

Buenas prácticas en visualización de datos

Víctor Pascual Cid

PID_00243143



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-Compartir igual (BY-SA) v.3.0 España de Creative Commons. Se puede modificar la obra, reproducirla, distribuirla o comunicarla públicamente siempre que se cite el autor y la fuente (FUOC. Fundació per a la Universitat Oberta de Catalunya), y siempre que la obra derivada quede sujeta a la misma licencia que el material original. La licencia completa se puede consultar en: <http://creativecommons.org/licenses/by-sa/3.0/es/legalcode.ca>

Índice

1. ¿Qué es la visualización de datos?	5
2. El proceso de visualización de datos	8
3. Objetivos de la visualización de datos	10
3.1. Comunicar	10
3.2. Explorar	11
3.3. Datos en contexto	12
3.4. Encontrar patrones y <i>outliers</i>	14
4. Reglas para desarrollar buenas visualizaciones	16
4.1. Empezar con preguntas	16
4.2. Gráficos autoexplicativos	16
4.3. La funcionalidad debe prevalecer frente a la estética	17
4.4. Uso de la interacción	19
4.5. La forma sigue a la necesidad	21
4.6. Preparar bien los datos	22
Resumen	24

1. ¿Qué es la visualización de datos?

La visualización de datos, como se entiende hoy en día, se empezó a estudiar a finales de los años noventa. Su temprana edad hace que esté en constante evolución y continuamente aparecen nuevas definiciones de lo que es y lo que no es. Colin Ware definió el concepto de visualización como:

«la representación gráfica de datos o conceptos, que tiene como resultado una imagen mental o un artefacto externo que ayude a la toma de decisiones».

C. Ware (2012). *Information Visualization, Third Edition: Perception for Design*. San Francisco: Morgan Kaufmann.

Del mismo modo, Alberto Cairo la ha definido como:

«la representación visual de información diseñada para permitir la comunicación, el análisis, el descubrimiento y la exploración».

A. Cairo (2016). *The Truthful Art*. New Riders.

Dentro de este campo, se pueden distinguir dos tipos de visualizaciones:

1) Visualizaciones estáticas: su principal función suele ser la de comunicar unos datos que pueden haber sido analizados previamente. Este tipo de visualizaciones debe ayudar a descubrir patrones y valores atípicos (en inglés, *outliers*) en los datos. Además, son aquellas visualizaciones que pueden ser utilizadas en formatos físicos como periódicos y revistas en papel. En la figura 1 se puede ver una visualización hecha por el periódico *El País* en la que se muestra la evolución de la tasa de paro en España en forma de gráfico de líneas, y la evolución del porcentaje de población activa en paro representado con un gráfico de barras.

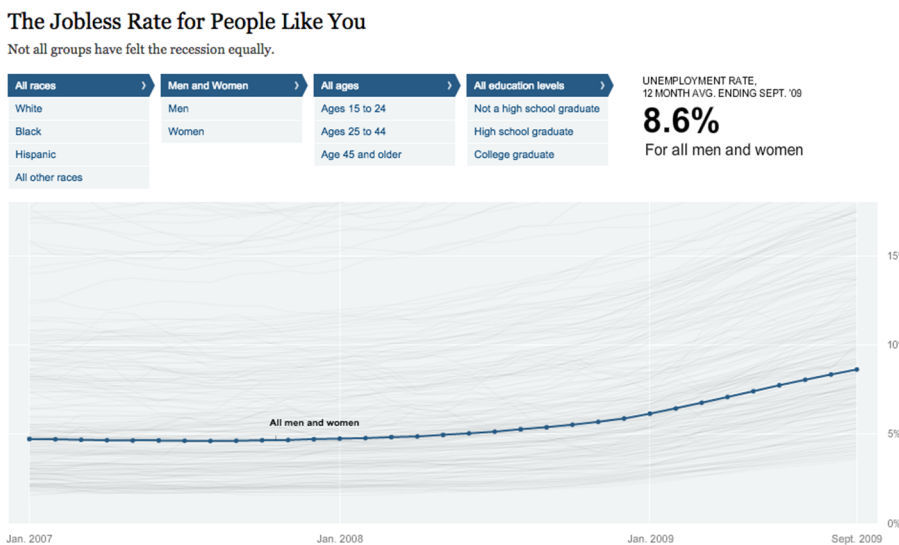
Figura 1. Evolución del número de parados y la tasa de paro en España.



Fuente: *El País*.

2) **Visualizaciones interactivas:** son aquellas que permiten a los usuarios interactuar con los datos. Esta propiedad hace que los datos puedan ser explorados por el usuario, al darle libertad para centrarse en aquello que más le interesa. Por ejemplo, acostumbrados a las clásicas representaciones de la evolución del paro, como lo que se mostraba en la figura anterior, la visualización interactiva del *New York Times* de la figura 2 permite al usuario elegir el sector de población del cual quiere ver la tasa de desempleo, mediante el uso del menú superior. De este modo, las visualizaciones interactivas permiten el descubrimiento y la exploración de los datos, además de comunicar el resultado de análisis previos.

Figura 2. Captura de pantalla de la visualización interactiva de la tasa de paro en Estados Unidos.



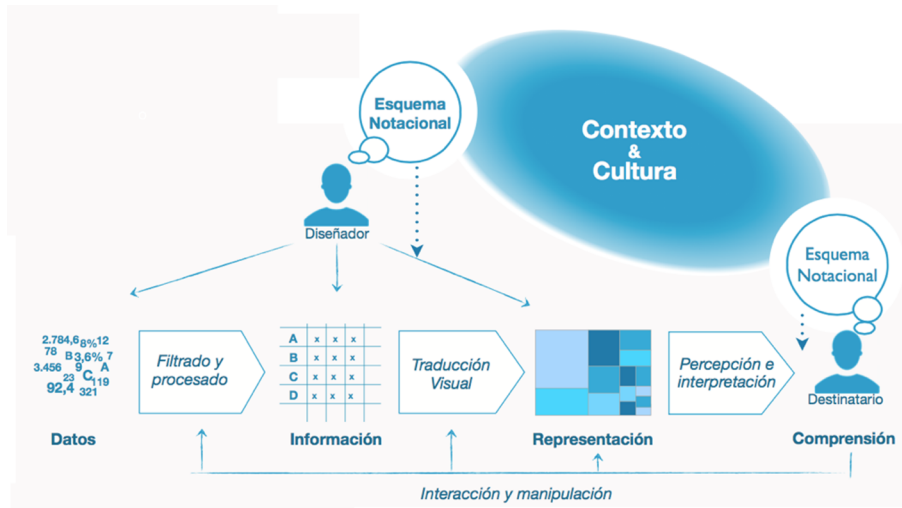
Fuente: *New York Times*.

