
Introducción al análisis visual mediante D3

Datos, *layouts*, visualizaciones

PID_00279535

Julià Minguillón Alfonso

Tiempo mínimo de dedicación recomendado: 2 horas



**Julià Minguillón Alfonso**

Doctor ingeniero en Informática por la Universitat Autònoma de Barcelona (UAB). Profesor agregado de los Estudios de Informática, Multimedia y Telecomunicación de la Universitat Oberta de Catalunya (UOC) en el ámbito de la ciencia de datos. Sus intereses docentes incluyen la programación, la minería de datos y la visualización, entre otros. En investigación, se dedica a analizar el comportamiento de los usuarios en entornos virtuales de aprendizaje y redes sociales, como la *Wikipedia*, con el objetivo de mejorar los procesos de apoyo al aprendizaje y la interacción con el entorno.

Segunda edición: septiembre 2021
© de esta edición, Fundació Universitat Oberta de Catalunya (FUOC)
Av. Tibidabo, 39-43, 08035 Barcelona
Autoria: Julià Minguillón Alfonso
Producción: FUOC



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

Índice

Introducción	5
1. La visualización como exploración de datos.....	7
2. La visualización como análisis preliminar.....	9
3. Herramientas para el análisis visual de datos	12
4. <i>Layouts D3 para el análisis visual de datos</i>	14
4.1. <i>Treemap</i>	14
4.2. <i>Bubble</i>	15
4.3. <i>Chord</i>	15
4.4. <i>Parallel sets</i>	16
4.5. <i>Force-directed graphs</i>	17
4.6. <i>Sankey.....</i>	17
4.7. <i>Sunburst</i>	18
4.8. <i>Parallel coordinates.....</i>	18
4.9. <i>Choropleth.....</i>	19
4.10. <i>Word cloud</i>	19
4.11. <i>Multiline graphs.....</i>	20
4.12. <i>Scatterplot</i>	21
Bibliografía	22

Introducción

En este material docente se introducen los conceptos básicos relacionados con el análisis visual de datos, así como el uso de una librería de software (D3) que permite visualizar datos de acuerdo a diversos formatos preestablecidos (*layouts*), proporcionando un entorno interactivo para la exploración y análisis visual de los mismos.

La visualización de datos es una herramienta muy eficaz para realizar un análisis preliminar de los mismos, aprovechando las capacidades del sistema visual humano para detectar y extraer conocimiento en forma de patrones, tendencias, *outliers*, etc. Para ello solamente es necesario representar los datos gráficamente de acuerdo al objetivo del análisis que se desee realizar, usando algún tipo predeterminado de visualización que se considere adecuada (por ejemplo, por haber sido satisfactoriamente usada con anterioridad en un conjunto de datos similar). Esto puede hacerse con herramientas generalistas como Tableau, pero estas nos limitan mucho en cuanto a los tipos de gráficos disponibles, su diseño y la posibilidad de ir más allá, combinando dos o más visualizaciones diferentes o bien añadiendo nuevos elementos gráficos. Por este motivo, exploraremos otras opciones más cercanas a la programación, en las que la visualización de datos se convierte en el resultado de un código que se ejecuta.

Asimismo, aunque la visualización de datos es una disciplina que requiere de conocimientos varios pertenecientes a distintos ámbitos (estadística, diseño gráfico, computación, psicología...), en este material docente describiremos principalmente modelos existentes para crear visualizaciones de una forma relativamente rápida y simplificada, que pueden servir de base para crear visualizaciones interactivas más complejas, como las usadas por Mike Bostock* en el *New York Times* para narrar historias mediante datos, por ejemplo. La idea es reaprovechar el código existente, modificándolo para generar nuestra visualización de datos.

*<https://goo.gl/aRFFhU>

Un estudio riguroso del proceso de visualizar datos requiere, entre otros aspectos, profundizar y comprender factores como la subjetividad intrínseca a las representaciones y el mapeo de datos, la importancia del aspecto visual y sus consecuencias cognitivas, etc. No obstante, es posible intentar reutilizar buenas prácticas en visualización de datos para explorar nuevos datos. En definitiva, se trata de ser capaces de «reciclar» las buenas prácticas.

De esta forma, sin tener que conocer librerías como D3 en profundidad, es posible adaptar visualizaciones ya existentes para visualizar nuevos datos, usan-

do el *layout* o configuración más apropiado en cada ocasión, en función de la naturaleza de los datos y de los objetivos de cada visualización. Para ello se proporcionan un conjunto de visualizaciones interactivas que incluyen plantillas (código HTML + CSS + *scripts* D3) que pueden ser modificadas para visualizar conjuntos de datos en formato CSV o JSON. Estas visualizaciones se ejecutan como páginas web y se muestran mediante un navegador, sin que sea realmente necesario disponer de un verdadero entorno de programación. Actualmente, esta idea se ha desarrollado en forma de lo que se conoce como *notebook*, un documento interactivo capaz de contener y ejecutar código (de diferentes lenguajes de programación), mostrando los resultados inmediatamente de forma interactiva. Es el caso del entorno Observable, creado por el ya mencionado Mike Bostock y por Melody Meckfessel, que permite la creación sencilla de visualizaciones de datos interactivas en forma de *notebooks* ejecutables.

En resumen, el objetivo de este material docente es proporcionar un mecanismo sencillo para transformar datos (en formato tabular o jerárquico) en visualizaciones interactivas que permitan su manipulación casi sin tener que programar, usando los diferentes *layouts* de D3 como un «molde» que genera la visualización a partir de los «ingredientes» adecuados (los datos en el formato preestablecido), ya sea como parte de una página web o en forma de *notebook* en Observable.

