

# La disciplina de la visualización de datos en 2017

## The discipline of data visualization in 2017

Mario Pérez-Montoro

**Pérez-Montoro, Mario** (2018). "La disciplina de la visualización de datos en 2017". *Anuario ThinkEPI*, v. 12, pp. 216-224.

<https://doi.org/10.3145/thinkepi.2018.32>

Publicado en *IweTel* el 22 de diciembre de 2017



**Resumen:** El avance tecnológico y conceptual está provocando cambios en la disciplina de la visualización de información. El objetivo de este trabajo es revisar algunas de las visualizaciones innovadoras recientes. Para abordarlas se han organizado en tres grupos. Por un lado, presentaremos una selección de las que han destacado por incorporar novedades conceptuales y nuevas propuestas de visualización. A continuación, mostraremos otras que han llamado la atención por la importancia del conjunto de datos que han visualizado. Por último, daremos cuenta de algunos de los avances tecnológicos que se han producido dentro de la disciplina a lo largo de 2017.

**Palabras clave:** Visualización de datos; 2017; Tendencias; Visualización de información; Visualización interactiva; Infografía; Visualizaciones multimedia; Conceptos; Contenidos; Tecnología.

**Abstract:** Recent advances in technology have provoked changes in the information visualization discipline. This article reviews some innovative visualizations and organizes them into three groups. First, we present a small selection of visualizations that incorporate conceptual novelties and new visual proposals. Second, we show some other ones that have drawn attention to the importance of the data set they have visualized. Finally, we account for some of the technological advances that were made in 2017 within the discipline.

**Keywords:** Data visualization; 2017; Trends; Information visualization; Interactive visualization; Infographics; Multimedia visualization; Concepts; Contents; Technology.

## 1. Introducción

En la edición anterior del *Anuario ThinkEPI* (Pérez-Montoro, 2017) inauguramos una serie de trabajos encaminados a mostrar algunas de las propuestas visuales desarrolladas durante el período de un año y que destacaban por sus planteamientos innovadores y su contribución a la evolución de la propia disciplina de la visualización de la información.

El objetivo de este trabajo es revisar algunas de esas visualizaciones innovadoras llevadas a cabo a lo largo de 2017. Para abordarlas, recuperaremos la clasificación que defendimos en el anterior trabajo. Así, por un lado, presentaremos

una selección de aquellas que han destacado por incorporar novedades conceptuales y nuevas propuestas de visualización. A continuación, mostraremos otras que han llamado la atención por la importancia del conjunto de datos que han visualizado. Por último, daremos cuenta de los principales avances tecnológicos que se han desarrollado dentro de la disciplina a lo largo de este año.

## 2. Conceptos

Comencemos revisando aquellas visualizaciones que queremos destacar por aportar nuevas soluciones conceptuales a la hora de representar visualmente datos.

Dentro de este grupo, en 2017, destacan visualizaciones relacionadas con:

- nuevas propuestas representacionales de eventos temporales anuales;
- análisis temático de textos;
- trazabilidad de eventos temporales que presentan variaciones en un ranking;
- examen visual del comportamiento de votantes en unas elecciones.

La primera de las visualizaciones que queremos destacar en este apartado surge del trabajo realizado en el seno del *Google News Lab*. Desde 2015, en esta iniciativa de *Google* se intenta asociar de forma coordinada a periodistas y artistas para concebir y desarrollar, utilizando como materia prima los datos generados en su buscador, reportajes dotados de una elevada creatividad visual.

En este contexto, Moritz Stefaner (especialista en visualización de información) decide emprender el proyecto *The rhythm of food*. <http://rhythm-of-food.net>

Junto a Simon Rogers y Alberto Cairo (*Google News Lab*), Yuri Vishnevsky (desarrollo) y Stefanie Weigele (ilustración), en este proyecto Stefaner intenta representar el “ritmo” de la búsqueda de información que presentan los usuarios de *Google* respecto al tema de la comida. La iniciativa comienza mostrando, utilizando la herramienta *Google Trends* y mediante el uso de gráficas de líneas (coloquialmente denominadas “fiebres”), las diferentes tendencias en el tiempo de las

búsquedas relacionadas con los principales tipos de dietas.

<https://trends.google.com/trends>

En esas gráficas podemos comprobar cómo va aumentando la curiosidad por las dietas veganas o la dieta mediterránea, o cómo la dieta baja en grasa ya no suscita el interés de antaño.

También muestra, mediante el uso de histogramas (gráficas de columnas continuas), la evolución anual de las búsquedas sobre un alimento concreto. Esos histogramas alimentan de forma dinámica lo que denominan “un reloj anual” (*year clock*) en el que se muestra como resultado final las búsquedas agrupadas por mes (figura 1). Esas agrupaciones permiten, entre otras cosas, identificar con claridad las temporadas naturales de cada alimento y explorar qué podemos consumir (como alimento de temporada) en cada uno de los meses que componen un año.

La segunda visualización que queremos destacar es *Inaugurate*, otro proyecto nacido en el *Google News Lab*.

<http://inauguratespeeches.com>

En esta iniciativa se intenta representar los discursos de investidura o inaugurales de los presidentes estadounidenses más buscados en *Google* desde 2004: George Washington, Thomas Jefferson, Abraham Lincoln, Franklin Roosevelt, John Kennedy, Richard Nixon, Ronald Reagan, George H. W. Bush, Bill Clinton, George W. Bush, Barack Obama y Donald J. Trump.