Escoger un tipo de visualización u otra dependerá del objetivo que se quiera alcanzar. Mientras las visualizaciones estáticas pueden comunicar datos de forma rápida y concisa, las interactivas permiten al usuario explorar y jugar con los datos.

2. El proceso de visualización de datos

La figura 3 representa el diagrama de visualización de datos creado por Juan Carlos Dürsteler y Yuri Engelhardt. Este diagrama representa todo el proceso que se deberá realizar a la hora de generar una visualización.

Figura 3. El diagrama de visualización de datos de J. C. Dürsteler e Y. Engelhardt rediseñado por Jaime Pérez.



J. C. Dürsteler

Sitio web: <http://infovis.net/>

Y. Engelhardt

Sitio web: <http://yuriengelhardt.com/>

El diagrama parte de la idea de que se dispone de unos datos que han sido recolectados previamente. Estos datos son filtrados y procesados hasta convertirlos en información, es decir, hasta tenerlos limpios y estructurados. A esta información se le aplica una traducción visual o, dicho de otro modo, se la representa gráficamente. Finalmente, es importante ser conscientes de que esta representación será consumida por un usuario destinatario, el cual obviamente no tiene por qué ser el mismo diseñador de la visualización. Por este motivo, es muy importante ser consciente de cuál es el perfil del destinatario del diseño de la visualización. No es lo mismo publicar una infografía en un periódico generalista que generar un *report* para el CEO de una empresa que necesita tomar una decisión de manera muy rápida. Por ello será importante ser conscientes de que el esquema notacional de cada usuario puede ser distinto, es decir, que cada uno puede interpretar ciertas cosas de manera distinta. Del mismo modo, hay que tener en cuenta que los diseñadores y los usuarios finales pueden compartir o no un mismo contexto o cultura.

Será, pues, de vital importancia tener al usuario final en mente para evitar posibles errores de interpretación de la visualización creada. Para ello, se deberán realizar distintos tests de usabilidad con potenciales usuarios para asegurarse de la adecuación del diseño seleccionado. Además, el diseño deberá también prever el tipo de dispositivo en el cual se utilizará la representación, puesto

que no es lo mismo diseñar para una pantalla grande, como la de un ordenador de sobremesa o un portátil, que hacerlo para un dispositivo móvil mucho más pequeño, donde es más difícil realizar según qué tipo de interacciones.

Finalmente, es importante enfatizar que, gracias a las tecnologías actuales, será posible desarrollar sistemas con los que el usuario podrá interactuar y manipular los datos en cualquier fase del proceso. Por ejemplo, como se verá más adelante, es muy común ofrecer visualizaciones interactivas en las que un usuario puede filtrar los datos utilizando un menú.

3. Objetivos de la visualización de datos

En los apartados anteriores se ha definido qué es la visualización de datos, explicado los diferentes tipos de visualizaciones y el proceso para llevarlas a cabo. A continuación nos centraremos en los principales objetivos que persigue una buena representación.

3.1. Comunicar

El objetivo principal de la visualización de datos es el de comunicar una idea o el resultado de un análisis hecho sobre estos datos. En este sentido, es interesante ser conscientes de que las figuras 4 y 5 muestran la misma información, pero utilizan distintas representaciones. Mientras la tabla (figura 4) ayuda a ver el detalle concreto de los valores, el gráfico de líneas (figura 5) permite ver la evolución temporal del valor de las acciones. Ambas son útiles, pero el gráfico de líneas aporta mucha más información a simple vista. Es decir, el gráfico «comunica mejor». A pesar de que es posible escoger entre diferentes tipos de representaciones, con la visualización de datos se busca seleccionar la más adecuada para transmitir la información.

Figura 4. Tabla de valores de las acciones de una empresa en el NASDAQ extraído de *Yahoo! Finance*. La ausencia de elementos visuales que llamen la atención del sistema visual preatentivo limita la comprensión de los datos.

Aug 2, 2016	105.60	106.07	104.04	104.07	20,954,862
Aug 1, 2016	104.41	106.15	104.41	106.05	37,499,000
Jul 29, 2016	104.19	104.55	103.68	104.21	27,733,700
Jul 28, 2016	102.83	104.45	102.82	104.34	39,869,800
Jul 27, 2016	104.27	104.35	102.75	102.95	92,344,800
Jul 26, 2016	96.82	97.97	96.42	96.67	56,239,800
Jul 25, 2016	98.25	98.84	96.92	97.34	40,382,900
Jul 22, 2016	99.26	99.30	98.31	98.66	28,313,700
Jul 21, 2016	99.83	101.00	99.13	99.43	32,702,000
Jul 20, 2016	100.00	100.46	99.74	99.96	26,276,000
Jul 19, 2016	99.56	100.00	99.34	99.87	23,779,900
Jul 18, 2016	98.70	100.13	98.60	99.83	36,493,900
Jul 15, 2016	98.92	99.30	98.50	98.78	30,137,000
Jul 14, 2016	97.39	98.99	97.32	98.79	38,919,000
Jul 13, 2016	97.41	97.67	96.84	96.87	25,892,200

Figura 5. Gráfico de líneas que representa la evolución de los valores de las acciones de una empresa en el NASDAQ de la aplicación Stock de Yahoo!. La codificación de los datos en una línea hace que la memoria de trabajo procese

todos los datos representados en la línea como un solo bloque, lo que facilita su comprensión.



Este fenómeno se debe a que el cerebro humano funciona de manera muy parecida a un ordenador, en el cual se dispone de memoria RAM (la que se utiliza para cargar programas que son utilizados en el momento, ya que es de rápido acceso) y del disco duro, que es el almacén de datos que podemos consultar cuando queramos, pero de más lento acceso. En lugar de memoria RAM, en el cerebro humano se encuentra la denominada *memoria de trabajo*, mientras que en lugar de disco duro se tiene la denominada *memoria a largo plazo*. Cuando se observa una visualización, el sistema visual humano envía la información de lo que se ve en bloques a la memoria de trabajo. Sin embargo, se suele decir que la memoria de trabajo tan solo puede almacenar, en promedio, unos siete bloques (dependiendo del tipo de información, tipo de actividad y otros factores). Estos bloques son seleccionados por la memoria icónica, que es la primera que responde a los estímulos visuales (como, por ejemplo, formas, colores, contrastes, curvaturas y tamaños). Cuando se ha asimilado la información de la memoria de trabajo, esta pasará a formar parte de la memoria a largo plazo.

Esto es lo que explica el hecho de que resulta mucho más fácil entender los datos mediante la figura 5, que codifica todos los números de una columna en una única línea temporal. Sin embargo, en la figura 4 no hay ningún elemento que destaque, nada que llame nuestra atención (active nuestra memoria icónica) y, por tanto, no es posible «cargar» toda esa información en la memoria de trabajo, lo que dificulta su comprensión.

