

La visualización de datos como recurso social: posibilidades educativas y de activismo.
Ana Sedeño Valdellós (España).¹

Resumen.

La visualización de datos es una materia interdisciplinar de investigación que viene constituyendo nuevas prácticas de creación en el arte, los medios de comunicación y otras facetas. Posee, además, múltiples aplicaciones, desde un punto de vista de aprovechamiento social. En primer lugar, se aborda el ámbito educativo con el análisis de las necesidades crecientes en competencias a través de estas nuevas técnicas; se define el término minería de datos educativos y se identifican ejemplos de empleo de los datos y su visualización en entornos educativos, un campo de estudio que se encuentra en expansión. En segundo lugar, se enmarca esta técnica como protagonista dentro de un movimiento global por la liberación de los datos de procedencia pública con un objetivo social transformador. Se analizan algunos ejemplos en esta dirección.

Palabras clave.

Visualización de datos, educación mediática, minería de datos, activismo social, economía del dato, periodismo de datos.

Abstract.

Data visualization is an interdisciplinary issue that is building new practices in art, mass media and others. Moreover it has multiple applications from a social point of view. Firstly, the educational sector is addressed in the analysis of the increasing skill needs through these new techniques; the term data mining education is defined; it's given some examples of use of the data and educational settings are identified, a field study that is expanding. On the second hand, this technique is framed as protagonist in a global movement for the release of public data with a changing social aim. Some examples are discussed in the text.

Keyframes.

Data visualization, media literacy, data mining, social activism, economy of data, data journalism.

Visualización de datos: introducción y posibilidades generales.

Para comprender el alcance de ámbitos nuevos de investigación y creación como el de la visualización de datos hay que recordar la importante cuestión de la diferencia entre la información y el conocimiento. Cuando se habla de información, nos referimos a los datos y hechos cuantificables que rodean a un fenómeno. El conocimiento, sin embargo, consiste en señalar las características propias de un fenómeno y crear formas de estructurar sus facetas y evaluar cuales son las consecuencias de este. Así se trata de interpretar o conceptualizar con los datos o la información. Los teoremas y las leyes físicas son formas de interpretar los datos, de dar estructura a la información.

La visualización de datos (*data visualization*) supone un modo de formalizar herramientas para convertir la información en conocimiento que combina técnicas provenientes de la estadística o la programación con el diseño artístico. Se trata de una materia interdisciplinar de investigación que viene constituyendo nuevas prácticas de creación en el arte, los medios de comunicación y otras facetas y ello con múltiples implicaciones sociales. Consiste en la aplicación de técnicas para seleccionar, procesar y poner a disposición de una audiencia una cantidad de datos, dándoles significado para que, mediante su visualización, se conozcan sus relaciones de causa o dependencia, con el objetivo de señalar, denunciar o establecer conocimiento sobre un procedimiento, un fenómeno, una acción... El grupo de investigación interdisciplinar *Visualizar: Visualización de datos*,² residente en Medialab-Prado habla de este ámbito como:

"una disciplina transversal que utiliza el inmenso poder de comunicación de las imágenes para explicar de manera comprensible las relaciones de significado, causa y dependencia que se pueden encontrar entre las grandes masas abstractas de información que generan los procesos científicos y sociales." (De Vicente, 2012).

Como apunta el teórico norteamericano Lev Manovich (2010), la visualización de datos tiene como antecedentes la pintura rupestre y los mapas y representa otra forma más de la

necesidad humana de encontrar patrones y de abstraer datos de nuestro entorno vital, con el objetivo de la supervivencia individual y de la especie:

Utilizaré el término visualización para las situaciones en las que unos datos cuantificados, que no son visuales en sí mismos, el resultado de sensores meteorológicos, la conducta del mercado de valores, el conjunto de direcciones que describen la trayectoria de un mensaje a través de una red de ordenadores, etc. se transforman en representaciones visuales (Manovich, 2008).

La visualización de datos es una operación más de las que permite el hecho de la digitalización, donde la información se reduce a datos binarios. Técnicas de trabajo informático como el mapeado o la animación en 3D son formas de visualización plenamente integradas en el universo informático y digital. Estas posibilidades se están convirtiendo en todo un campo de investigación social y de creación artística.