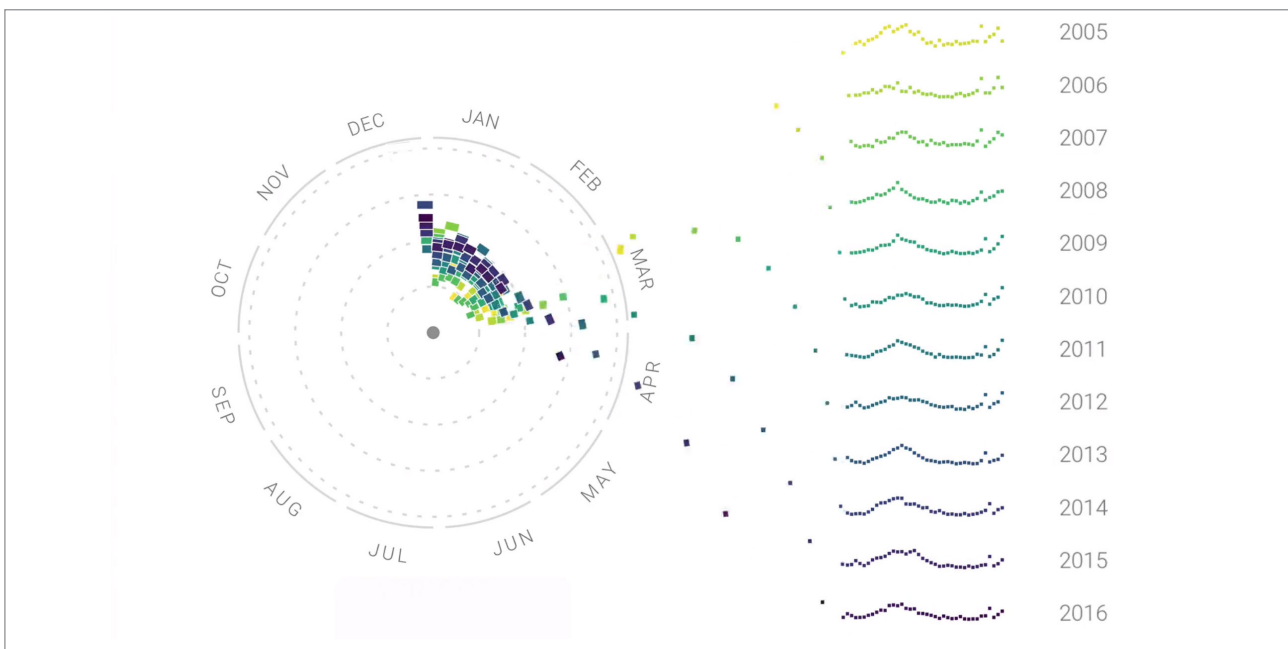


Figura 1. Transformación de los histogramas (tendencias mensuales por año de las búsquedas sobre albaricoques) en un diagrama en forma de reloj anual que pone el acento en las búsquedas sobre esa fruta agrupadas por mes *The rhythm of food*. <http://rhythm-of-food.net>

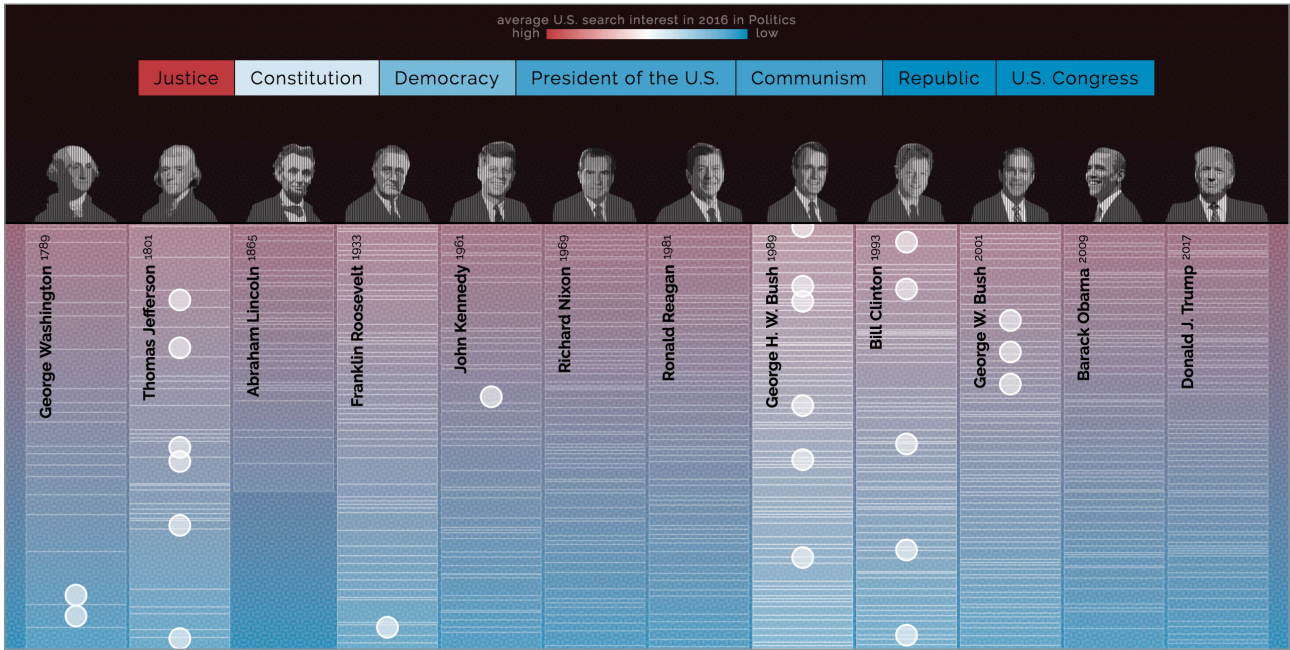


Figura 2. Visualización de los discursos de investidura analizando los temas sobre política y comparando esos temas con el interés de búsqueda actual  
Inaugurate. <http://inauguratespeeches.com>

Con esa visualización se consigue extraer los temas más comunes mencionados en los discursos (por la frecuencia de sus términos). Esos temas se organizan bajo una clasificación (política, finanzas, emociones, religión, sociedad y otros) y se comparan (mediante un código de color) con el interés de búsqueda promedio de los datos de búsqueda de Google en 2016, comprobando la

vigencia actual de esos mismos temas (figura 2).