3.2. Explorar

En 1977, el estadístico John W. Tukey propuso el concepto de análisis de datos exploratorio (a menudo llamado EDA como acrónimo de *exploratory data analysis*). La propuesta de Tukey consiste en analizar los datos mediante la realización de sucesivas visualizaciones que ayuden a entender sus características principales.

Sin embargo, en los últimos veinte años el concepto de *explorar* ha tomado una nueva dimensión gracias a los avances tecnológicos. Y es que la rápida capacidad de cálculo de los dispositivos de hoy en día permite la existencia de visualizaciones interactivas en las cuales la representación visual puede cam-

Referencia bibliográfica

G. A. Miller (1956). «The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information». *Psychological Review*.

Para saber más

exploratory data analysis

biar de forma, o proporcionar nueva información a partir de las interacciones del usuario. La potencia de este concepto reside en el hecho de que, mientras se interactúa con los datos, la visualización podrá también mostrar nuevas perspectivas de esos datos. Estos nuevos puntos de vista pueden ayudar al descubrimiento de hechos que ni tan siquiera se estaban buscando, lo que la convierte en una herramienta de análisis de gran capacidad.

Para entender más la potencia de este concepto, se recomienda dedicar unos minutos a explorar la visualización del *New York Times* «How Different Groups Spend Their Day». Esta visualización muestra el porcentaje de ciudadanos americanos que dedican tiempo a actividades como dormir, comer, trabajar, etc. a lo largo del día. Antes de empezar a utilizar la visualización, es importante pensar qué se desea saber sobre los datos que representa. Tal vez ¿qué hacen los americanos cuando salen de trabajar? O si ¿duermen más los hombres que las mujeres?, o ¿cómo cambian los porcentajes entre los que no tienen hijos y los que tienen más de dos?

Como usuarios de esta visualización, una vez se haya interactuado con ella, es importante recapacitar sobre todas aquellas otras cosas que se han ido descubriendo gracias a las capacidades interactivas del gráfico. Así, es posible suponer que los datos más cruciales para un trabajo u objetivo concreto están representados de manera interactiva e imaginar el conocimiento que se podría llegar a extraer de ellos.

3.3. Datos en contexto

Ben Shneiderman propuso en 1996 el *information-seeking mantra* o **mantra de la búsqueda de información**:

«Overview first, zoom and filter, then details-on-demand»

B. Shneiderman (1996). «The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations». En: *Proceedings of the IEEE Symposium on Visual Languages*. Washington: IEEE Computer Society Press (págs. 336-343).

Lo que propone Shneiderman con su mantra es que para buscar información hay que proporcionar primero una vista general de los datos, para luego filtrarlos y obtener detalles de aquello que más interesa. Representar una vista general de todos los datos es lo que permite tener un contexto sobre ellos, de modo que sea posible entender, por ejemplo, si un valor concreto es grande o pequeño en función de cómo es en el conjunto de datos.

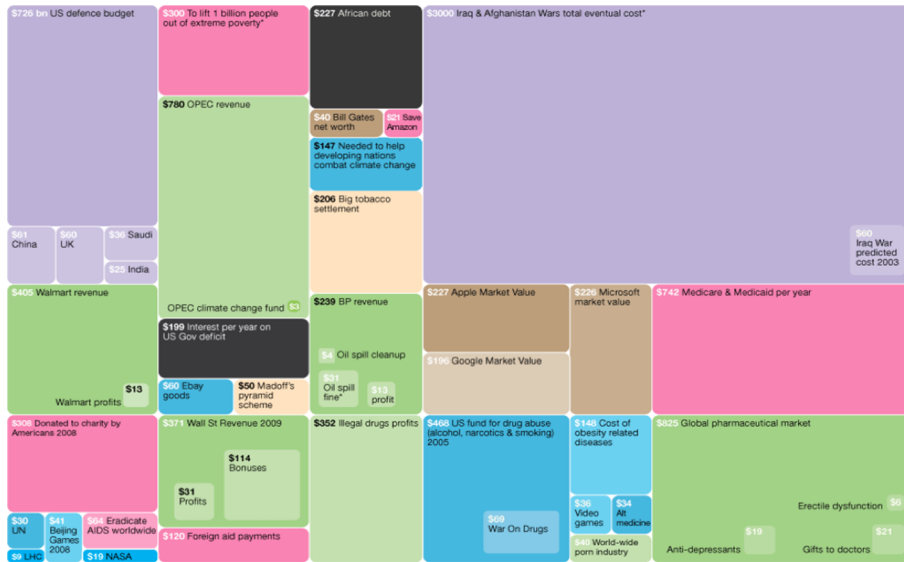
En este sentido, el «Billion Dollar-O-Gram» de la figura 6 es una visualización diseñada por David McCandless con el objetivo de hacer tangibles grandes valores que cada día se pueden oír en la prensa. A menudo pasa que cuando se mencionan volúmenes de dinero (o de cualquier otra magnitud) tan grandes se hace difícil llegar a entender y comparar cuánto es poco, mucho o muchísimo. En esta representación se pueden ver conceptos como lo que se predijo que costaría la guerra de Irak, frente a lo que ha costado hasta la fecha (recuadros

New York Times

Sitio web: <http://www.nytimes.com/interactive/2009/07/31/business/20080801-metrics-graphic.html>

en color violeta en la esquina superior derecha), o que el valor de Apple y Google en el mercado es similar (cuadrados marrón oscuro y marrón claro en el centro).

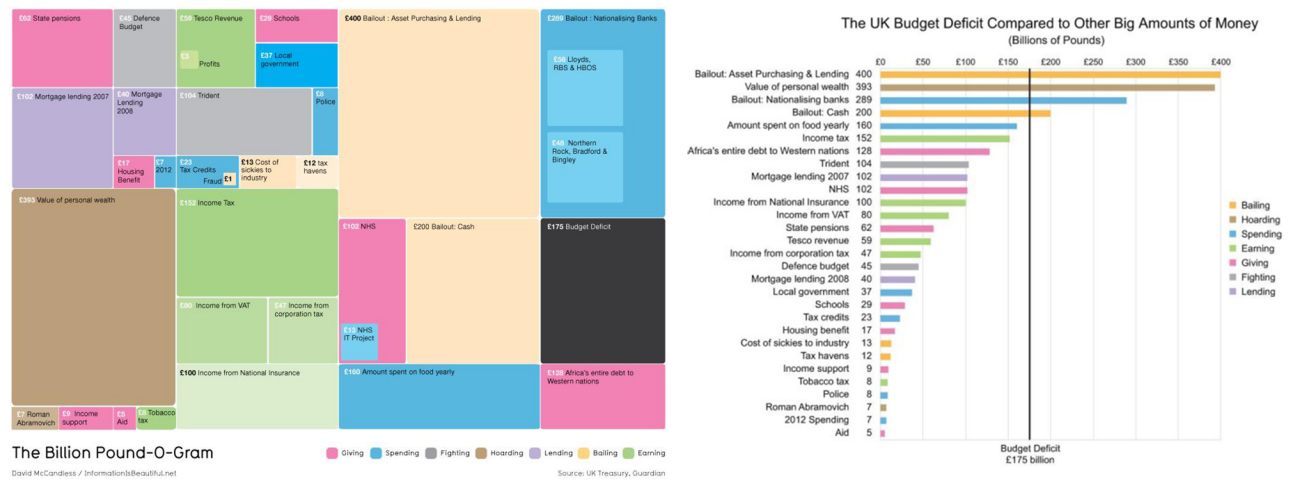
Figura 6. «The Billion Dollar-O-Gram» pone distintos valores en contexto.



