



Universitat d'Alacant
Universidad de Alicante

XIII JORNADES DE XARXES D'INVESTIGACIÓ EN DOCÈNCIA UNIVERSITÀRIA

Noves estratègies organitzatives i metodològiques en la formació
universitària per a respondre a la necessitat d'adaptació i canvi



JORNADAS
DE REDES DE INVESTIGACIÓN
EN DOCENCIA UNIVERSITARIA **XIII**

Nuevas estrategias organizativas y metodológicas en la formación
universitaria para responder a la necesidad de adaptación y cambio

ISBN: 978-84-606-8636-1

Coordinadores

María Teresa Tortosa Ybáñez

José Daniel Álvarez Teruel

Neus Pellín Buades

© **Del texto: los autores**

© **De esta edición:**

Universidad de Alicante

Vicerrectorado de Estudios, Formación y Calidad

Instituto de Ciencias de la Educación (ICE)

ISBN: 978-84-606-8636-1

Revisión y maquetación: Neus Pellín Buades

Publicación: Julio 2015

Una aplicación Web para la exploración interactiva e intuitiva de las pirámides de población de España

B. Zaragozaí Zaragozaí; J. Torres Prieto; A. Belda Antolí

Instituto Universitario de Informática
Instituto Interuniversitario de Geografía
Depto. Ciencias de la Tierra y Medio Ambiente
Universidad de Alicante

RESUMEN

En esta comunicación se destaca el potencial pedagógico de las TICs en la enseñanza de los contenidos sobre la estructura y dinámica de la población. Sin embargo, el uso de estas tecnologías no se ha generalizado tanto como sería esperable. La complejidad de las herramientas y la ausencia de interfaces de usuario intuitivos podrían explicar esta escasa adopción. En este sentido, se ha desarrollado el prototipo de una aplicación web para que los estudiantes puedan explorar la estructura de la población de España de un modo interactivo y dinámico. Dicha aplicación permite consultar individualmente todas las pirámides de población y las gráficas de tendencia de los municipios de España, ajustar los parámetros de visualización y exportar sus consultas.

Palabras clave: aplicación web, pirámides de población, constructivismo, aprendizaje interactivo, datos locales

1. INTRODUCCIÓN

La visualización de datos es una disciplina transversal que utiliza el inmenso poder de comunicación de las imágenes para explicar de manera comprensible los patrones que se pueden encontrar entre los grandes volúmenes de datos que generan los procesos científicos y sociales de hoy en día. Recientemente, el auge del Big Data y la creciente conciencia de su gran potencial en el mundo de los negocios han propiciado un mayor desarrollo de la visualización como disciplina. Estas tendencias han hecho necesario que muchos profesionales sin conocimientos técnicos específicos necesiten aplicar procesos de Big Data en sus organizaciones. La visualización de datos y una serie de nuevas herramientas hace que esto sea posible. Por ejemplo, plataformas y proyectos como Platfora, Datameer, Clearstory, Chartio, Visual.ly, Vizify o Tableau, entre otros, han hecho que estas técnicas avanzadas de visualización para Big Data sean asequibles para todo tipo de organizaciones. Por otra parte, también existen proyectos de software libre que permiten el desarrollo de visualizaciones personalizadas, entre los que D3.js, Paper.js, Raphael.js o R, son algunos de los más importantes.

1.1 Aprendizaje basado en visualizaciones

Las técnicas de visualización de datos también han sido aplicadas como apoyo a metodologías de aprendizaje constructivistas (Day, 2012). La mayor parte de las visualizaciones se utilizan en un marco en el que los alumnos construyen el conocimiento de forma que tenga sentido para ellos. La teoría constructivista es la base para una amplia gama de enfoques de aprendizaje activo, como el aprendizaje basado en la investigación y el aprendizaje basado en problemas (Beringer, 2007). Estos enfoques inductivos se basan en que los estudiantes construyan teoría y generalizaciones a partir de estudios de caso y ejemplos, en lugar de enfoques más tradicionales en los que los estudiantes aprenden la teoría y luego estudian un pequeño conjunto de ejemplos. Los estudiantes suelen ser más comprometidos en su aprendizaje activo de lo que son en los enfoques tradicionales, pero el impacto de estos enfoques sobre el aprendizaje no está claro y se carece de estudios suficientes (Day, 2012).

El impacto educativo de los métodos de visualización depende de dos componentes: la mejora de la formación con la visualización y el despliegue de la visualización en el aula. En estudios anteriores se ha comprobado que hay una desconexión significativa entre la creencia

intuitiva de que la visualización mejora el aprendizaje de los estudiantes y la voluntad y capacidad de los docentes para desplegar la visualización en sus aulas (Naps et al., 2003).