Para poder comprenderlo de forma adecuada hay que entender que:

- cada rectángulo representa una oración (se puede obtener la oración pasando el cursor por encima);
- la altura de un rectángulo se basa en el número de palabras en una oración;

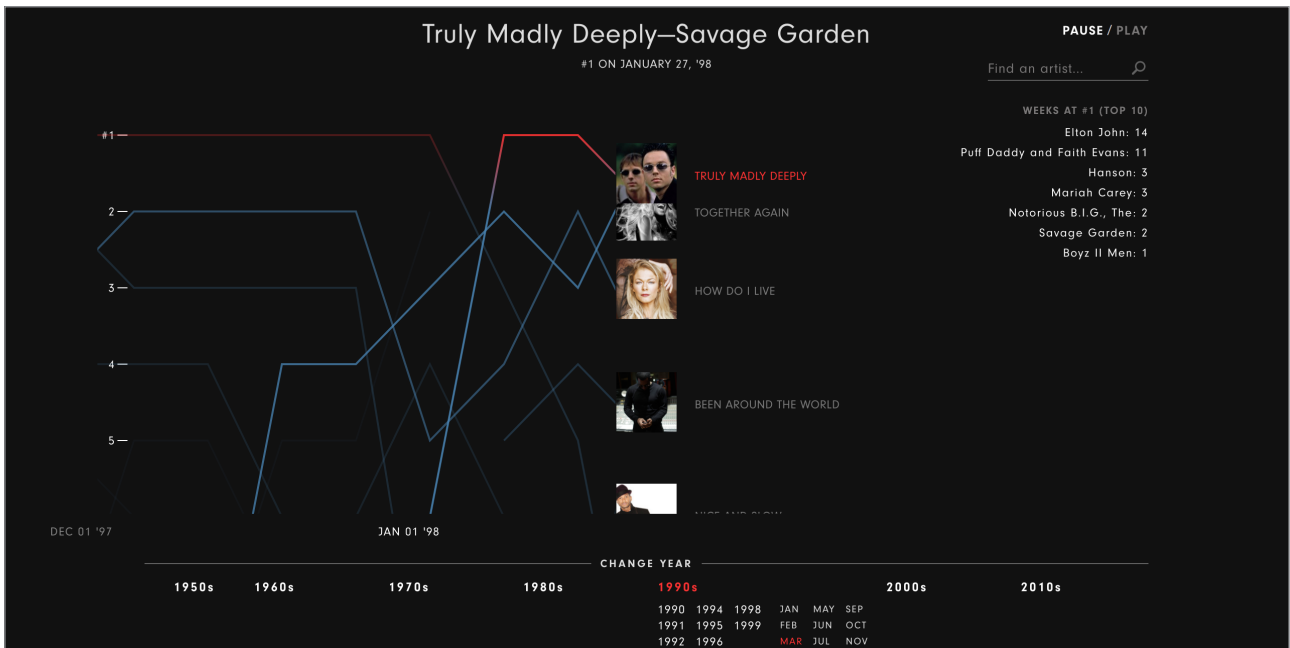


Figura 3. Visualización de la evolución de las 5 primeras canciones de la lista de éxitos de la primera semana de enero de 1998

How music taste evolved. <https://pudding.cool/2017/03/music-history/index.html>

- una columna de oraciones representa un discurso de inauguración completo;
- la altura de una columna representa el número total de palabras en un discurso;
- cada círculo representa la mención de un tema (ofreciéndose más información al respecto al pasar el cursor sobre el círculo);
- la posición vertical de un círculo representa su ubicación en el discurso;
- algunos círculos presentan marcas adicionales que señalan que ese es el tema más mencionado en ese discurso.

La tercera visualización que queremos destacar ha sido realizada por el equipo de *The pudding* e intenta representar de forma original la evolución de los números uno de las listas de éxitos musicales desde 1958 hasta 2016. Concretamente, una vez que le damos al *play* y elegimos un año y mes, la página muestra en una columna en forma de ranking las 5 primeras canciones de la lista de éxitos de esa semana utilizando la carátula del disco al cual pertenece como representación de cada una de esas canciones. El número uno de esa semana suena de fondo hasta que es desbancado por otro tema en alguna semana posterior. Cada canción, al avanzar las semanas, va dejando un rastro en forma de línea que permite observar la evolución de su posición en el ranking desde que entra en el top 5 hasta que desaparece de mismo (figura 3).

La última visualización que queremos destacar en este primer apartado ha sido publicada por el periódico *The New York times* en su sección

de interactivos. Se trata de una visualización integrada en un artículo (*How Britain voted*) que analiza cómo votaron los británicos en las elecciones generales de 2017 y en las que eligieron a Theresa May como primera ministra.

Se trata de una gráfica en la que se ubican, en el espacio generado por dos ejes cartesianos, flechas con diferente orientación, tamaños y colores. En el eje de abscisas (X) se ubica la edad media de los votantes; en el de ordenadas (Y), el porcentaje de votantes. Cada flecha representa el porcentaje de votantes de una edad media concreta en una localidad determinada. El color rojo nos indica que han votado a los laboristas; el azul, a los conservadores. Si la flecha apunta hacia arriba, el porcentaje de votantes ha aumentado respecto a las elecciones anteriores; si apunta hacia abajo, éste ha disminuido. Esas diferencias en el porcentaje se representan por la longitud de la flecha y su valor exacto (el valor actual y el anterior) nos aparece si pasamos el ratón por encima de la misma. Por la distribución de colores, vemos que el voto de la población de mayor edad sustenta la candidata conservadora, mientras que los jóvenes votan la de los laboristas (figura 4).

### 3. Contenidos

El segundo de los apartados lo vamos a dedicar a resaltar una serie de visualizaciones que han destacado por intentar representar de forma visual conjuntos de datos que despiertan el interés de una forma intrínseca.

La primera de las visualizaciones que queremos destacar aborda uno de los temas sociales

