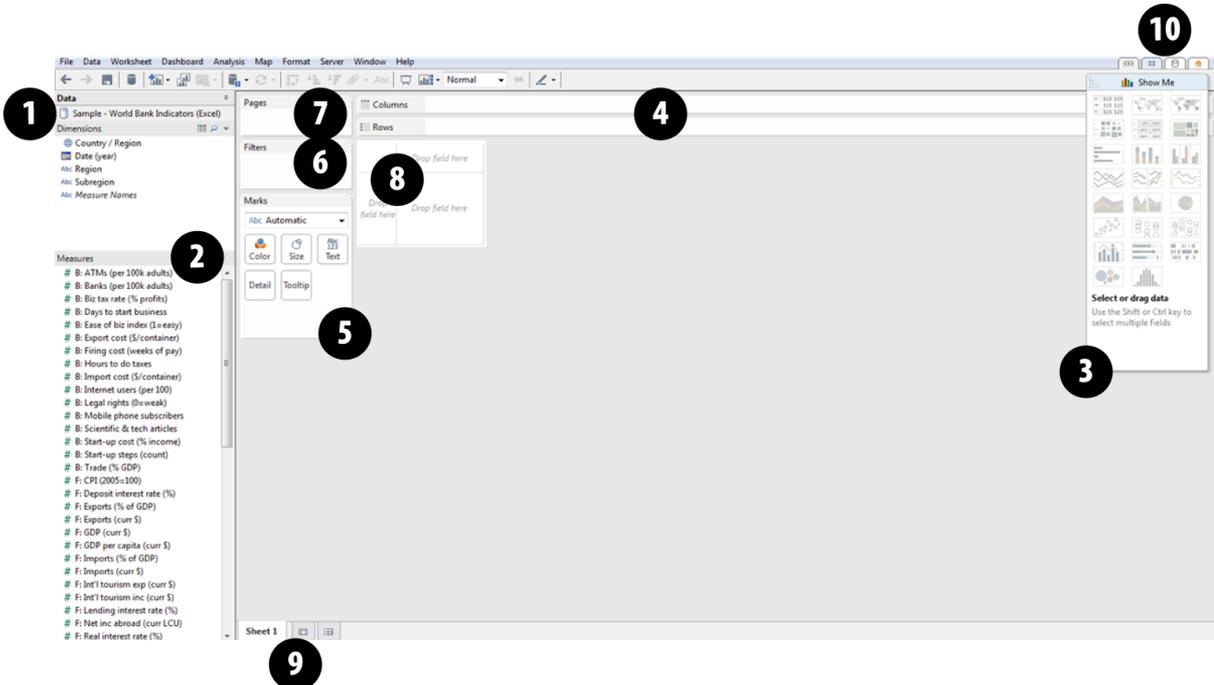
Tableau se basa como Excel en libros de trabajo (workbooks) los que pueden contener:

- Hojas (Sheet), son las visualizaciones. Cada hoja contiene una visualización individual
- Cuadros de mando (Dashboards). Contienen una combinación de visualizaciones y otros elementos, como etiquetas, imágenes o links que nos permiten enriquecer la información. Nos permite interactuar entre las diferentes visualizaciones del cuadro de mandos de manera interactiva.
- Historias (Stories), son una secuencia de visualizaciones que nos permiten definir una secuencia de visualización, de esta manera podemos asegurar que el lector siga el desarrollo deductivo de nuestra presentación, haciendo más sencillo codificar y enfatizar los mensajes.

Al abrir Tableau vemos la siguiente pantalla, que nos permite fácilmente iniciar el trabajo.



Cuando abrimos una fuente de datos, o una archivo Tableau nos encontramos con el siguiente interfaz de usuario.



- 1.- Lista de fuentes de datos, pueden ser más de una.
- 2.- Dimensiones y Medidas. Campos disponibles en las fuentes de datos, para realizar la visualización.
- 3.- Muestrame o ShowMe. Nos muestra las recomendaciones de visualización para los datos seleccionados.
- 4.- Tarjeta para las columnas y las filas. Nos muestra que se está mostrando en cada eje, así como la función de agregación que se está utilizando
- 5.- Tarjeta de máscaras. Nos proporciona el control de la visualización, color, tamaño, etiquetas, etiquetas emergentes, texto y formas.