Según Price et al. (1998), hay cuatro roles en el despliegue de técnicas de visualización en el proceso educativo: desarrolladores, diseñadores, docentes y estudiantes. Cada uno de estos roles desarrolla funciones específicas en las visualizaciones. Por ejemplo, los desarrolladores de herramientas de visualización están interesados en optimizar sus desarrollos para los otros tres roles. Los diseñadores de visualización se esfuerzan por lograr visualizaciones que sean valiosas para una gran audiencia. Los docentes quieren integrar las visualizaciones entre sus materiales, siempre que estas sean capaces de mejorar su enseñanza y la motivar el aprendizaje de los estudiantes. Finalmente, los estudiantes esperan aprender los conceptos mejor y de una manera más divertida.

Según Naps et al. (2003), la mayor parte de los estudios sobre herramientas de visualización se han centrado en los roles del desarrollador y diseñador, mientras que la investigación en Ciencias Sociales se ha centrado en la efectividad de la visualización para mejorar el aprendizaje de los estudiantes. En cambio, hay pocas investigaciones que se hayan centrado en las necesidades del docente. Por ejemplo, según Shaffer et al. (2011), las encuestas anuales realizadas en el SIGCSE'10 indicaron que, a pesar de la opinión positiva de la mayor parte de los encuestados (90%), solamente el 10% de confirmaron haber utilizado frecuentemente las visualizaciones entre sus materiales. En cambio, un porcentaje más amplio indicaron haber utilizado estas técnicas de una manera ocasional (50-75%). Según los encuestados, los principales impedimentos para la integración de las visualizaciones en el aula serían:

- Problemas para encontrar buenos ejemplos
- Aprender a utilizar nuevas herramientas
- Carencia de herramientas de desarrollo efectivas
- El tiempo necesario para integrar las visualizaciones dentro de los contenidos de un curso más amplio.
- Dudas respecto al valor pedagógico de las visualizaciones.
- Limitaciones tecnológicas en las aulas.

El mayor impedimento para la adopción de estas técnicas por parte de los docentes es el tiempo necesario para encontrar buenos ejemplos, aprender, instalar y desarrollar visualizaciones y luego integrarlas en un curso. El impacto global de las visualizaciones en el

aula será mínimo hasta que más docentes opten por integrar estas técnicas en sus clases. Finalmente, queda por aportar evidencias que respalden los beneficios de las distintas visualizaciones en el aprendizaje, pero los estudios basados en encuestas son costosos y difíciles de mantener en el tiempo.

Estrechamente relacionado con el impacto de estas tecnologías en el aula, se ha apuntado que el uso de datos locales en las lecciones facilita que los estudiantes participen más activamente en el aprendizaje (Liu et al., 2008). Los estudiantes muestran una curiosidad natural acerca de sus áreas locales (Kim et al. 2013). Por ejemplo, según Kim et al (2013), el uso de datos locales es fundamental para implementar con éxito los SIG en la educación secundaria. Por este motivo, no basta con generar buenas visualizaciones, sino que estas deben incorporar datos locales de un modo sencillo y fácil de explotar. No obstante, la falta de estándares y la multitud de medios para compartir los datos públicos dificultan esto y requieren el 80% del tiempo necesario (Extracción, Transformación y Carga) para realizar análisis de estas características (Maimon & Rokach, 2010).

1.2 Objetivos

En este estudio se quiere avanzar hacia la superación de los obstáculos mencionados y para ello se quiere desarrollar un primer prototipo de una aplicación Web, con abundantes datos locales disponibles y de sencilla integración en un curso de Ciencias Sociales. Evidentemente, siendo una Web App, no serían necesarios procesos de instalación, ni conocimientos tecnológicos específicos. Finalmente, este tipo de visualizaciones deberían registrar las consultas de los usuarios y ser capaces de proporcionar datos sobre su uso y entender mejor cómo aprenden los usuarios. La plataforma tecnológica descrita hará posible que en futuros estudios se pueda responder a las dos hipótesis que se han introducido en los párrafos anteriores: (1) ¿En qué medida puede mejorar el aprendizaje con el uso de visualizaciones de datos? y (2) ¿Cómo de importante es la utilización de datos locales para aumentar el interés de los alumnos?