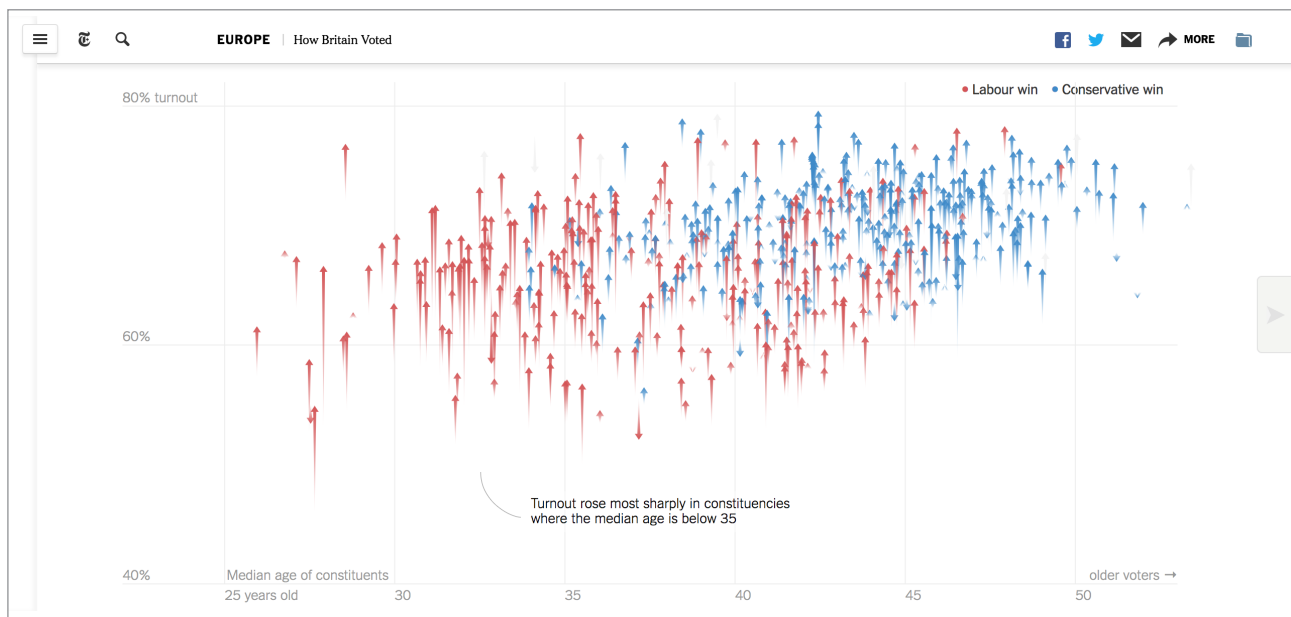


Figura 4. Visualización que analiza cómo votaron los británicos en las elecciones generales de 2017

*How Britain voted*

<https://www.nytimes.com/interactive/2017/06/08/world/europe/british-general-election-results-analysis.html>

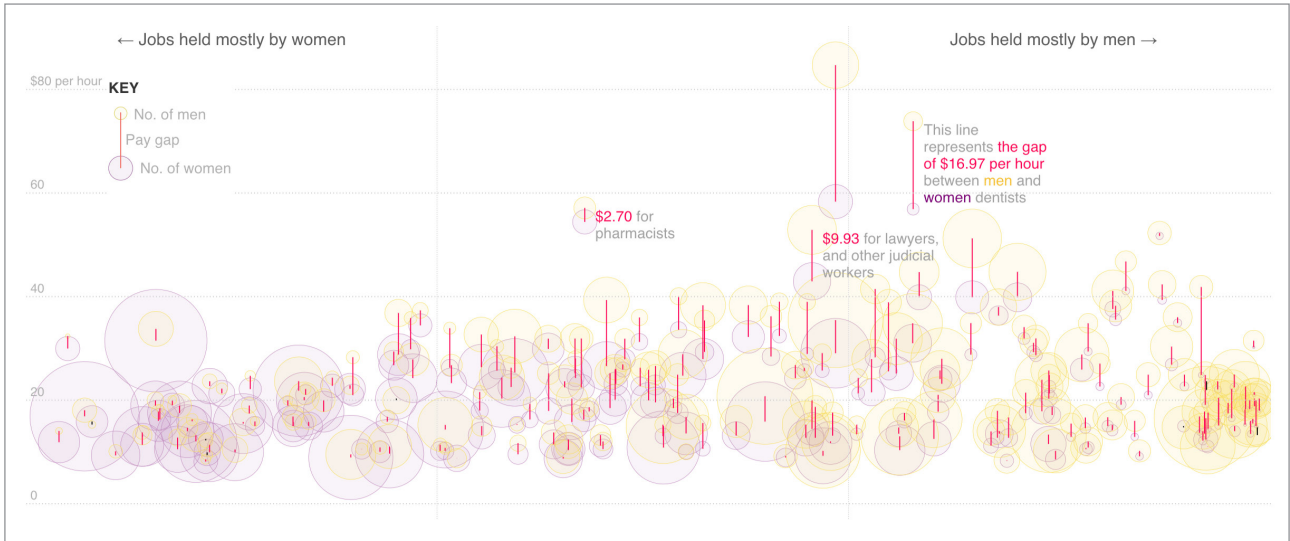


Figura 5. Análisis visual de la brecha salarial (en el precio por hora) entre hombres y mujeres por profesión  
*Can we talk about the gender pay gap?*  
<https://goo.gl/cXgNof>

más importantes: las diferencias salariales de género en el contexto laboral. Se trata de un trabajo publicado en el periódico estadounidense *The Washington post* y creado por el español Xaquín G. V. (reputado periodista interactivo, de datos y visual que ha dirigido el departamento de *Visuals* en *The guardian* y ha trabajado en el *National geographic*, *The New York times*, *Newsweek* y *El mundo*).

En este trabajo periodístico (*Can we talk about the gender pay gap?*) se analiza de forma visual (mediante la estrategia narrativa del *scroll telling* (navegación mediante el uso del cursor)) las diferencias salariales que sufren las mujeres frente a los hombres. Así, muestra que por la brecha de sueldo en un mismo puesto de trabajo, las mujeres trabajan (de media) más de dos meses gratis al año, las

diferencias del precio por hora del trabajo de una mujer frente a un hombre por profesión (figura 5) o cómo la formación universitaria no consigue equilibrar esas diferencias, entre otras cosas.