6.- Tarjeta de filtros. Filtros que estamos aplicando en la visualización, pueden ser dimensiones o medidas.

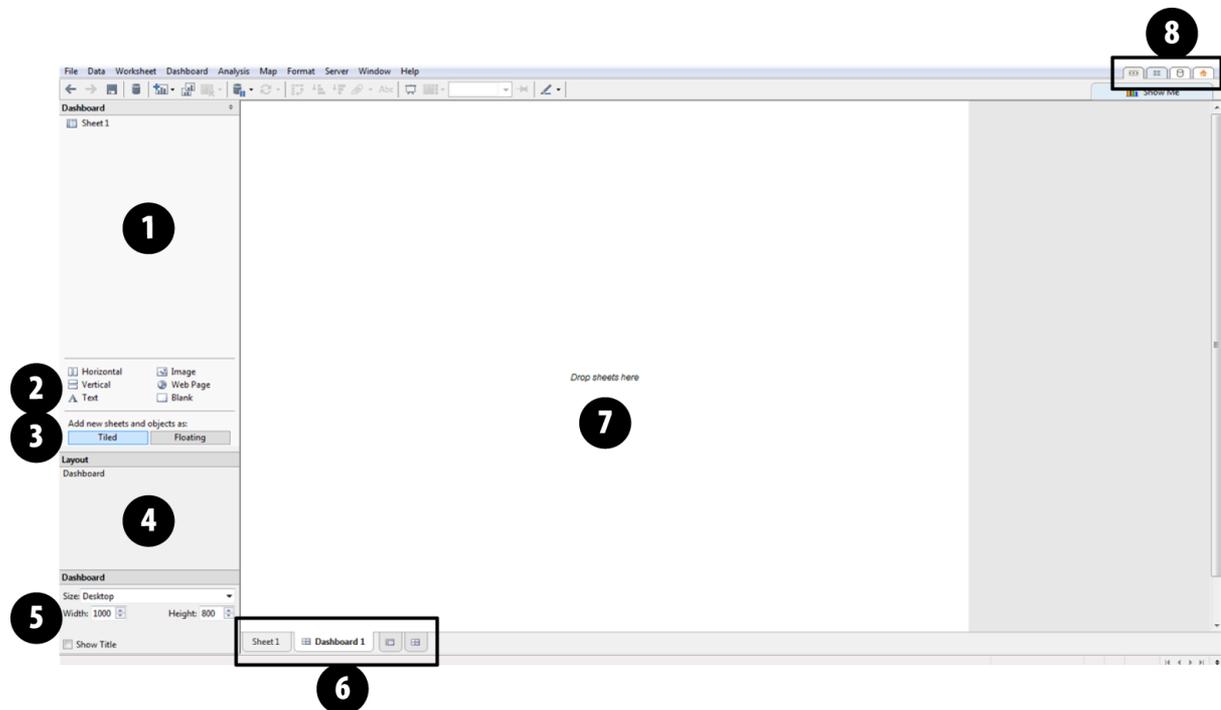
7.- Tarjeta de páginas. Filtra la visualización por pasos o animaciones basadas en un campo

8.- Vista . Zona donde aparece la visualización

9.- Hojas y cuadros de mandos. Lista las visualizaciones y cuadros de mandos disponibles

10.- Accesos rápidos. Nos permite acceder a conectar datos, Muéstrame y otras funcionalidades.

Cuadros de mando.



1.- Lista de visualizaciones creadas en este libro de trabajo.

2.- Objetos que podemos añadir al cuadro de mandos (imágenes, texto etc)

3.- Formato para el cuadro de mando, mosaico o flotante

4.- Muestra todas las hojas y objetos incluidos en el cuadro de mandos

5.- Control de tamaño para ajustarlo al medio de visualización

6.- Muestra todo lo que se ha creado y nos permite crear nuevos cuadros de mandos o visualizaciones

7.- El cuadro de mando.

8.- Accesos rápidos.

Aplicación visual basada en ventanas, tiene dos ediciones pública y profesional. La versión pública permite conectarse a archivos y grabar localmente las visualizaciones. Mientras que la versión profesional permite conectarse a una gran variedad de fuentes de datos, permite guardar las visualizaciones en un servidor, o publicarlas mediante Tableau online (servidores públicos o servidores Tableau). Es la versión recomendada para realizar analítica de negocios.