Fuente: <http://www.sickchirpse.com/wp-content/uploads/2013/01/Billion-Dollar-O-Gram-Dave-McCandless-USA-version.jpg>

Es muy importante ser conscientes de que esta visualización posibilita tener una vista general de los datos, y ayuda a poner en contexto cada uno de los valores, lo que relativiza algunos de esos números tan grandes. Sin embargo, es vital entender que, como se puede ver con los cuadrados que simbolizan Apple y Google, esta representación no es buena para hacer comparaciones precisas, ya que el cerebro humano no está preparado para comparar áreas. Sin embargo, sí que lo está para longitudes o distancias, como se puede observar en el rediseño de la visualización mostrado en la figura 7.

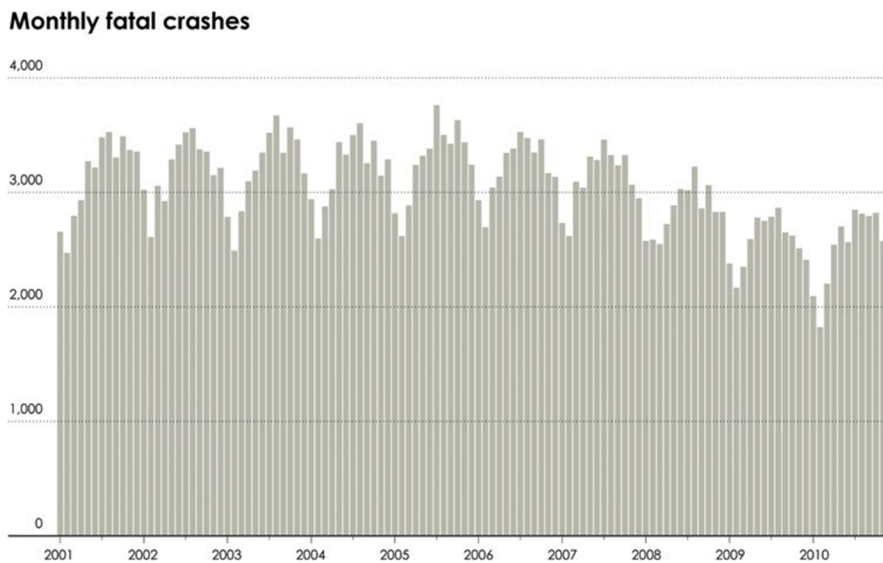
Figura 7. Propuesta de rediseño del «Billion Pound-O-Gram» para facilitar la comparación hecha por Stephen Few. La idea consiste en utilizar un gráfico de barras horizontales, el cual facilita mucho la comparación entre los distintos valores.



3.4. Encontrar patrones y outliers

Otro objetivo fundamental de la visualización de datos es el de facilitar el descubrimiento de patrones u *outliers* (término inglés para referirse a ‘valores atípicos’). El gráfico de la figura 8 es un clásico ejemplo en el que se puede ver la estacionalidad de los accidentes en Estados Unidos desde 2001 hasta 2011. Además de poder apreciar que los accidentes aumentan durante los meses de verano, ya que existe un patrón anual, también se puede observar un descenso de accidentes a partir de 2008.

Figura 8. Accidentes mortales por mes en Estados Unidos.



Fuente: *Data Points: Visualization That Means Something*.

Sin embargo, merece la pena reflexionar sobre este objetivo desde una perspectiva más abierta. La figura 9 muestra una captura de pantalla tomada del proyecto «Flight Patterns» de Aaron Koblin. El autor disponía de la posición GPS de miles de aviones que sobrevolaban el espacio aéreo norteamericano durante varias horas, y lo representó por medio de puntos de color que dejan un pequeño trazo de su trayectoria, usando colores más claros (es decir, más blancos) cuando el avión se acercaba al suelo.

Referencia bibliográfica

N. Yaw (2013). *Data Points: Visualization That Means Something*. Somerset: Wiley.

Aaron Koblin

Sitio web: <http://www.aaronkoblin.com/>

Figura 9. «Flight Patterns» de Aaron Koblin. Captura del vídeo. http://www.aaronkoblin.com/work/flightpatterns/FPWeb_Final_3.mov



Gracias a las codificaciones visuales utilizadas en el proyecto, esta representación revela un claro patrón conocido por todos: el mapa de Estados Unidos. Y no solo eso, sino que también revela la posición de los aeropuertos de Estados Unidos, que son aquellos puntos que sobresalen en todas las rutas que se aprecian en el mapa.

Es evidente que no hacen falta estos datos para descubrir el mapa. Sin embargo, es importante recordar que aquí no se está representando un mapa, sino las trayectorias de un conjunto de aviones, las cuales llevan el mapa implícito. La genialidad de este proyecto es que **cuando se representan los datos de manera adecuada, se propicia el descubrimiento de patrones ocultos en ellos.**

4. Reglas para desarrollar buenas visualizaciones

Diseñar una visualización implica decidir qué codificaciones visuales se utilizarán para representar los datos. Las principales codificaciones visuales que existen son: posición, forma, color y movimiento¹. A continuación se detalla una serie de reglas a tener en cuenta para la elección de dichos elementos. Como se podrá ver, muchas de estas reglas son muy sencillas de aplicar, y generalmente se podrán implementar mediante cualquier herramienta que pueda trabajar con datos.

⁽¹⁾Para más detalles «Preattentive Visual Properties and How to Use Them in Information Visualization».