Un fenómeno cercano al uso de estas posibilidades es el periodismo de datos, nuevo sector profesional en los medios de comunicación que se encuentra en plena formalización, como indica la pirámide invertida de Paul Bradshaw (2011). Para Flores (2009, p. 78), este proceso de infomorfosis, que están teniendo los mensajes informativos se produce independientemente del soporte (web, podcast, blogs, telefonía, etc.) por donde se emiten dichos mensajes. De ahí que las tecnologías no sólo afectan a las formas del periodismo, sino además a los contenidos”. Para Flores y Porto (2012), el Periodismo transmedia se define como una “una forma de lenguaje periodístico que contempla, al mismo tiempo, distintos medios, con varios lenguajes y narrativas a partir de numerosos medios y para una infinidad de usuarios”.

Periódicos como *The New York Times* o el británico *The Guardian* han convertido esta práctica en elemento diferenciador e incorporan en su información todo tipo de fórmulas para el tratamiento y la visualización de datos, facilitando su velocidad de lectura y comprensión, conscientes de la infoxicación (la cantidad abrumadora de información) que sufre el lector. Esta rama del periodismo emplea aplicaciones como Google Refine, Maps,

Google Fusion Tables, Google Earth, Hojas de cálculo, Docs, Dipity, DataWrapper, Mapas cartográficos (para geoposicionamiento), unas libres y otras propietarias, pero el catálogo se incrementa exponencialmente. Este proceso de transformación del periodismo se encuentra paralelo a la necesidad de mayor participación del individuo en las redes sociales y/o ciudadanas. Con tales herramientas los individuos participan en la difusión de los hechos noticiosos, tanto de una manera individual (a través de cuentas propias), como aportando otros datos al trabajo de un periodista profesional.

Por otro lado, la gestión y utilización de datos de todo tipo casi está convirtiéndose en un sector de lo que se llama “Nueva Economía”, “una nueva forma de producción y consumo, consecuencia de los cambios tecnológicos relacionados con la información, las comunicaciones y la globalización” que hacen referencia a “la forma en que las nuevas tecnologías de la información mejoran la eficiencia de todos los aspectos de la economía, especialmente de las empresas tradicionales.” (De Vicente, 2012, p. 10).

El ámbito del Big data parece centrar la atención de todo un nuevo sector de la economía, pues supone la aparición de “conjuntos de datos cuyo tamaño les sitúa más allá de las posibilidades de captura, almacenamiento, gestión y análisis de las herramientas de bases de datos más comunes” (McKinsey, 2013). Además, se encuentra en el núcleo de la llamada Internet de las cosas, que supone que cualquier objeto de la vida cotidiana (máquinas, electrodomésticos...) necesitan y generan información sobre su funcionamiento y su interacción con los humanos y por tanto, esta puede ser empleada para mejorarlos.

Visualización de datos en educación.

A continuación se abordan las potencialidades que esta nueva herramienta tecnológica ofrece a la educación, especialmente para determinados tipos, como el *e-learning*. Para comenzar con este tema, resulta esclarecedor contar con algunas conclusiones de

investigación que prueban la idoneidad de la visualización de información a la hora de su aprovechamiento intelectual. Ruddle y su equipo (2002) afirman que “los gráficos, si están bien contruidos, pueden hacer el proceso cognitivo humano de la construcción de una imagen mental más fáciles”.³ Shaw realizó un experimento para averiguar si las nuevas formas de presentación de la información y los datos podían mejorar la percepción y retención de los mismos y llegó a estas conclusiones (Anagastaki, 2013):

- ✓ Las personas que trabajan con técnicas de mapeo visual usaron un promedio de 19% menos de recursos cognitivos.
- ✓ Fueron un 17% más productivo y 4.5% más capaces de recordar los detalles que cuando se utiliza el software tradicional equivalente.
- ✓ Los grupos que trabajan juntos en un proyecto utilizan un promedio de 10% menos de recursos cognitivos.
- ✓ Fueron un 8% más productivos y recordaron 6,5% más de datos cuando emplearon visualizaciones, en comparación con técnicas tradicionales.