La segunda de las visualizaciones que queremos destacar analiza el grado de fidelidad a la realidad que presentan una serie de películas basadas en hechos reales. Se trata de un trabajo publicado en la página *Information is beautiful*, recurso de referencia en el campo de la visualización de información. En esta propuesta se seleccionan una serie de películas basadas en hechos reales, mostrándonos su porcentaje de fidelidad a esos hechos (*Based on a true story?*). Las películas son representadas como una barra generada por la agregación de pequeños rectángulos que representan cada una de las escenas que conforman

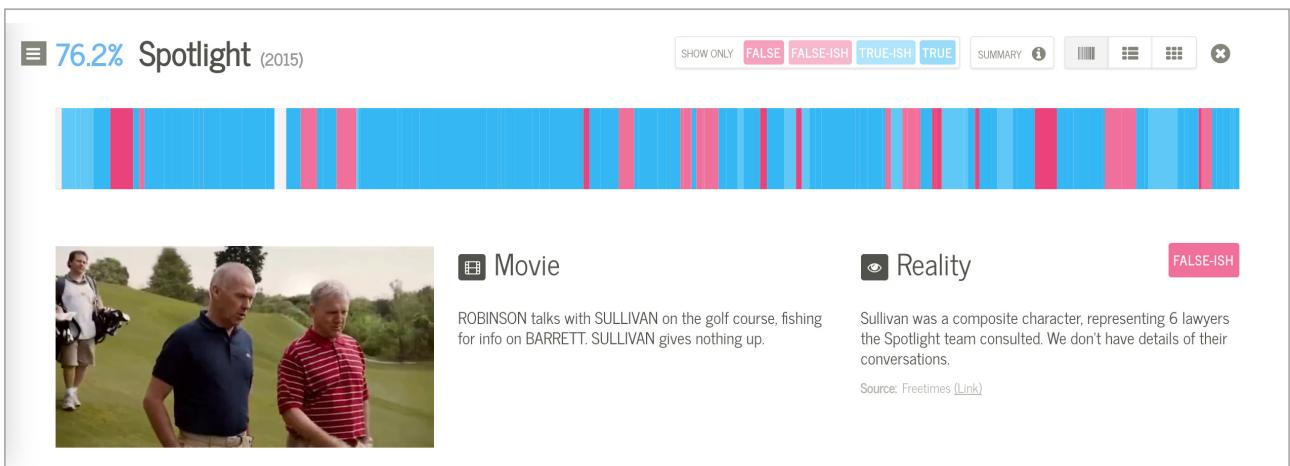


Figura 6. Visualización del grado de fidelidad a la realidad por escena de la película *Spotlight* sobre los actos de pederastia de la iglesia católica en Boston  
*Based on a true story?* <http://www.informationisbeautiful.net/visualizations/based-on-a-true-true-story>

esa obra audiovisual. A partir de una pequeña leyenda de colores, podemos identificar qué escenas recogen fielmente la realidad, cuáles se acercan a ésta, cuáles son recreaciones libres y cuáles, aun siendo recreaciones libres, incluyen algún tipo de referencia real (figura 6). Si colocamos el cursor sobre una de esas escenas, se muestran los personajes que participan, la trama que aborda y el hecho real con el que está relacionado.

La tercera de las visualizaciones que destacamos en este apartado puede ser considerada como una propuesta de meta-análisis

de los principales productos visuales que se suelen utilizar a la hora de representar información. De nuevo bajo el paraguas del *Google News Lab*, la diseñadora Anna Vital nos presenta un gráfico interactivo, denominado *The visualization universe* (figura 7).

<http://visualizationuniverse.com>

Con este trabajo nos muestra cómo los diferentes tipos de gráficos han variado en popularidad en los últimos 12 meses, según la cantidad de personas que buscaron ese gráfico

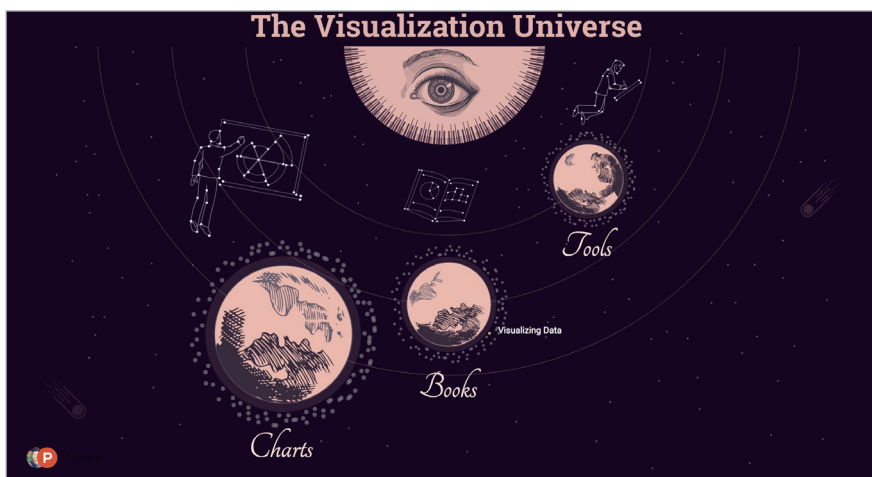


Figura 7. El proyecto *The visualization universe* de la diseñadora Anna Vital

*The visualization universe*

<http://visualizationuniverse.com>

en *Google*. La propuesta está organizada para que se pueda navegar por diferentes criterios como el tipo de gráfico, libros, herramientas, interés de búsqueda y cambio en ese interés. Este último revela qué es lo que más crece en popularidad. En el gráfico interactivo se muestra que los gráficos de barras, los diagramas de Gantt y los histogramas son los tipos de gráficos más buscados. También hay secciones que evalúan la popularidad de los datos sobre libros y herramientas de visualización (*The visual display of quantitative information* de Edward Tufte fue