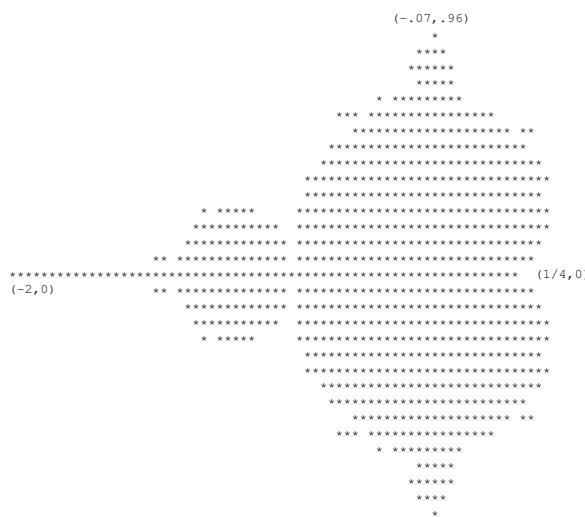
Enlace de interés

Para conocer más sobre Observable, véase el siguiente enlace:
<https://observablehq.com/about>

1. La visualización como exploración de datos

Aplicaciones para visualizar datos han existido desde hace mucho tiempo, aunque se han popularizado recientemente dadas las posibilidades que ofrece el hardware disponible en la actualidad, el cual ha ido doblando su capacidad, velocidad y resolución a lo largo de los años. No hace muchos años, sin dicha capacidad computacional, la generación de gráficos por ordenador era muy limitada, tanto por la resolución gráfica como del tiempo de cómputo necesario para generar los gráficos. De hecho, las primeras visualizaciones debían utilizar los recursos básicos del juego de caracteres para representar gráficos. Lanzadas al mercado en el año 1981 con el IBM PC, las primeras tarjetas gráficas CGA tenían una resolución de 320 x 200 píxeles y una paleta de dos bits o cuatro colores. Menos de cuatro décadas después, el hardware disponible permite generar gráficos de hasta 7680 x 4320 píxeles (lo que se conoce como 8K), con veinticuatro bits de profundidad de color RGB y una frecuencia de refresco de 60 Hz.

Figura 1. Primera visualización del conjunto de Mandelbrot, por Brooks y Matelski, 1978



Fuente: <http://mrob.com/pub/muency/brooksandmatelski.html>

Un ejemplo de la evolución de la capacidad y las posibilidades ofrecidas por las tarjetas gráficas lo proporciona la visualización de objetos fractales, construidos a partir de un cierto algoritmo y unos datos o parámetros de entrada. Se trata de descripciones algorítmicas muy sencillas que generan imágenes muy complejas, con un grado de detalle virtualmente infinito. Cuando Benoit Mandelbrot presentó a finales de los setenta el conjunto que lleva su nombre, tuvo que utilizar las limitadas capacidades de representación de la época, como se puede ver en la primera imagen creada del mismo por Ro-

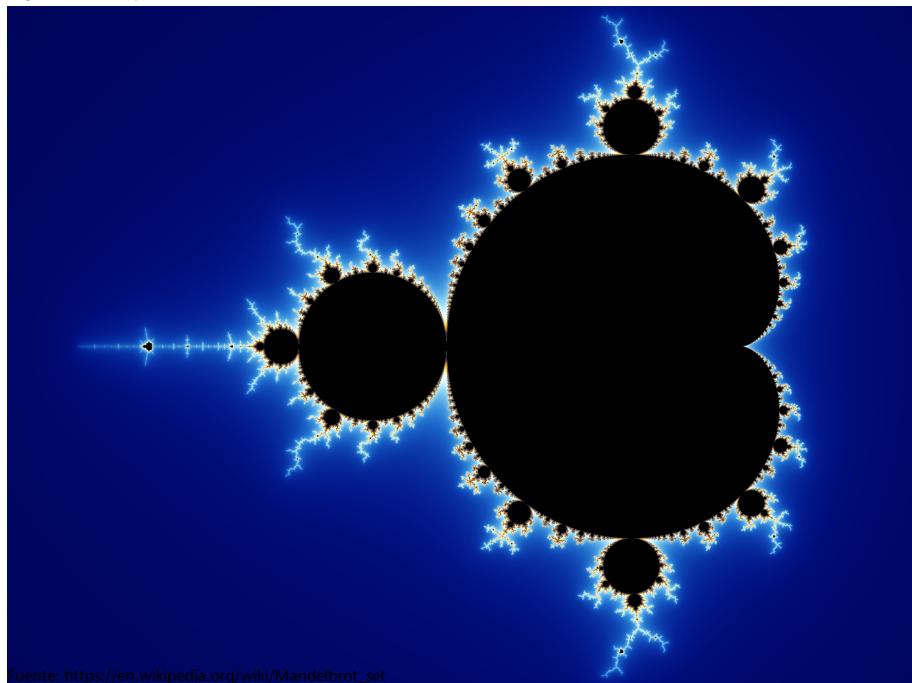
Enlace de interés

Para conocer más sobre los objetos fractales, véase el siguiente enlace:
<https://ca.wikipedia.org/wiki/Fractal>

bert W. Brooks y Peter Matelski, mostrada en la figura 1, usando solamente texto. Sin embargo, fue con la popularización del ordenador personal y sus capacidades gráficas cuando se pudo visualizar por primera vez el conjunto de Mandelbrot en detalle, a unos niveles de resolución insospechados anteriormente. Eso permitió ver su complejidad y descubrir algunas propiedades interesantes (patrones repetitivos pero cambiantes con detalle infinito, o su conectividad, entre otras).

El conjunto de Mandelbrot es un buen ejemplo de cómo la visualización de datos transforma la comprensión de un concepto, posibilitando el descubrimiento de conocimiento hasta el momento desconocido (la forma del conjunto en el plano complejo y el nivel de detalle, en este caso). Concretamente, Benoit Mandelbrot conjeturó que el conjunto que lleva su nombre no era conexo, y fue mediante la visualización del mismo que se descubrieron los pequeños filamentos que hacen que el conjunto de Mandelbrot sea conexo. Sin el uso de ordenadores para visualizarlo con una resolución diferente, hubiera sido imposible.

Figura 2. Conjunto de Mandelbrot en alta resolución



Fuente: https://en.wikipedia.org/wiki/Mandelbrot_set

Por lo tanto, la visualización de un conjunto de datos es normalmente una primera etapa indispensable para comprender mejor su naturaleza, detectando rápidamente aspectos tan intrínsecos a los datos como simetrías, localizaciones, valores extremos, continuidad, suavidad, agrupaciones, etc. Mediante dicha exploración es posible hacerse una idea preliminar que puede permitir afinar los objetivos del análisis y la posterior visualización de los resultados, en un ciclo continuo.