De hecho, ya se vislumbra la necesidad en la formación de profesionales que gestionen y dominen esta información numérica, el conocido como científico de datos o data scientist. Como ha indicado Pat Hanrahan (Woods, 2011), prácticamente todos los trabajadores necesitarán poseer competencias en relación a los datos. De igual manera, en el Informe *Future Work Skills 2020* (Davis, Fidler y Gorbis, 2011), del Phoenix Research Institute, proporciona diez grandes bloques de competencias casi todas relacionadas con los datos: dar sentido, inteligencia social, pensamiento adaptativo, competencias interculturales, pensamiento computacional, alfabetización en nuevos medios, transdisciplinareidad, mentalidad de diseño, gestión de la carga cognitiva, colaboración virtual. Surge aquí un nuevo ámbito de formación sobre el que reflexionar.

El impacto educativo de la visualización no sólo depende de qué tan bien los estudiantes aprenden cuando lo usan, sino de cómo sea aceptado e integrado en las clases por los

profesores. Según Naps et al. (2003), la integración de las técnicas de visualización en la enseñanza en clase, sin embargo, ha quedado muy por debajo de su potencial, debido a una falta de comprensión de las necesidades del instructor. El trabajo describe estas necesidades y ofrece orientaciones tanto para el uso efectivo de visualizaciones como para su evaluación e impacto en el estilo de aprendizaje.

La visualización de datos se encuentra en los temas más requeridos de MOOCs de todo el mundo, desde que el Centro Knight de Datos propuso el curso "Introducción a la Infografía y Visualización de Datos", con el instructor Alberto Cairo. Los MOOCs sobre visualización se dirigen hacia la adaptación del conocimiento sobre uso de la información a cualquier profesión o interés mediante técnicas narrativas (formas de presentar información que permitan contar historias) adecuadas a cada ámbito. Otras instituciones como el Data Driven Journalism han puesto en marcha cursos como el *Doing Journalism with Data: First Step, Skills and Tools*, con gran éxito internacional.⁴

La minería de datos o *data mining* se está aplicando ya como un campo interdisciplinar en educación bajo el nombre de minería de datos educativos (EDM o *educational data mining*). Se trata de desarrollar métodos de exploración de datos que llegan desde ambientes educativos. Su propósito es la comprensión de los matices de aprendizaje de los diferentes alumnos y la identificación de sus preferencias para los resultados educativos. También puede emplearse a un nivel macrogenérico, desde la administración educativa, para evaluar la situación de este tipo de entornos y sacar conclusiones sobre los resultados de las políticas en su contexto real.

La EDM puede ser descrita como la combinación de las ciencias de la computación, la estadística y la educación. Existen variados métodos, unos son interdisciplinarios con la minería de datos general (como la predicción, el *clustering*, la detección de valores atípicos, la minería de relaciones, el análisis de redes sociales, los análisis textuales) y otros que son particulares de la minería de datos educativa como la destilación de datos para juicio

humano, el descubrimiento con modelos, la factorización de matriz no negativa y el rastreo de conocimiento (Romero and Ventura, 2013, pp. 21-22).

Existen dos libros principales sobre EDM. *Data Mining in E-Learning* donde se analiza el fenómeno en este tipo de entorno educativo que cuenta con técnicas como Gismo o CourseVis, que permiten construir imágenes digitales animadas y/o interactivas, y el *Handbook of Educational Data Mining* (Romero et al., 2010). En cuanto a los *journals*, existen en torno a once publicaciones interesadas directamente en estos temas, algunas con alto impacto e indexadas en la Journal Citation Report (JCR): *Journal of the Learning Sciences* y *Computer and Education*. Los temas de mayor interés de la EDM en estas publicaciones son la organización de las clases para buscar la eficiencia educativa, la identificación de quienes se benefician del feedback del profesor y de sus consejos y ayuda, la decisión de qué tipo de feedback o ayuda es más efectiva y los detalles sobre cómo animar a estudiar a los alumnos y a buscar material útil de manera individual o colectiva (Romero y Ventura, 2013, p. 18). Además, los autores han descrito los temas que se encuentran en la agenda de la investigación en este campo (Romero y Ventura, 2013): métodos y entornos genéricos, minería de datos educaciones, minería de los procesos educativos, adaptación y personalización de datos, desarrollo de software educacional; evaluación de intervenciones de aprendizaje, emoción, afecto y elección, teoría pedagógica e integración de datos... entre otros.