4.1. Empezar con preguntas

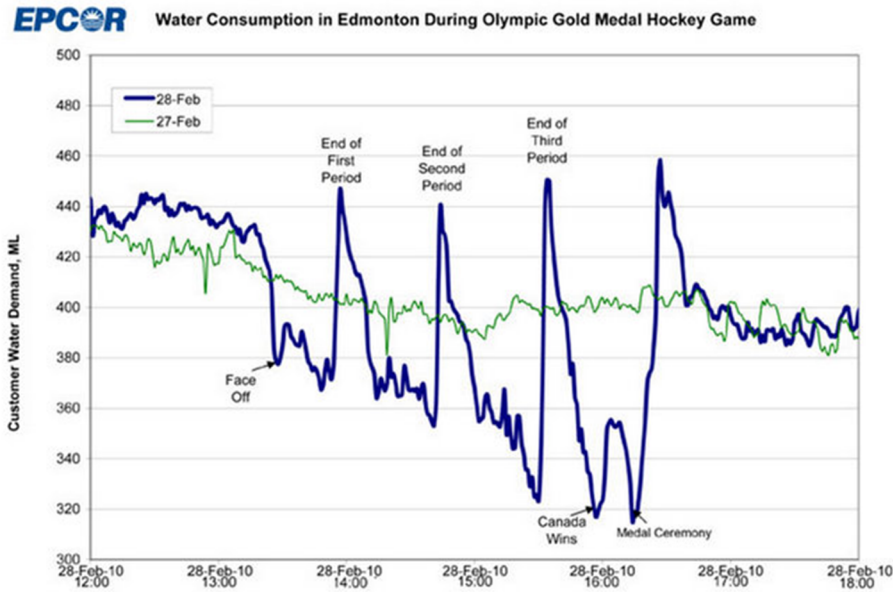
Una visualización puede tener muchas formas distintas. Por ejemplo, un conjunto de datos que contenga información sobre accidentes en España podrá tener la localización del accidente, el número de ocupantes del vehículo, el número de víctimas o heridos, o el tipo de vehículo. Está claro, pues, que en función de lo que se quiera saber (en función del «objetivo informacional») se podrán representar estos datos en un mapa, en un gráfico de barras que acumule el número de accidentes por número de ocupantes, o bien el número de heridos o víctimas.

Por lo tanto, tener preguntas concretas que se quieran hacer a los datos es de vital importancia para crear una representación visual que ayude a responderlas correctamente.

4.2. Gráficos autoexplicativos

Por lo general, es deseable que el usuario sea capaz de entender una visualización sin leer el texto que la acompaña (como en el caso de una noticia periodística) o sus instrucciones. Para ello se dispone de una herramienta elemental que mejora mucho la legibilidad de las visualizaciones: los **textos** o **etiquetas**.

El siguiente gráfico de la figura 10 muestra dos líneas temporales: la verde representa el consumo de agua en la ciudad canadiense de Edmonton un día laborable cualquiera, mientras que la azul muestra el consumo de agua el día de la final de *hockey* sobre hielo de las olimpiadas de invierno, en la que participaba la selección nacional de Canadá. Viendo el gráfico, y gracias a las etiquetas, es muy fácil poder ver cómo el consumo de agua está íntimamente relacionado con el devenir del partido, puesto que en cada final de periodo (un partido de *hockey* consta de tres periodos) hay un repunte en el consumo de agua (presumiblemente debido a que los espectadores hacen una «parada técnica» en el baño, mientras que no la hacen durante el partido).

Figura 10. Visualización del consumo de agua en Edmonton realizada por la empresa EPCOR.

Fuente: <http://www.zdnet.com/article/infographic-water-consumption-in-edmonton-canada-during-olympic-gold-medal-hockey-game/>

Gracias a las etiquetas y al título de la representación es muy fácil entender la historia que se esconde tras estos datos. Y es que no solamente se puede ver que, a medida que avanzaba el partido, más gente aprovechaba (presumiblemente) para ir al baño, sino que además hubo un periodo extra de desempate justo después del final del tercer periodo. Finalmente, cuando Canadá ganó, algunos espectadores aprovecharon para volver al baño justo antes de poder ver la entrega de medallas. Sin las etiquetas, sería necesario conocer el contexto para relacionar el consumo con el partido.

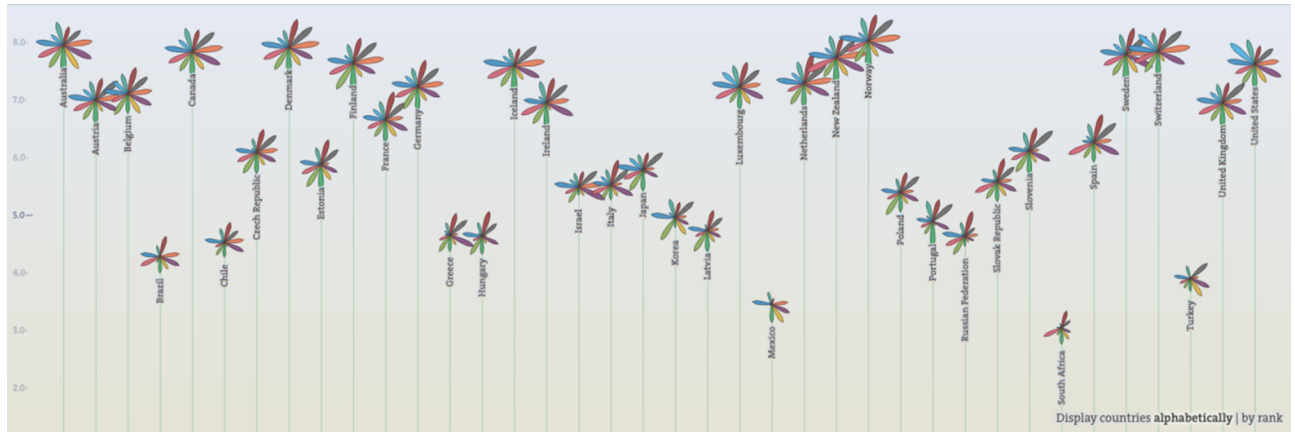
4.3. La funcionalidad debe prevalecer frente a la estética

No hay ninguna duda de que crear visualizaciones bellas tiene un poder de persuasión muy importante. El «Better Life Index» de la OECD, mostrado en la figura 11, es un ejemplo de visualización muy atractiva que representa gráficamente el índice para comparar países creado por esta institución. El uso de una metáfora (la flor) para la representación de cada país resulta muy interesante desde el punto de vista estético, aun cuando pueda presentar problemas con respecto a la información que se representa gráficamente. Se trata, de hecho, de una visualización interactiva muy compleja que combina diferentes indicadores (uno por pétalo) y un indicador ponderado de los mismos (la posición de la flor).