2. La visualización como análisis preliminar

La misma falta de conocimiento sobre la naturaleza del conjunto de Mandelbrot puede suceder con otros datos en general, los cuales son de naturaleza compleja y pueden combinar diferentes aspectos al mismo tiempo, entre otros: ser multidimensionales, ir ligados a restricciones espacio-temporales, longitudinales (que evolucionan en el tiempo), multimodales (combinando diferentes fuentes y orígenes), así como provenir de la ejecución de múltiples procesos paralelos y/o modelos. Visualizar datos incluye gestionar toda esta complejidad para convertirlos en información, es decir, obtener respuestas a las preguntas u objetivos de la visualización. El análisis visual no substituye a la estadística clásica o la construcción de modelos de minería de datos, sino que aporta una perspectiva diferente basada en las capacidades del sistema visual humano.

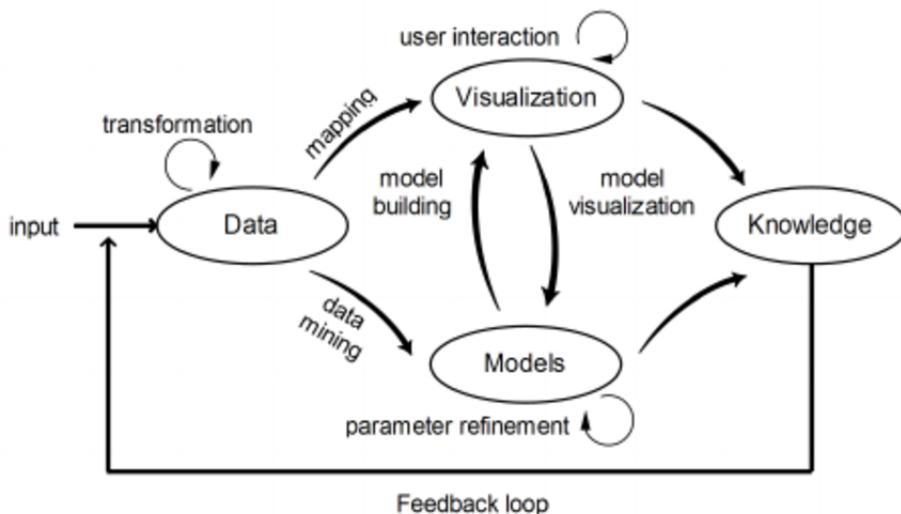
Así, el objetivo de la visualización es mostrar la naturaleza de los datos, facilitando su comprensión y posterior exploración. Se trata, entonces, de realizar un análisis visual preliminar para detectar los aspectos clave presentes en los datos: distribuciones de cada variable, valores extremos, relaciones entre variables, tendencias, patrones, *outliers*, etc. Para ello es necesario poder disponer de un entorno gráfico que permita visualizar datos usando diferentes proyecciones, combinando herramientas estadísticas con modelos generados a partir de los datos, desde descriptores estadísticos hasta el resultado de un algoritmo de clasificación no supervisado, por ejemplo, variando los parámetros del mismo.

En este sentido, la evolución de la visualización de datos no se ha centrado solamente en la capacidad de generar gráficos complejos con mayor resolución en un breve lapso de tiempo, sino que ha ido incorporando elementos interactivos en la propia visualización, en forma de operaciones básicas (selección, filtrado, etc.). De acuerdo al trabajo de Keim y otros publicado en 2008, el análisis visual de datos se fundamenta en un mantra que es una versión modificada del propuesto por Ben Shneiderman en 1996:

*"Analyse First –
Show the Important –
Zoom, Filter and Analyse Further –
Details on Demand"*

Así, el proceso de análisis visual consiste en un ciclo continuo que se inicia en los datos y sus posibles transformaciones, y que se bifurca en dos aproximaciones complementarias, la visualización y la construcción de modelos estadísticos o de minería de datos, entre las cuales existe un diálogo con el objetivo de extraer conocimiento que pueda ser usado para iterar el proceso de análisis visual con un mayor nivel de detalle o complejidad, tal y como muestra la figura 3. La capacidad de interacción debe permitir al usuario de la visualización realizar, al menos, las operaciones básicas definidas por Ben Shneiderman (vista general, zoom, filtro y selección).

Figura 3. Proceso de análisis visual definido por Keim *et al.* (2008)



Desde una perspectiva de análisis visual, las dos primeras etapas definidas en la figura 3 son la transformación (o adaptación) de los datos y su visualización, incluyendo en esta segunda interacción. Por lo tanto, una vez establecido el objetivo del análisis visual de los datos, se trata de seleccionar un tipo de visualización interactiva que permita realizar dicha exploración preliminar.

El conjunto de herramientas estadísticas para analizar, describir y resumir datos es enorme, existiendo diferentes versiones en función de la naturaleza de los datos y el objetivo a alcanzar. Sin pretender ser exhaustivos, un análisis preliminar debe plantearse, al menos, las siguientes preguntas:

- ¿Se trata de analizar una sola variable o bien una combinación de dos o más variables? En función del número de variables que formen parte del análisis (es decir, de la visualización), las posibilidades son muy diferentes, así como el resto de preguntas a plantearse.
- ¿Se trata de variables continuas o categóricas (incluyendo ordinales, nominales y binarias)? Para cada tipo de variable o pareja de variables existe un test estadístico diferente en función del objetivo. Es decir, debemos pensar qué estamos comparando o midiendo.

- ¿Se están analizando las variables originales o bien ha habido una etapa de transformación intermedia y/o son el resultado de un modelo construido con anterioridad? En ambos casos, la visualización debería tener en cuenta dichos procesos intermedios.

No obstante, en este material docente nos centraremos principalmente en datos multidimensionales categóricos, en las que las posibilidades para mostrar la naturaleza de los datos y las relaciones entre una o más variables permiten diferentes representaciones más allá de los típicos histogramas y gráficos de barras, así como una cierta manipulación mediante el uso de elementos interactivos. También veremos, sin embargo, algunos ejemplos de datos de carácter más cuantitativo, combinados con datos categóricos.