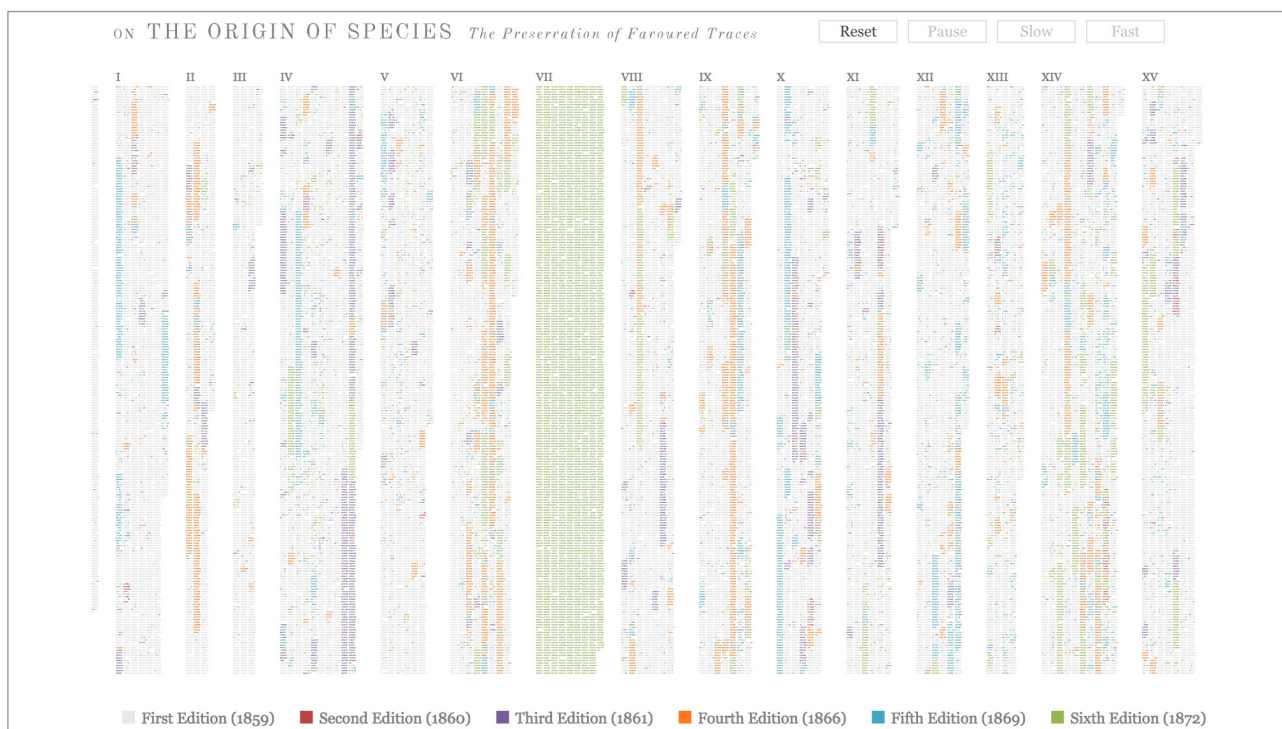


Figura 8. Visualización de los cambios introducidos en cada una de las ediciones la obra *El origen de las especies* de Charles Darwin

*Fathom*. <https://fathom.info/traces>

el libro más buscado, y *Excel* y *PowerPoint* fueron las principales herramientas utilizadas).

Queremos terminar este apartado reportando de forma breve un par de visualizaciones que pueden ser de interés por los datos que representan.

La primera entronca con propuestas ya comentadas en nuestra participación en el anterior número de *Anuario ThinkEPI* (Pérez-Montoro, 2017). Concretamente, con las visualizaciones que analizaban las diferentes ediciones de la obra *El origen de las especies* de Charles Darwin (1859). En este caso, tenemos representada toda la obra por la agregación de cada una de sus páginas. Cada guión representa una página (cuyo texto se despliega al pasar el cursor) y los colores nos muestran los cambios introducidos en cada una de las ediciones de la obra.

La segunda se centra directamente en una de las operaciones básicas y críticas involucradas en el proceso de la gestión de la información: la descripción de los recursos documentales.

En esta propuesta se intenta representar la brecha entre la fecha de la creación de la descripción bibliográfica (formato MARC) de una obra y la fecha de la publicación de esa misma obra (figura 9). Con este análisis visual podemos constatar, entre otras muchas cosas, que sólo una pequeña porción de las obras ha sido descrita de forma contemporánea a su publicación.

#### 4. Tecnología

El último de los apartados está dedicado a destacar algunas de las novedades introducidas en el ámbito de las tecnologías relacionadas con la disciplina de la visualización de datos.

La primera de las novedades es el lanzamiento oficial de la versión 2 de *Vega-Lite*, un lenguaje para crear visualizaciones de información interactivas de una forma sencilla. *Vega-Lite* permite la generación de visualizaciones como un conjunto de codificaciones que mapean los campos de datos con las propiedades de las marcas gráficas. Utiliza un formato JSON que compila según especificaciones completas en el lenguaje *Vega* superior. Incluye soporte para transformaciones de datos como la agregación, el agrupamiento, el filtrado y la clasificación, así como transformaciones visuales como el apilamiento y el facetado en pequeños múltiplos. También ofrece soporte para combinaciones flexibles de gráficos e interacciones, como la visión panorámica, el zoom, el filtrado interactivo o la selección vinculada.

La segunda de las novedades que queremos destacar es un proyecto de inteligencia artificial que ha obtenido uno de los *Google DNI* (*Digital News Initiative*) *Fund*, una iniciativa creada por la compañía del buscador para ofrecer apoyo al periodismo de alta calidad en Europa a través de la innovación en el espacio digital. Concretamente,