OECD

Sitio web: <http://www.oecdbetterlifeindex.org/>

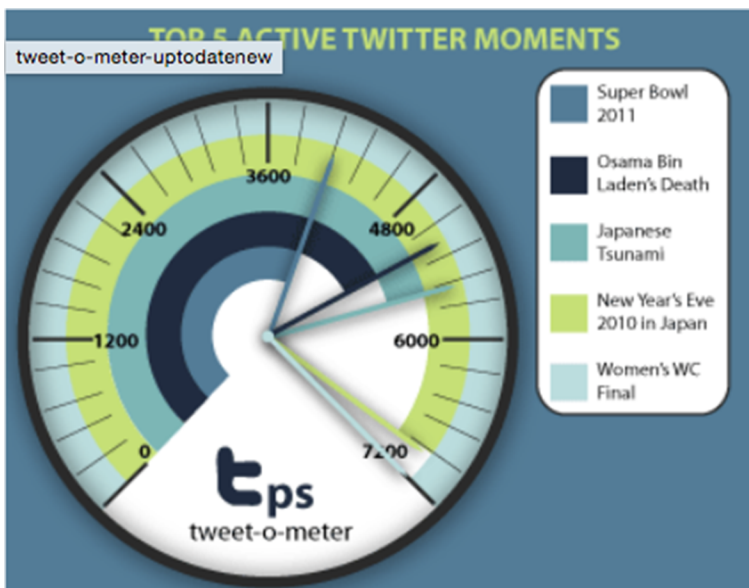
Figura 11. Flores representando barras en el «Better Life Index» de la OECD.



Fuente: <http://www.oecdbetterlifeindex.org>

Sin embargo, es de vital importancia entender que el objetivo principal de la visualización de datos, en un entorno analítico, debe ser el de generar conocimiento sobre los datos con los que se trabaja. Obsérvese, por ejemplo, la representación de la figura 12, que muestra los momentos más activos en Twitter durante el año 2011. ¿Cuál ha sido el tercer momento más activo?

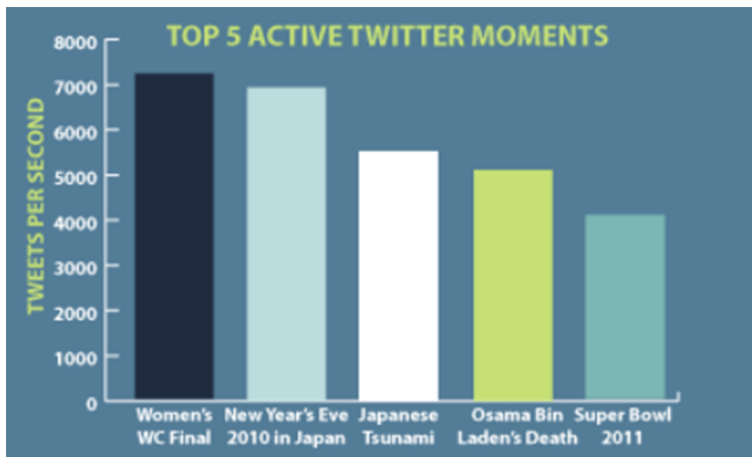
Figura 12. Los cinco momentos más activos en Twitter en 2011 representados en una metáfora parecida a un cuentakilómetros.



Fuente: <https://www.smashingmagazine.com/2011/10/the-dos-and-donts-of-infographic-design/>

Probablemente, para contestar a la pregunta se han necesitado varios segundos, e incluso ha podido ser algo confuso interpretar el gráfico. Esto es porque la representación utilizada no es la más adecuada para el «objetivo informacional» que se pretende. Si se intenta hacer el mismo ejercicio con el gráfico de la figura 13, se puede observar que resulta mucho más sencillo.

Figura 13. Representación alternativa al cuentakilómetros que transmite mejor la información.



Es importante darse cuenta de que, en ambas visualizaciones, el valor de tuits por segundo se ha codificado con una barra. Mientras en la figura 12 el valor a representar se ha codificado mediante un ángulo (cuantos más tuits, más larga es la barra circular o segmento de arco), en la figura 13 se ha codificado con la distancia de la barra. De esta manera, la longitud de las barras de la figura 12 no es proporcional a los valores que representan², lo que causa confusión. El uso del color de fondo también es un aspecto que debe tenerse en cuenta en este caso en concreto.

⁽²⁾ Hay que recordar que la longitud de la circunferencia es $2 \times \pi \times r$. De este modo, aquellas barras circulares que estén en el exterior tendrán una longitud mayor que las interiores dado un mismo ángulo, de ahí la falta de proporcionalidad entre los diferentes segmentos de arco.

4.4. Uso de la interacción

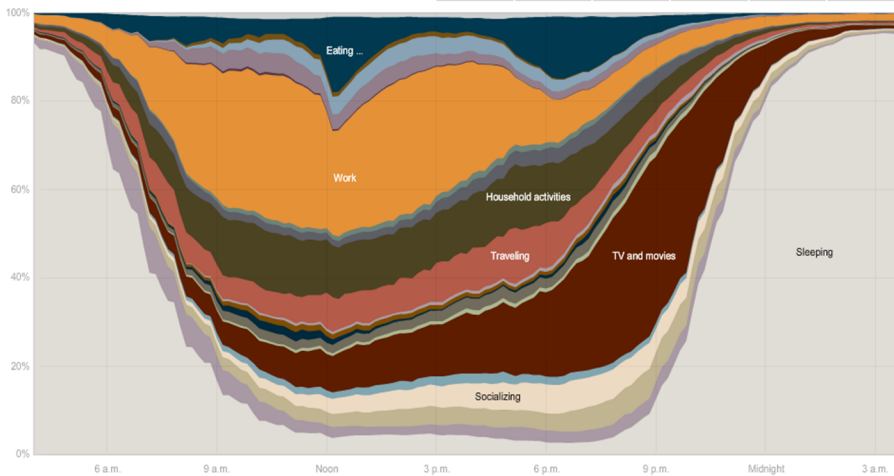
Anteriormente se ha comentado la existencia de dos tipos de visualizaciones: las estáticas y las interactivas. Es muy importante ser consciente de que el uso de interacción puede aportar muchos beneficios a una visualización. Una de las limitaciones que sufren los gráficos estáticos es que «solo» pueden comunicar un conjunto concreto de datos de una única manera, mientras que en los gráficos interactivos se puede permitir al usuario que decida en qué fijarse. Además, hay veces en que la decisión que se toma a la hora de decidir la forma que tendrán los datos implica no ser todo lo precisos que se desearía en algún aspecto concreto de estos. Obsérvese, por ejemplo, la visualización de la figura 14, ya utilizada anteriormente en el apartado 3.3. pero ahora plasmada como visualización estática:

Figura 14. «How Different Groups Spend Their Time».

Everyone

Sleeping, eating, working and watching television take up about two-thirds of the average day.