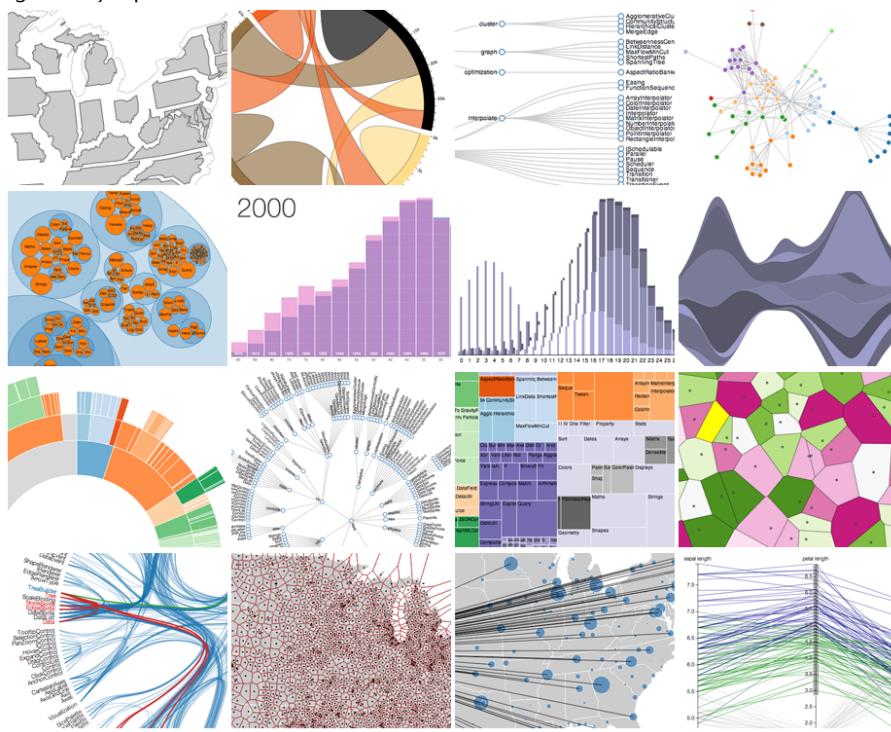
Es importante recordar que un análisis visual no sustituye a un análisis estadístico, dado que visualmente no somos capaces de determinar con la precisión necesaria si dos cantidades son ligeramente parecidas o diferentes. Puede decirse que es un análisis rápido de «baja resolución», que nos permite hacernos preguntas que después trataremos de responder con las herramientas estadísticas y algorítmicas adecuadas.

3. Herramientas para el análisis visual de datos

La creación de visualizaciones interactivas puede abordarse desde dos aproximaciones diferentes: la primera es utilizar herramientas de software diseñadas específicamente para ello, siendo QlikView, Tableau y Power BI de las más populares en la actualidad, principalmente debido a su potencia y flexibilidad, aunque realmente nos limitan las opciones y el aspecto final a los determinados por la herramienta. La segunda aproximación es utilizar las capacidades gráficas de algún lenguaje de programación o librería para mostrar datos de forma más o menos interactiva, como D3 (mostrado en la figura 4), Processing o R+ggplot2, entre otros*, dependiendo en este caso de un desarrollo *ad hoc* que puede ser costoso y complejo, ya que hacen falta ciertos conocimientos de programación. Una tercera vía que combina ambas aproximaciones es la utilización de herramientas software que generan visualizaciones de datos en forma de código que puede ser ejecutado independientemente de la herramienta usada para crearlas, como Quadrigram, por ejemplo.

*<https://goo.gl/PnOq4a>

Figura 4. Ejemplos de diferentes visualizaciones creadas con D3



Fuente: <https://github.com/d3/d3/wiki/Gallery>

En la actualidad, la visualización de datos interactiva cuenta con un nuevo aliado, el cual elimina la necesidad de crear aplicaciones específicas y nos dota de una interfaz visual coherente para su cometido. Se trata de los navegadores

web, que visualizan páginas que contienen un código fuente que construye (mediante un proceso de renderización) la visualización cuando la página es accedida y cargada. Una página web es una combinación de CSS (hojas de estilo que determinan el aspecto de los elementos de la página), contenido HTML propiamente dicho y código JavaScript que permite manipular el DOM (del inglés Document Object Model, es decir, la estructura de la propia página web vista como un documento estructurado jerárquicamente), generando nuevos contenidos que se incrustan dinámicamente, incluyendo código HTML y gráficos vectoriales (SVG, o Scalable Vector Graphics). El gráfico (o mejor dicho, la forma de construirlo) es parte de la página, y es visualizado cuando el navegador ejecuta las órdenes necesarias para mostrar el contenido de la página.

De forma análoga, últimamente se ha desarrollado el concepto de *notebook*, que no es más que un documento que puede contener código que puede ejecutarse en el navegador o en un entorno de programación, que muestra los resultados como parte del propio documento e incluye la creación de visualizaciones de datos.

De esta manera, generar una visualización de datos interactiva puede verse como la creación de un documento o de una página web construida dinámicamente y que visualiza dichos datos de acuerdo a una configuración preestablecida. En este sentido, D3 (o también D3.js) es una librería JavaScript que permite manipular datos en diferentes formatos (tablas, CSV o JSON, entre otros) y generar gráficos vectoriales de forma dinámica que pueden ser incrustados en la página web para su manipulación, incluyendo elementos de interactividad, tanto por lo que respecta a la interfaz del usuario como al uso de transiciones que aportan dinamismo a la visualización (Murray, 2013). D3 se puede usar dentro de un *notebook* Observable sin problema y a este se puede acceder desde un navegador, simplificando la carcasa HTML+CSS necesaria para visualizarlo.

Dada su flexibilidad, D3 puede utilizarse para crear cualquier tipo de visualización interactiva, generando los elementos gráficos a partir de los datos que alimentan la visualización, desde gráficos de barras simples hasta complejas visualizaciones combinando diferentes elementos gráficos (Bostock *et al.*, 2011). Uno de los aspectos más interesantes de D3 es la incorporación de la interacción como parte de la propia visualización, de forma que esta deviene la interfaz de acceso a los datos, permitiendo su manipulación de acuerdo a las operaciones básicas definidas por Shneiderman (1996).