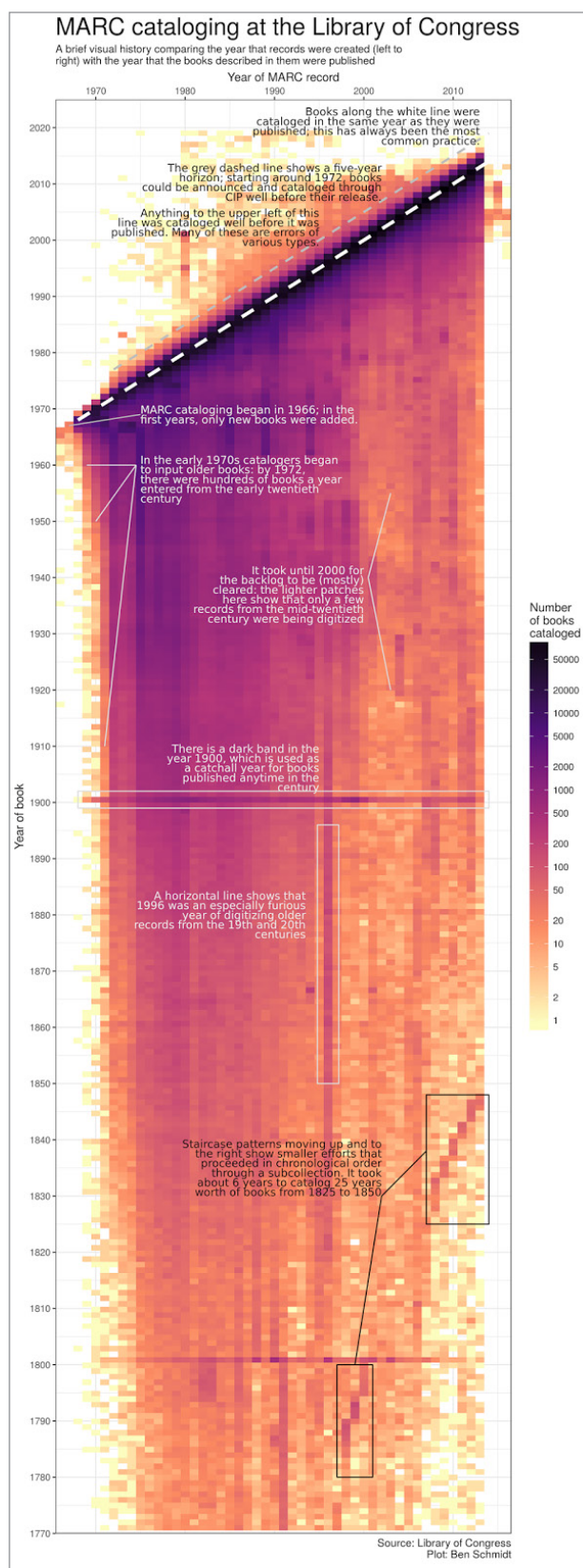


Figura 9. Análisis visual del grado de correspondencia entre la fecha de la creación de la descripción bibliográfica (formato MARC) de una obra y la fecha de la publicación de esa misma obra

A brief visual history of MARC cataloging at the *Library of Congress*

<http://sappingattention.blogspot.com.es/2017/05/a-brief-visual-history-of-marc.html?m=1>

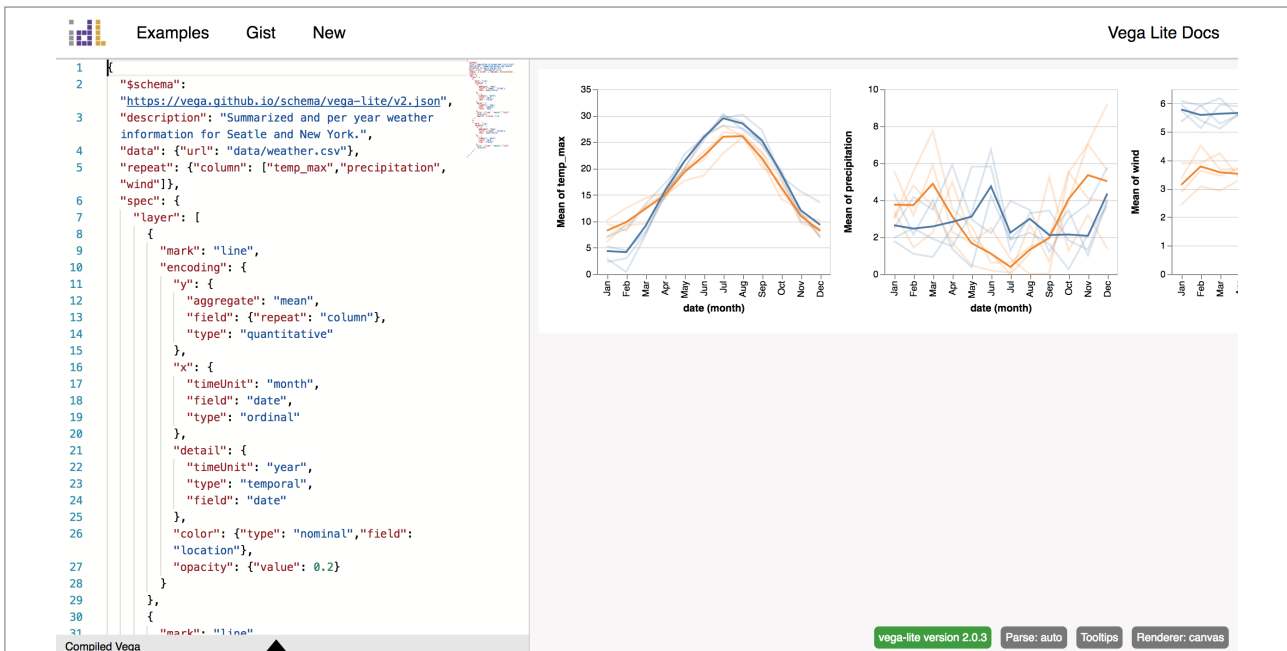


Figura 10. Ejemplo de visualización interactiva sobre el tiempo en Seattle y New York a lo largo de un año realizada con versión 2 de Vega-Lite Versión 2 de Vega-Lite. [https://vega.github.io/editor/#/examples/vega-lite/repeat\\_layer](https://vega.github.io/editor/#/examples/vega-lite/repeat_layer)

dentro del grupo *Prodigioso Volcán* se va a crear una plataforma de generación automática de gráficos mediante el uso de *Machine Learning Comprehension*. <http://www.prodigiosovolcan.com>

Esta iniciativa permite la generación sencilla e instantánea de la mayoría de los gráficos que publican los medios de comunicación de

forma recurrente (tasas de desempleo, impuestos, cotizaciones bursátiles, variaciones del PIB, audiencias de televisión, resúmenes financieros o resultados de elecciones políticas, entre otras muchas más).

La tercera de las novedades destacables se encuentra relacionada con los datos de investigación. Concretamente, queremos poner