Everyone	Employed	White	Age 15-24	H.S. grads	No children
Men	Unemployed	Black	Age 25-64	Bachelor's	One child
Women	Not in lab...	Hispanic	Age 65+	Advanced	Two+ children

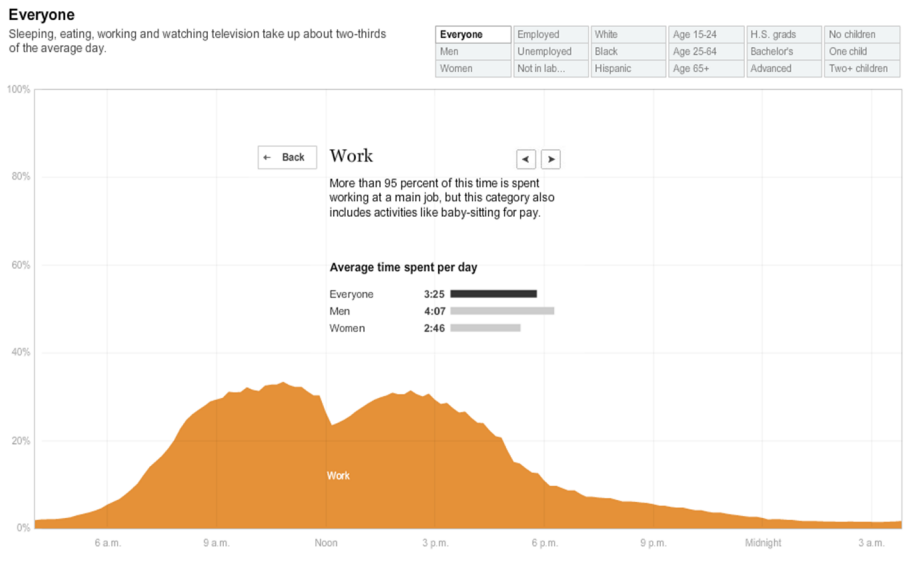


Fuente: *New York Times*.

Esta representación es un gráfico de áreas apiladas en el que cada área coloreada representa un tipo de actividad que realizan los norteamericanos durante el día. El eje de las x representa el tiempo, que empieza a las cuatro de la mañana, que es cuando se asume que empieza el día para muchas personas y también el de menor actividad. El eje de las y es el que ayuda a identificar el porcentaje de población que está realizando cada actividad, representada por un color diferente. El gráfico es bastante claro e interesante, y gracias a la interactividad proporcionada por el menú superior se pueden ver los datos relativos al segmento de población en el que se está interesado, pues permite cambiar entre un sector u otro fácilmente, para ver así los cambios en las distintas actividades.

Hasta aquí esta visualización es muy buena, sin embargo, la elección de la codificación de los datos hace que sea muy difícil, por ejemplo, comparar si hay más gente que está trabajando a las diez de la mañana que a las doce del mediodía. Eso es porque las áreas de las distintas actividades están apiladas y **no están referenciadas en el cero**. Sin embargo, los diseñadores de esta visualización crearon otra interacción para solucionar este problema: con un clic en cualquier área, esta queda referenciada en el cero y la comparación que se deseaba realizar pasa a ser mucho más sencilla de ejecutar, como muestra la figura 15.

Figura 15. Selección de un tipo de actividad que permite realizar otro tipo de análisis visual mediante el uso de la interacción.



Así pues, se puede observar que la interacción puede ayudar a solucionar problemas inherentes a la codificación visual elegida y así proporcionar mejor y mayor información, desde diferentes perspectivas.

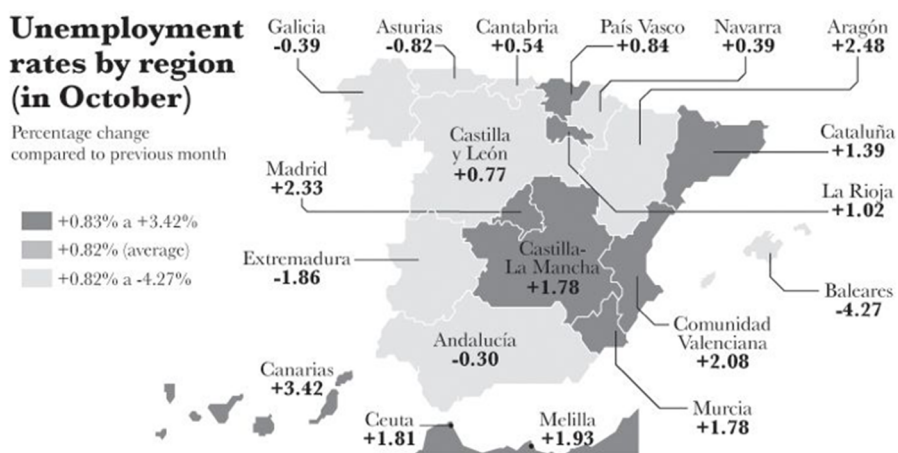
4.5. La forma sigue a la necesidad

Como ya se ha comentado con anterioridad, es importante tener claro cuál es el objetivo de la visualización, para poder así elegir aquella codificación visual que sea más adecuada para este. Alberto Cairo pone un buen ejemplo en su libro *The Functional Art*, en el cual se muestra la visualización de la figura 16, sobre datos de empleo.

Alberto Cairo

Sitio web: <http://albertocairo.com/>

Figura 16. Representación de la tasa de paro por comunidad autónoma.

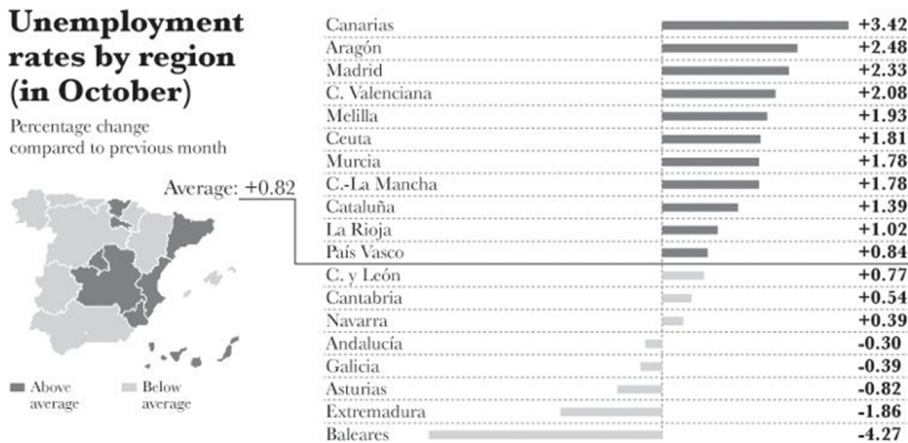


Fuente: *The Functional Art*.

Dado el porcentaje de cambio en la tasa de paro por comunidad autónoma respecto al mes anterior, parece razonable pensar que el objetivo principal de una visualización de dichos datos debería permitir descubrir qué comunidades autónomas han mejorado más y cuáles están teniendo más problemas en comparación con el mes anterior. Sin embargo, la figura 16 hace que esta tarea

sea bastante complicada, puesto que solo se han utilizado tres gradaciones de color, y dentro de las comunidades con un mismo color no queda más remedio que inspeccionar los números, intentar memorizarlos, y después intentar inferir el orden de los valores. En definitiva, esta visualización ayuda más bien poco a responder a la pregunta más básica que se esconde dentro de estos datos.