Las posibilidades que ofrece D3 para visualizar datos categóricos (principalmente, pero también de cualquier otro tipo) son muy interesantes, justificando su elección para un análisis exploratorio inicial e, incluso, ser el esqueleto inicial de una visualización compleja. Para ello dispone de un catálogo de configuraciones (*layouts*) muy extenso, como se verá a continuación.

4. *Layouts* D3 para el análisis visual de datos

Como ya se ha comentado, D3 ofrece una buena colección de *layouts* ya predefinido que pueden utilizarse para crear visualizaciones interactivas con relativamente poco esfuerzo, tal y como muestra la figura 4.

Los *layouts* son configuraciones predeterminadas que permiten crear gráficos.

La idea básica es reutilizar las visualizaciones ya existentes para adaptarlas a los datos que desean visualizarse (y viceversa, adaptar los datos a los requerimientos de la visualización), de forma que se utiliza cada *layout* como un «molde» o «caja negra» que transforma datos en una visualización interactiva.

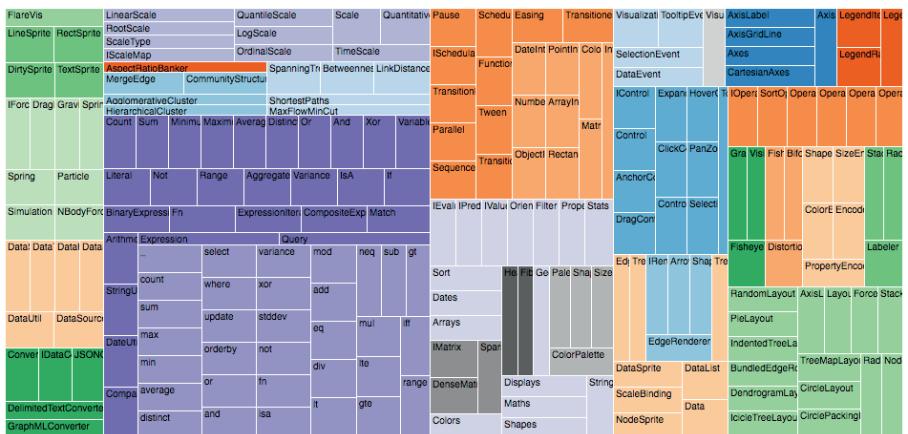
A continuación se muestra una selección de *layouts* de D3 obtenidos de la propia galería (<https://github.com/d3/d3/wiki/Gallery>), que permiten representar datos y realizar una exploración sencilla de los mismos mediante el uso de elementos interactivos integrados en la propia visualización. Como complemento a este material docente, se ha preparado un ejemplo para cada uno de estos *layouts*, que incorpora un mayor nivel de detalle y una visualización interactiva sobre unos datos reales que pueden ser descargados para reutilizar la visualización. Hemos mantenido los nombres originales en inglés porque es como se conocen estos gráficos en el ámbito de la visualización.

Enlace de interés

Podéis encontrar algunos ejemplos de *layouts* en la página siguiente:
<https://datavizcatalogue.com/>, así como ejemplos interactivos en
<http://oer.uoc.edu/VIS/D3/>

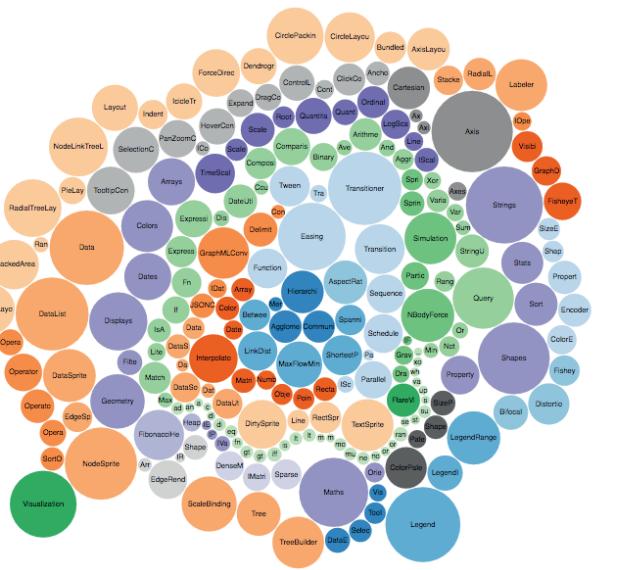
4.1. Treemap

Este tipo de visualización permite observar cómo se distribuyen los valores de una o más variables categóricas en un área, normalmente rectangular, permitiendo realizar comparaciones sencillas por tamaño o proporción, así como detectar la existencia de patrones en el caso de combinar dos o más variables. Por ejemplo, se puede visualizar cómo se distribuyen los presupuestos de un país en las diferentes partidas que lo componen, observando claramente qué apartados o regiones se llevan la mayor parte.

Figura 5. Ejemplo de visualización de *treemap*

4.2. Bubble

Se trata de una visualización jerárquica que permite visualizar las relaciones entre una o más variables categóricas, de forma similar a un *treemap* pero con una representación basada en círculos. El aprovechamiento del espacio es menor que en el caso del *treemap*, pero las estructuras jerárquicas presentes en los datos pueden detectarse mejor, especialmente cuando la estructura jerárquica no está equilibrada. Además, visualmente resultan más atractivas que los *treemaps*, que pueden quedar muy apretados si hay muchas variables o categorías dentro de cada variable.

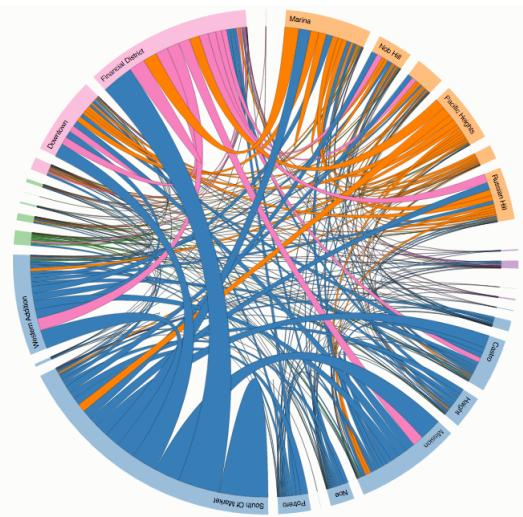
Figura 6. Ejemplo de visualización de *bubble*

4.3. Chord

En este caso se trata de una visualización radial que utiliza una corona circular donde se distribuyen los valores de una variable categórica, de acuerdo a su probabilidad o frecuencia. Entre cada par de valores existe un arco (*chord*)

que permite conocer la relación entre ambos valores, también proporcional al número de elementos que participan en dicha combinación. Por ejemplo, es posible visualizar las combinaciones más comunes que aparecen en las matrículas de los estudiantes, analizando las parejas de asignaturas más frecuentes. Es otra manera de representar una tabla de contingencia entre dos variables.