Otro ámbito en creciente interés es el de las herramientas de visualización de datos. Tag Galaxy, VUE (Visual Understanding Environment), Many Eyes o Revisit son empleadas en universidades como Stanford, Tufts o British Columbia. Algunas están basadas en Flickr (Tag Galaxy) o permiten visualizar mensajes de twitter y su contenido asociado (Revisit) de manera directa y en tiempo real. La Visual Understanding Environment (VUE)⁵ es un proyecto Open Source de base en la Tufts University. Se centra en crear técnicas flexibles para integrar recursos digitales y gestionarlos para apoyar la enseñanza y la investigación.

Visualización de datos y activismo ciudadano.

La visualización de datos se ha convertido en una técnica básica de generación de herramientas efectivas de comunicación por medio de estrategias de diseño, infografía y demás modos de organización visual de datos. Entre otras aplicaciones como las mencionadas, la visualización ha respondido también a cierta necesidad de construcción colectiva de la información y de su puesta en disposición a la ciudadanía nacional o internacional. Su evolución y su intencionalidad en torno a la regeneración democrática de las sociedades contemporáneas en los últimos años le ha dado una dimensión de fenómeno o movimiento (Álvarez, Gertrudix, y Rajas, 2014; Borger, Hoof, Costera y Sanders, 2013). De hecho se encuentra en el ámbito de la denominada tecnopolítica, disciplina que trata de pensar desde una perspectiva transdisciplinaria cómo la tecnología se aplica o puede aplicarse a objetivos comunitarios o políticos, incorporando conocimientos de las ciencias de la complejidad, las ciencias sociales, las redes y los datos.

Estas ideas se encuentran amparadas bajo ciertos conceptos teóricos producidos por las ciencias sociales para describir nuevos procesos sociales. Autores como Howard Rheingold (2004) han analizado la inteligencia colectiva o mente enjambre, que emerge en procesos de movilización multitudinaria mediados por las TICs. Por su parte, Borge-Holthoefer et al. (2011) han estudiado movimientos como Occupy Wall Street, el 15M o la primavera árabe como fenómenos donde las redes sociales se han servido de las herramientas de visualización para explicar los comportamientos colectivos.

El movimiento por el Government Open Data pretende que los datos e información puedan ser usados y distribuidos gratuitamente. Además, se trabaja para que sea el gobierno y la administración las que produzcan y controlen la producción de datos y la pongan al servicio de la ciudadanía. Según esta organización, los datos que ofrece un gobierno son abiertos si cumplen con estos ocho principios (Opengovdata, 2007)⁶ completos, primarios, oportunos, accesibles, procesables automáticamente, no discriminatorios y no propietarios y libres de

licencias. A estos se añaden siete principios más, que solicitan que los datos deben ser online y gratuitos, permanentes, compartidos, con una presunción de apertura, documentados, seguros como apertura y diseñados con un input público. Todas estas condiciones parecen describir las condiciones en las que los datos podrían ser plenamente manipulables para objetivos activistas y con una intención social transformadora y, por tanto, dispuestos fácilmente por cualquier herramienta, incluso básica, de visualización.

Otras organizaciones trabajan bajo la idea de que la liberación de los datos de todo tipo pero sobre todo de las instituciones públicas es un derecho fundamental de las personas en la sociedad de la información. Así la World Bank Institute⁷ opina que “El gobierno abierto asegura el acceso de los ciudadanos a la información, datos, procesos gubernamentales, involucra a los gobiernos de forma más eficaz en la voluntad y capacidad para responder a los ciudadanos y trabajar en colaboración para resolver los problemas difíciles de gobernanza”. Por su parte, a Open Knowledge Foundation⁸ ha producido a modo de eslogan el mejor mensaje sobre el empoderamiento a través de los datos: “observa cómo los datos pueden cambiar el mundo, un mundo donde el conocimiento crea el poder de la mayoría, no de unos pocos” (Álvarez, Gertrudix y Rajas, 2014).