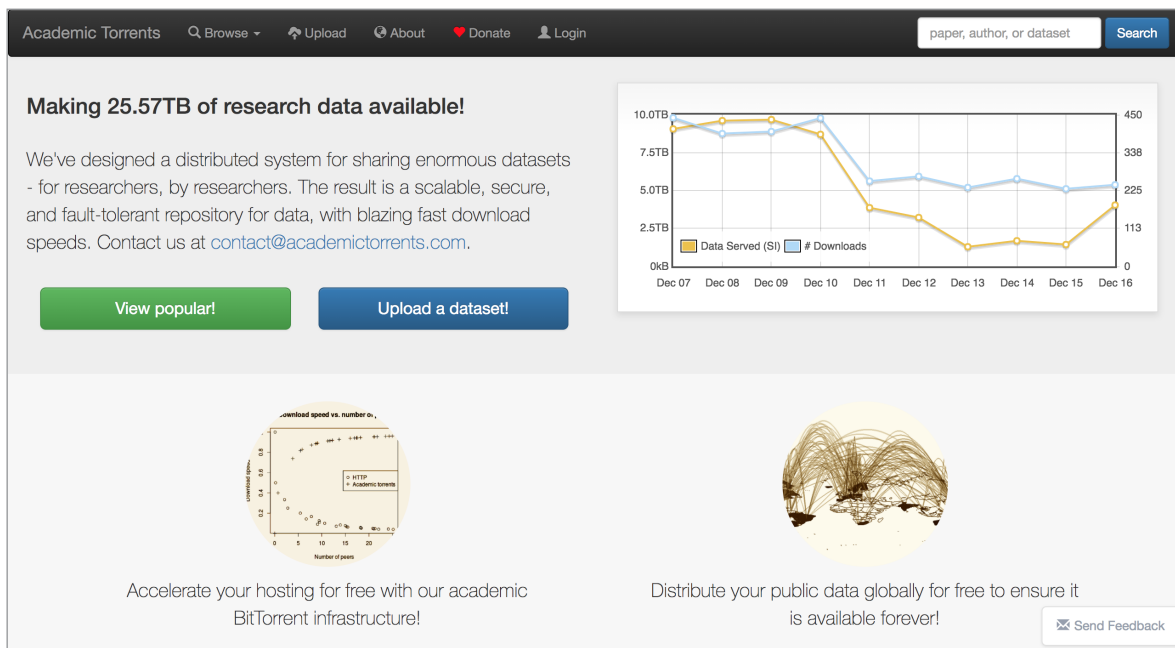


Figura 11. Página principal de *Academic Torrents*, un repositorio de datos de investigación *Academic torrents*. <http://academictorrents.com>



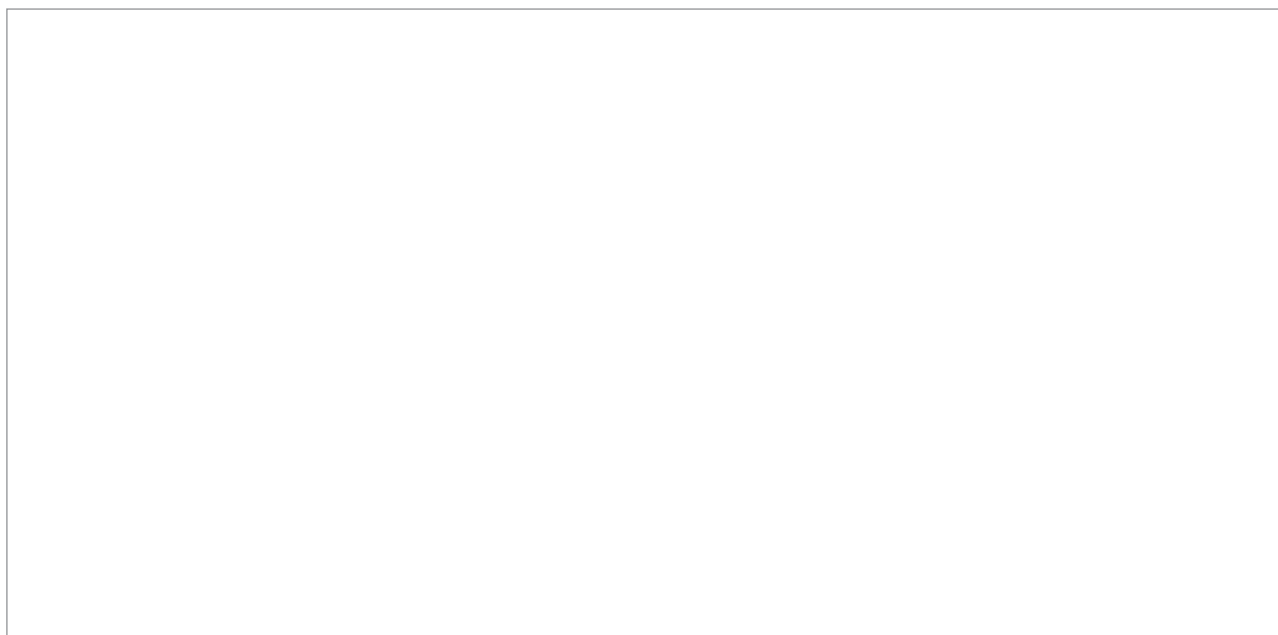


Figura 12. Monografía *Data visualization for social science*, de Kieran Healy

*Data visualization for social science*

[http://socviz.col/?utm\\_campaign=Data%2BElixir&utm\\_medium=email&utm\\_source=Data\\_Elixir\\_148](http://socviz.col/?utm_campaign=Data%2BElixir&utm_medium=email&utm_source=Data_Elixir_148)

en valor la creación de una plataforma de *torrents* de datos de investigación (de 25,57 TB de tamaño). En esta plataforma es posible subir para compartir y obtener de forma distribuida conjuntos de datos de investigación dentro de un repositorio escalable, seguro y con velocidades de descarga rápidas. El repositorio incorpora un sistema de navegación por categorías y un sistema de búsqueda para la interrogación del sistema.

Por último, queremos concluir este breve apartado dedicado a la tecnología destacando un libro que puede resultar interesante para aquellos que quieran introducirse en la disciplina de la visualización y en una de las herramientas principales, R, que suelen utilizarse para la generación de representaciones visuales. Se trata de una buena monografía (con una versión en abierto) sobre R y *ggplot2* aplicados al ámbito de las ciencias sociales.

Su autor, Kieran Healy (*Duke University*), nos ofrece un manual ilustrado con multitud de ejemplos que nos permitirá comprender los principios básicos de la visualización de datos, adquirir los

conocimientos prácticos para poder discriminar por qué algunos gráficos y figuras funcionan bien (mientras que otros pueden no informar o engañar de forma intencionada), generar una amplia gama de gráficos en R usando *ggplot2* y poder refinar las propuestas para conseguir una comunicación eficiente.

## 5. Referencias

**Pérez-Montoro, Mario** (2017). “Visualización de información en 2016: conceptos, contenidos y tecnología”. *Anuario ThinkEPI*, v. 11, pp. 284-296. <https://doi.org/10.3145/thinkepi.2017.55>

**Mario Pérez-Montoro**

*Universitat de Barcelona*

*Departament de Biblioteconomia, Documentació y Comunicació Audiovisual*  
[perez-montoro@ub.edu](mailto:perez-montoro@ub.edu)