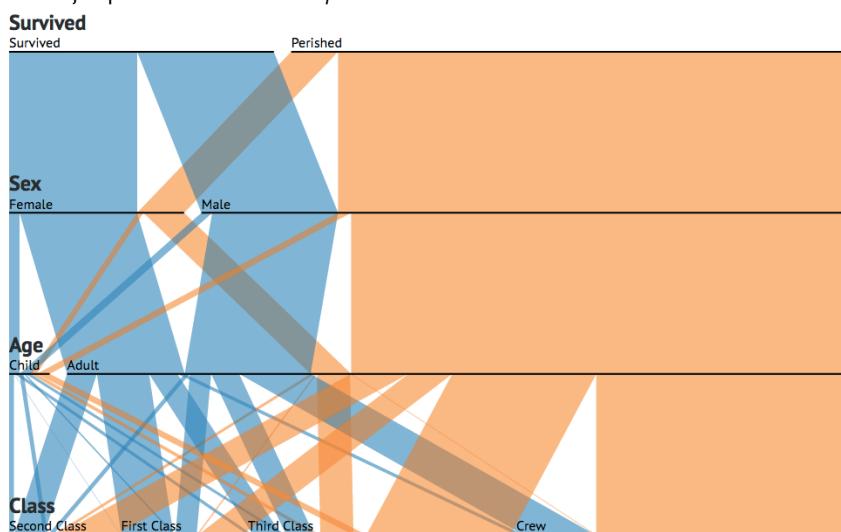
Figura 7. Ejemplo de visualización de *chord*



4.4. Parallel sets

Son útiles para representar datos categóricos, permitiendo visualizar tanto las frecuencias de aparición de cada valor para cada variable como las combinaciones de valores entre variables. En este segundo caso, se pueden utilizar como un análisis visual complementario al uso de tablas de contingencia para más de dos variables. Un ejemplo típico es la segmentación de estudiantes respecto a sus atributos: género, grupo de edad, estudios previos, carrera escogida, etc.

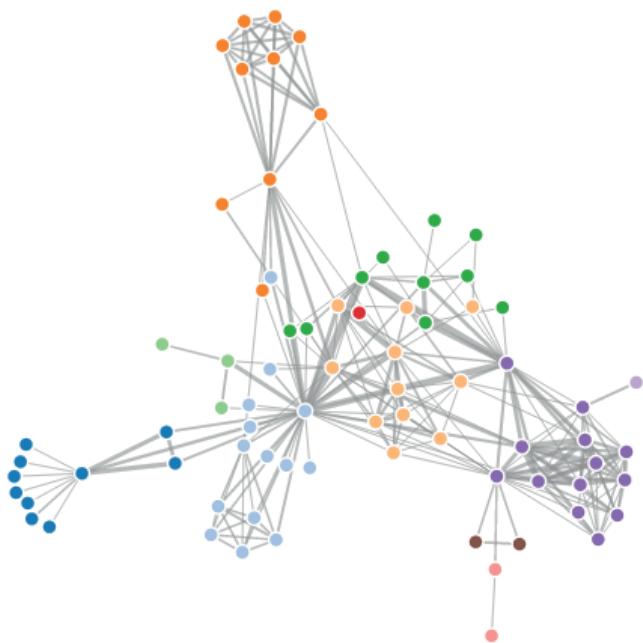
Figura 8. Ejemplo de visualización de *parallel sets*



4.5. Force-directed graphs

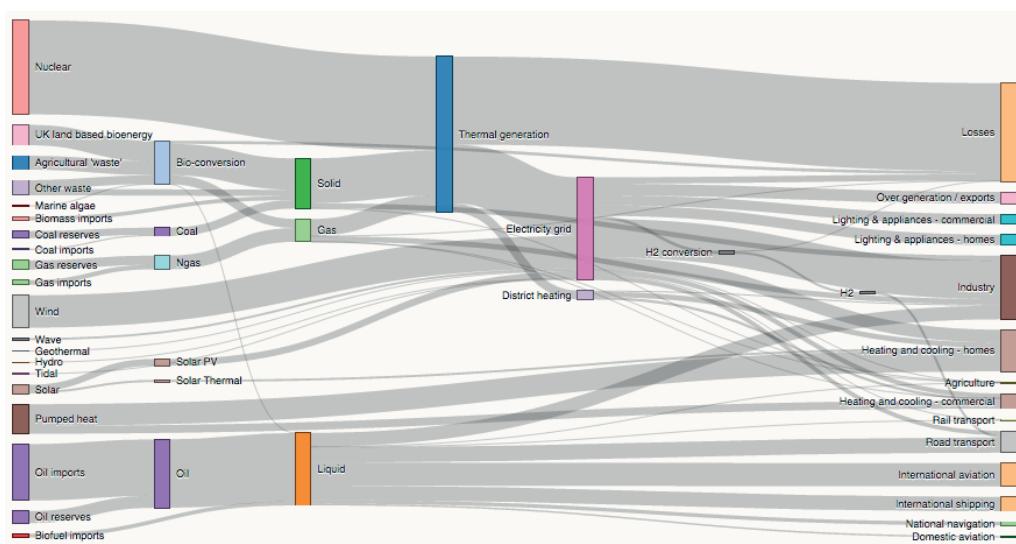
A veces es más interesante visualizar las relaciones entre elementos que los atributos de los elementos en sí mismos. Los grafos permiten visualizar dichas relaciones, utilizando una disposición que refleja, de cierta manera, la distancia (entendida como similitud o disimilitud) entre los diferentes elementos, pero de forma relativa, no absoluta. La visualización muestra los diferentes grupos o clústeres que se forman, permitiendo establecer categorías de forma sencilla. Por ejemplo, es posible visualizar un grafo que visualiza las coautorías entre autores que publican conjuntamente, donde cada autor es un nodo del grafo y cada arco entre nodos representa el número de veces que un autor ha compartido autoría con otro. Los elementos también se pueden representar de color y tamaño diferente, en función de uno o más atributos que se quieren representar en el grafo.