El proyecto Atlas de la Innovación en España⁹ ha construido una herramienta online que permite conocer cómo se distribuyen los agentes empresariales innovadores, especialmente Pymes, en el país. Se pretende informar con una representación visual interactiva sobre las empresas de base tecnológica y actividad I+D+I en España desde una perspectiva geográfica y sectorial.

El proyecto atNight¹⁰ explora las posibilidades que ofrece el Big Data para diseñar y tratar colaborativamente las ciudades y aportar información sobre cómo utilizar más eficientemente la energía.

Por su parte, Reveal-it¹¹ propone revelar información específica de los procesos de consumo de los ciudadanos y su interrelación con estructuras energéticas en ambientes urbanos. El objetivo es recopilar información sobre los procesos energéticos y cómo mejorarlos, así como relacionarlos con patrones de comportamiento humano respecto a la

energía y los intereses que existen en torno a su producción, distribución y comercialización en diferentes países.

Hacks y hackers¹² creado por Burt Herman, Richard Gordon y Aron Pilhofer propone un sitio para investigar sobre cómo utilizar la información para participar de manera activa en la vida social a través de los medios de comunicación y las asociaciones políticas.

MercaMadrid¹³ es una topología visual de la red de flujos de alimentos que hacen visibles los cambios de tendencias alimenticias en los últimos años en Madrid (y por extensión en España, como capital del país).

En el caso de Many eyes¹⁴ presenta un sitio público para obtener visualizaciones interactivas de datos desde la web hasta en doce formas. Contiene también foros donde se debate sobre los datos a modo de comunidad *online*.

Conclusiones.

El texto ha tratado de describir la visualización de datos como una técnica que permite una disposición de los datos para presentarlos de manera organizada ante diferentes públicos y con distintos objetivos y que por esta razón adquiriera el estatuto de conocimiento, comprensible y válido para su interpretación y utilización con objetivos de transformación social. Como a cualquier herramienta de tratamiento de la información para su transformación en conocimiento, pueden aplicársele diversos usos: entre ellos se encuentran también los relacionados con las posibilidades del empoderamiento ciudadano y en el entorno educativo.

En este último caso, está siendo empleado por instituciones educativas para conocer sus posibilidades en la educación *e-learning* así como en una creciente reflexión acerca de las

necesidades en habilidades en torno a los datos, que se prevén crecientes en los próximos años. La minería de datos educativa se muestra como una técnica que permite conocer diferentes aspectos de las interacciones pedagógicas y que, en la actualidad, se encuentra en expansión.

En la actualidad, se observa una firme creencia en el poder del conocimiento colectivo como motor de cambio. La tecnopolítica da nombre al conjunto de posibilidades de aplicación de las herramientas de visualización respecto a objetivos políticos de alcance global, que definen una ciudadanía global consciente de las problemáticas generales que la globalización de los comportamientos y los procesos sociales ejercen en todos los seres humanos, independientemente de su localización geográfica. En este caso, la visualización ha respondido a cierta necesidad de construcción colectiva de la información, con objetivos de regeneración democrática y de disposición libre de la información pública, para la mejora del control de la ciudadanía a los poderes políticos nacionales o globales.

Bibliografía.

Álvarez, Sergio, Gertrudix, Manuel (2011). Contenidos digitales abiertos y participación en la Sociedad Digital. *Enl@ce. Revista Venezolana de Información, Tecnología y Conocimiento*, 8 (2), 79-93.

Álvarez García, Sergio, Gértrudix Barrio, Manuel y Rajas Fernández, Mario (2014). La construcción colaborativa de bancos de datos abiertos como instrumento de empoderamiento ciudadano. *Revista Latina de Comunicación Social*, 69, 661-683. Recuperado el 13 de abril de 2015. Disponible en http://www.revistalatinacs.org/069/paper/1029_UR/32es.html

Anagostaki, Eva (2013). *Visual for thought: why visualization should be used more in the classroom*. Master of Media. University of Amsterdam Recuperado el 24 de marzo de 2015. Disponible en <http://mastersofmedia.hum.uva.nl/2013/03/20/visual-for-thought-why-data-visualization-should-be-used-more-in-the-classroom/>

Anton Bravo, Antón (2013). El periodismo de datos y la web semántica, *CIC*, vol. 18, 99-116.