Figura 17. Rediseño de la figura 16 hecho por Alberto Cairo.



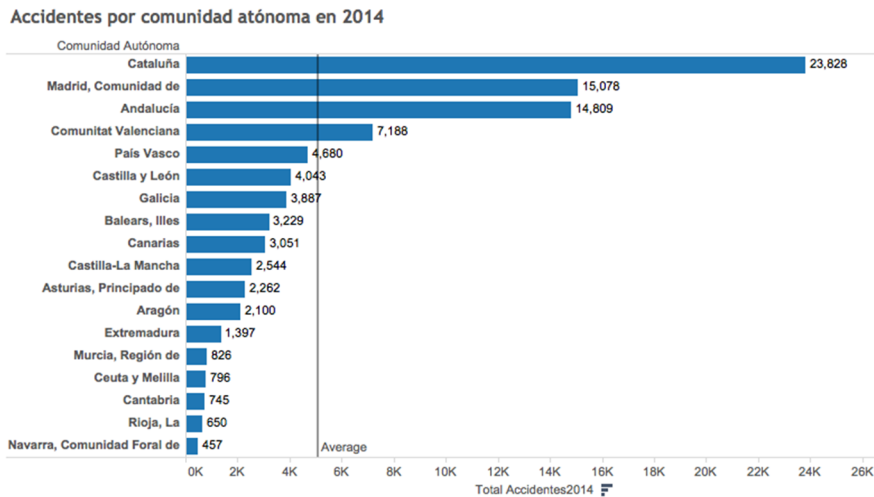
Fuente: *The Functional Art*.

La figura 17 muestra el rediseño de la figura 16 hecho por el autor, en el que propone la utilización de un gráfico de barras que, como se puede observar, es mucho más claro y comunica mejor los datos disponibles que el mapa, el cual queda como elemento que proporciona contexto geográfico.

4.6. Preparar bien los datos

Para hacer una buena visualización de datos es de vital importancia disponer de buenos datos. Esto significa que se tendrá que trabajar con los datos antes de representarlos, con el fin de poder mostrar el mensaje que se esconde detrás de ellos. Un claro ejemplo de preparación de datos es el proceso de normalización. En la siguiente visualización se representa el número de accidentes mortales por comunidad autónoma en el año 2014. En la figura 18 se puede ver que Cataluña, Madrid y Andalucía son las tres comunidades con mayor número de accidentes mortales. Sin embargo, estas tres comunidades autónomas son las más pobladas, por lo que tiene sentido que también sean las que tienen más tráfico y, por lo tanto, más accidentes. Esta información es útil pero no permite, por ejemplo, hacer un análisis que posibilite a la Dirección General de Tráfico decidir en qué comunidad autónoma es necesario aplicar nuevas medidas para prevenir accidentes.

Figura 18. Accidentes mortales por comunidad autónoma en 2014.

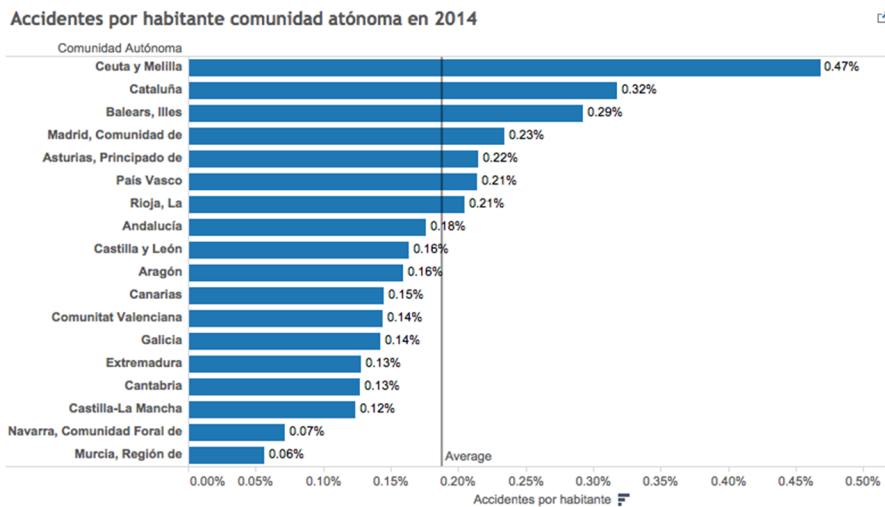


Fuente: Elaboración propia.

En el gráfico de la figura 19 se ha aplicado una normalización, calculando el número de accidentes por habitante. De este modo, se puede ver que Ceuta y Melilla³ pasan a estar en la primera posición de la clasificación, y que, por ejemplo, las islas Baleares se sitúan por encima de Madrid.

⁽³⁾También hay que tener en cuenta que Ceuta y Melilla tienen muy pocos habitantes, lo que implica que la tasa de accidentes por habitante puede variar muy fácilmente.

Figura 19. Accidentes mortales por habitante para cada comunidad autónoma en 2014.



Fuente: Elaboración propia.

Resumen

A modo de sumario, para hacer una buena visualización existe una serie de reglas básicas que deben tenerse en cuenta:

- Es muy importante tener claro qué se quiere que cuente una visualización, quién va a ser el consumidor de esta y en qué dispositivo la va a utilizar.
- Un conjunto de datos podrá ser representado de muchas formas distintas. La misión del analista o diseñador será escoger aquella visualización que mejor ayude a cumplir con el objetivo informacional del usuario final.
- Existen dos tipos fundamentales de visualización: las estáticas y las interactivas. Por lo general, las primeras se centran más en comunicar datos o hechos mediante recursos visuales, mientras que las segundas permiten un análisis de datos más exploratorio.
- Las visualizaciones de datos en un entorno analítico tienen como prioridad ayudar a generar conocimiento en torno a los datos, y no la mera generación de imágenes atractivas.

Para aprender a generar visualizaciones de datos no hay más alternativa que practicar con el ejemplo. Cuantas más visualizaciones se hagan, más se conseguirá interiorizar todas aquellas reglas que, al final, determinan la diferencia entre una buena y una mala visualización. Y, por supuesto, estar al corriente de las nuevas tendencias y herramientas para su desarrollo. Se trata, por lo tanto, de seguir el ejemplo descrito por Enrico Bertini en forma de ciclo continuo: aprende, copia, critica, crea y expón tus visualizaciones.

Referencia bibliográfica

Enrico Bertini (2010). «How to Become a Data Visualization Expert: A Recipe».