Figura 9. Ejemplo de visualización de *force-directed graphs*



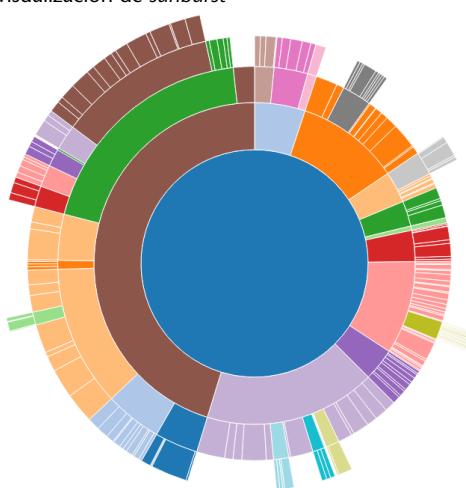
4.6. Sankey

En este caso se trata de visualizar un flujo de datos entre diferentes conceptos de acuerdo a ciertos criterios, normalmente espaciales o temporales, aunque también puede usarse para mostrar cambios del valor de una o más variables en función de otra, normalmente el tiempo. En este caso, permite comparar valores que puede tomar una variable respecto al total, mostrando las proporciones de cada posible valor.

Figura 10. Ejemplo de visualización de *sankey*

4.7. Sunburst

Este tipo de visualización es equivalente al *treemap* pero utilizando una disposición radial en lugar de rectangular. Al igual que *bubble*, el aprovechamiento del espacio es menor, pero la disposición radial conlleva un orden que puede aprovecharse para mostrar los elementos de acuerdo a algún parámetro. Por otra parte, el hecho de alejarse del centro también implica una cierta ordenación que permite establecer prioridades entre los diferentes elementos visualizados. Finalmente, se pueden mostrar diferentes niveles de profundidad para cada segmento, por lo que es una representación más apta que *treemap* para estructuras jerárquicas muy desbalanceadas.

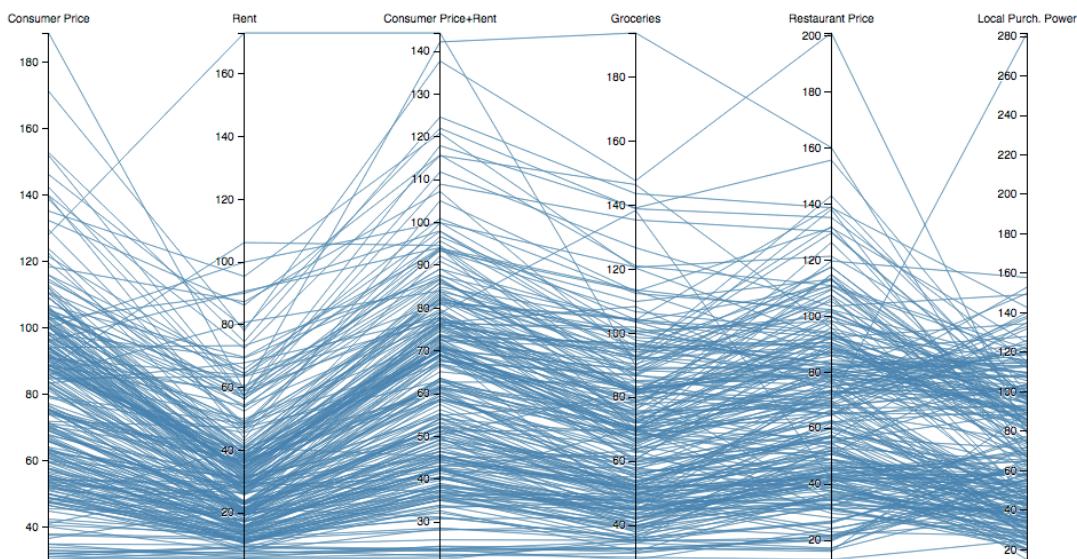
Figura 11. Ejemplo de visualización de *sunburst*

4.8. Parallel coordinates

Aunque a primera vista pueda parecer confusa, esta visualización es muy utilizada dado que aprovecha las capacidades del sistema visual humano para detectar patrones y tendencias usando líneas y posiciones relativas, facilitando

su interpretación en clave de la correlación entre variables. Permite representar un espacio multidimensional más allá de la típica representación en dos dimensiones ortogonales, de ahí su nombre de 'coordenadas paralelas'.

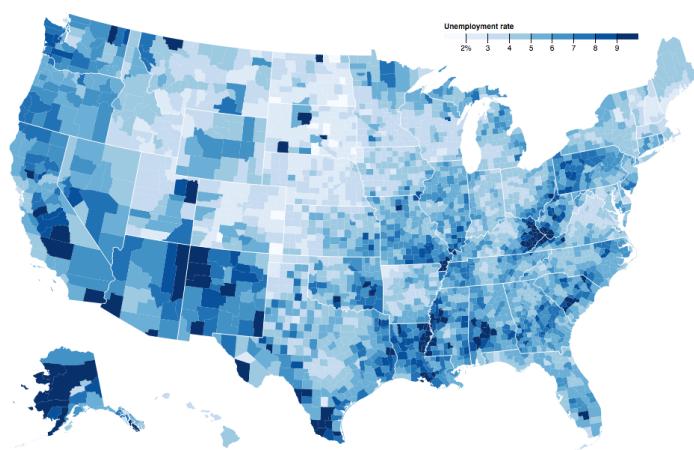
Figura 12. Ejemplo de visualización de *parallel coordinates*



4.9. Choropleth

En muchos casos los datos a visualizar tienen un origen relacionado con un área o región geográfica. Este tipo de visualización permite superponer una gradación de color a un mapa, mostrando para cada segmento un valor posible en forma de la intensidad del color. De esta manera es posible detectar fácilmente agrupaciones donde se concentran los valores extremos de la variable utilizada para la visualización.