Borger, Merel, Hoof, Anita, Costera, Irene, Sanders, José (2013). Constructing participatory journalism as a scholarly object. *Digital Journalism*, 1(1), 17-134.

Borge-Holthoefer, Javier et alt. (2011). Structural and dynamical patterns on online social Networks: The spanish may 15th movement as a case study. *Plos One*, 6(8). Recuperado el 12 de diciembre de 2014. Disponible en www.plosone.org/.../info%3Adoi%2F10.1371%2Fjournal.pone.0023883

Bradshaw, Paul (2011). La pirámide invertida del periodismo de datos. *Blog Tejiendo redes*. Recuperada el 23 de noviembre de 2014 Disponible en <http://tejiendo-redes.com/2011/07/07/la-piramide-invertida-del-periodismo-de-datos/>

Cairo, Alberto (2011). *El arte funcional: Infografía y visualización de información*. Madrid: Alamut.

Davies, Anna, Fidler, Devin y Gorbis, Marina (2011). Future Work Skills 2010. Palo Alto, California: Institute for the Future for the University of Phoenix Research Institute. Recuperado el 28 de enero. Disponible en http://www.iftf.org/uploads/media/SR-1382A_UPRI_future_work_skills_sm.pdf

De Vicente, Jose Luis et alt. (2012). *Sectores de la nueva economía 20+20. La economía del dato*. Madrid: EOI.

Flores, Jesus Miguel (2009). Nuevos modelos de comunicación, perfiles y tendencias en las redes sociales. *Comunicar*, 33, 73-81.

Manovich, Lev (2010). What is visualization?. *Poetess Archive Journal*, 2.1.

Manovich, Lev (2008). La visualización de datos como nueva abstracción y antisublime. *Estudios visuales*, 5 (24/7), 126-136.

Manovich, Lev (2013). Visualizing Vertov. Disponible en <http://lab.softwarestudies.com/2013/01/visualizing-vertov-new-article-by-lev.html>

McKinsey Global Institute (2013): *Big Data: The Next Frontier for Innovation, Competition and Productivity*. Recuperado el 5 de abril de 2015. Disponible en http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx

Naps, Thomas L., Cooper, Stephen, Koldehofe, Boris y Les, C. (2003). Evaluating the educational impact of visualization. iTiCSE-WGR 03 Working group reports from ITiCSE on Innovation and technology in computer science education. Nueva York: ACM.

Porto, Denis y Flores, Javier (2012). *Periodismo transmedia*. Madrid: Fragua.

Rheingold, Howard (2004). *Multitudes inteligentes: la próxima revolución social*. Barcelona: Gedisa.

Romero, Cristobal & Ventura, Sebastian (2013). "Data mining in education". *WIREs Data Mining and Knowledge Discovery*, 3(1), pp. 12-27. Recuperado el 3 de marzo de 2015.

Ruddle, R., Brodlie, K., and Dimitrova, V. (2002). *Communication, visualisation and interaction*. Leeds: University of Leeds, School of Computing. Handbook of the course.

Woods, David (2011, 30 de noviembre). Tableau Software's Pat Hanrahan on What is a Data Scientist. *Forbes*, Disponible en www.forbes.com/sites/danwoods/2011/11/30/tableau-software-pat-hanrahan-on-what-is-a-data-scientist/

¹ Doctora en Comunicación Audiovisual y Profesora en el Departamento de Comunicación Audiovisual y Publicidad de la Universidad de Málaga (España).

² <http://medialab-prado.es/article/visualizar>

³ "Graphics, if well constructed, may make the human cognitive process of constructing a mental image of the route easier". (2002, Traducción propia).

⁴ <http://www.datadrivenjournalism.net/course/>

⁵ vue.tufts.edu

⁶ <http://opengovdata.org>

⁷ <http://wbi.worldbank.org/wbi/>

⁸ <https://okfn.org/>

⁹ <http://www.atlasvisualdeinnovacion.com>

¹⁰ <http://www.atnight.ws/contact.php#.UiYm2IXTNrU>

¹¹ <http://bit.ly/reveal-it>

¹² <http://hackshackers.com>

¹³ <https://mercamadridvisualizar.wordpress.com/proyecto-2/colaboran/>

¹⁴ www.manyeyes.com

R
y
P