Figura 13. Ejemplo de visualización de *choropleth*



4.10. Word cloud

Una alternativa a la representación de las frecuencias absolutas usando gráficos de barras o histogramas es el uso de lo que se conoce como *word cloud* o

nube de palabras. Se trata de una visualización no reduccionista, dado que se muestran los elementos del conjunto de datos tal cual, en este caso, las palabras de un texto. Cuantas más veces aparece una palabra en un texto, más grande aparece representada en el *word cloud*, de forma que es fácil identificar rápidamente cuáles son los términos más populares. Es, por tanto, una especie de histograma de frecuencias que sustituye las barras por los tamaños de las palabras representadas. Otros atributos como el color o, incluso, la tipografía pueden usarse para mapear otras características del conjunto de datos, mientras que la posición y la rotación suelen ser libres, y vienen dadas por la forma final que quiere obtenerse, es decir, de la propia nube. Se utilizan mucho para resumir textos y hacerse una primera idea al respecto.

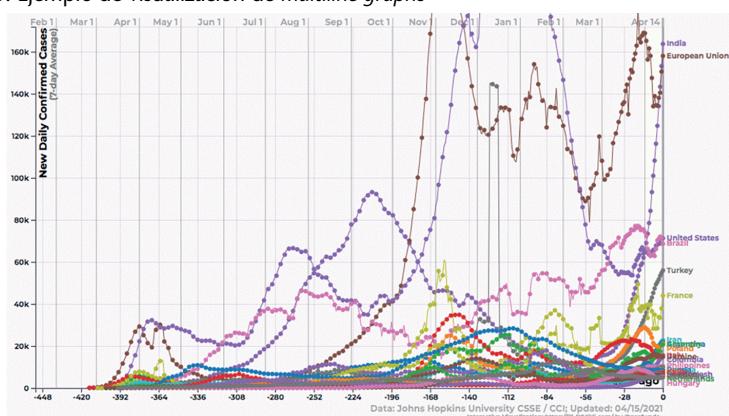
Figura 14. Ejemplo de visualización de *word clouds*



4.11. Multiline graphs

Los gráficos de línea múltiples son aquellos en los que diferentes variables comparables entre sí se dibujan usando el mismo eje X o Y y la misma escala, de forma que sea posible al mismo tiempo detectar patrones y comparar tendencias, normalmente a lo largo del tiempo. En este caso, las variables que representar suelen ser cuantitativas y, en función de la naturaleza de los datos, puede darse la opción de usar una escala lineal o logarítmica para representar crecimientos en lugar de cantidades.

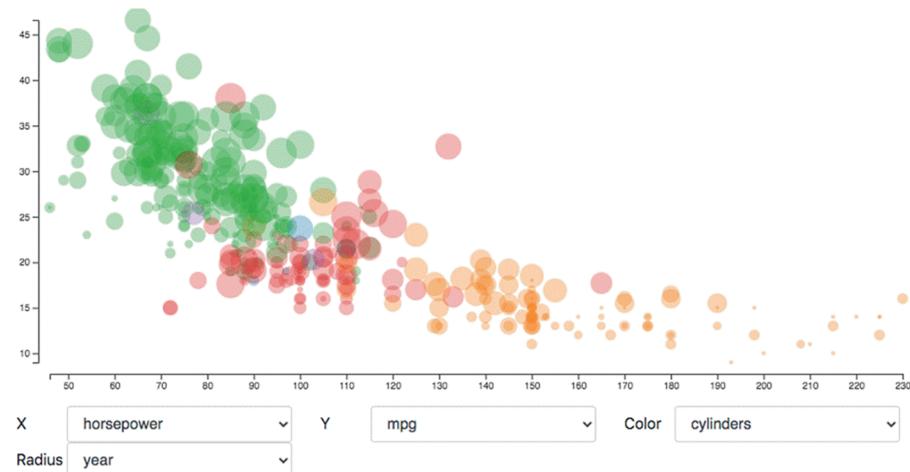
Figura 15. Ejemplo de visualización de *multiline graphs*



4.12. Scatterplot

Para representar la relación entre dos (o a veces tres) variables numéricas, se utiliza lo que se conoce como un *scatterplot*. Se trata de un gráfico muy sencillo en el que se dibujan los puntos x e y del conjunto de datos original, donde x e y son las dos variables de las que quiere analizarse su comportamiento conjunto, en el caso de dos dimensiones. Para cada combinación de x e y , se dibuja un punto en el gráfico, de forma que rápidamente pueden verse las agrupaciones, las tendencias y también los datos atípicos. Además, para cada punto, pueden mapearse otras variables del conjunto original en forma de sus atributos, que pueden ser la forma, el tamaño y el color, por ejemplo, de forma que pueden construirse gráficos muy complejos que nos permiten ver la naturaleza del conjunto de datos que estamos analizando. Si el usuario puede elegir las variables que se mapean en cada uno de los ejes y atributos, se trata de una herramienta muy poderosa para hacer un análisis visual preliminar.

Figura 16. Ejemplo de visualización de *scatterplot*



Bibliografía

Bostock, M.; Ogievetsky, V.; Heer, J. (2011). D³ data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12), 2301-2309.

Brooks, R., Matelski, J. P. (1981). «The dynamics of 2-generator subgroups of PSL (2, C)». En: *Riemann surfaces and related topics: Proceedings of the 1978 Stony Brook Conference* (vol. 97, págs. 65-71). Princeton Press University.

Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., Melançon, G. (2008). «Visual analytics: Definition, process, and challenges». En: *Information visualization* (págs. 154-175). Springer Berlin Heidelberg.

Murray, S. (2013). *Interactive data visualization for the Web*. O'Reilly Media, Inc.

Shneiderman, B. (1996). «The eyes have it: A task by data type taxonomy for information visualizations». En *Proceedings IEEE Symposium on Visual Languages* (págs. 336-343). IEEE.

Páginas web

- D3: <https://d3js.org>
- QlikView: <http://www.qlik.com/es>
- Tableau: <http://www.tableau.com/>
- Processing: <https://processing.org/>
- Ggplot2: <https://ggplot2.tidyverse.org/>
- Quadrigram: <http://www.quadrigram.com/>
- The Data Visualization Catalogue: <http://www.datavizcatalogue.com/>