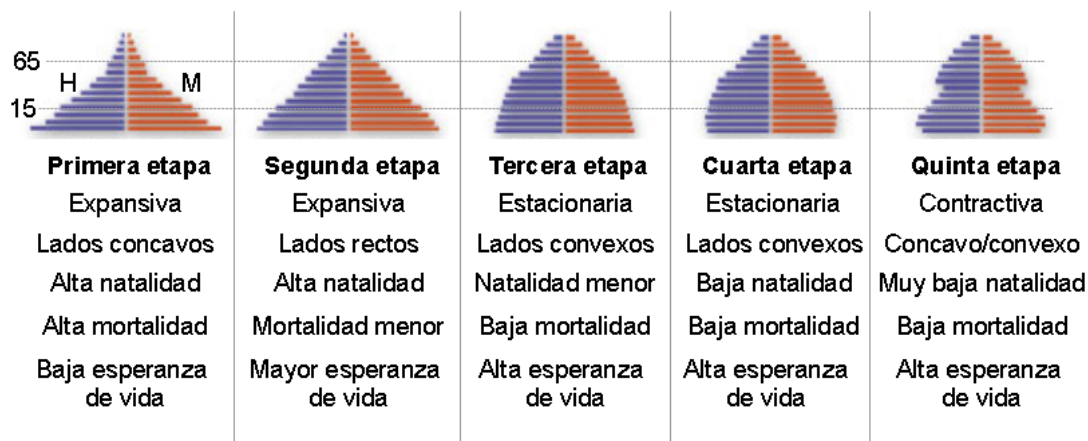
2. DESARROLLO DE LA CUESTIÓN PLANTEADA

2.1 Pirámides de población

Las pirámides de población son un punto de partida excelente para generar visualizaciones que cumplan todos los requisitos comentados en la introducción. En este caso

de estudio no es necesario diseñar una nueva visualización, por lo que su integración entre los materiales de un curso de Ciencias Sociales no será complicada. Las pirámides de población tienen un gran poder explicativo y se han utilizado durante varias décadas en la enseñanza de la estructura de la población.

Figura 1. Pirámides de población y sus formas según el modelo de Transición Demográfica



Una pirámide de población muestra la estructura por edad y sexo de la población de un territorio y puede proporcionar información acerca de la estabilidad política y social, así como el desarrollo económico. No obstante, también se pueden mostrar otras variables relacionadas en forma de “barras apiladas” (tamaño de la vivienda, nivel educativo, ingresos, por edad y sexo, entre otras posibilidades disponibles en los censos). La población se distribuye a lo largo del eje horizontal de un gráfico de barras, con la población masculina a la izquierda y la femenina a la derecha. Ambas se dividen en grupos de edad, habitualmente grupos de 5 años, representados como barras horizontales a lo largo del eje vertical, con los grupos de edad más jóvenes en la parte inferior y el más antiguo en la parte superior. La forma de la pirámide de población evoluciona gradualmente con el tiempo, dependiendo de la fecundidad, la mortalidad y las tendencias de la migración. Las pirámides de población tienen diferentes formas que indican hechos destacables sobre la población del país o región estudiada. Estos gráficos son útiles porque dan una idea muy visual de la esperanza de vida, las tasas de natalidad y mortalidad, entre otros aspectos. Además, las pirámides muestran cambios en el tiempo, por lo que pueden ayudar a los dirigentes a planificar políticas sociales de acuerdo a modelos como el de la Transición Demográfica (ver Figura 1).

Dada su amplia difusión, no faltan ejemplos de visualizaciones interactivas de pirámides de población. Entre otras visualizaciones, es posible encontrar desarrollos en Javascript, visualizaciones en la plataforma Tableau y también en Shiny. En la página web

de populationpyramid.net se pueden consultar un gran número de pirámides de población a partir de los datos del Banco Mundial. En esta web se pueden consultar las pirámides nacionales y regionales, entre 1950 y 2100, siendo en su mayor parte datos estimados por dicho organismo. En la plataforma Tableau Public es posible encontrar dos tipos de paneles interactivos, uno más tradicionalⁱ y una visualización más moderna que muestra una mayor cantidad de datos en un único vistazoⁱⁱ. Finalmente, también existen visualizaciones de pirámides de población desarrolladas con R y Shiny, para estudiar la población de Italiaⁱⁱⁱ.

Como se ha mencionado, existe una cierta variedad de visualizaciones de pirámides desarrolladas con distintas tecnologías. Sin embargo, el único elemento limitante es el acceso a los datos, ya que estos proceden de múltiples orígenes, formatos y proporcionan distintos niveles de detalle temporal (p. ej. datos por año o datos por grupos de edad), o espacial (p. ej. escala nacional, municipal o inframunicipal). Además, como consecuencia de las tecnologías utilizadas para el desarrollo de estas visualizaciones (“lado del servidor/lado del cliente”), no es posible conocer el uso que se hace de éstas (p. ej. gráficas más consultadas, selección por áreas geográficas, interés por las estructuras atípicas, entre otras opciones). Así pues, para poder crear una única plataforma en la que consultar pirámides de distintos orígenes será necesario implementar un almacén de datos (datawarehouse) con herramientas específicas que faciliten la importación, gestión, explotación y representación de datos de población de distintas fuentes.

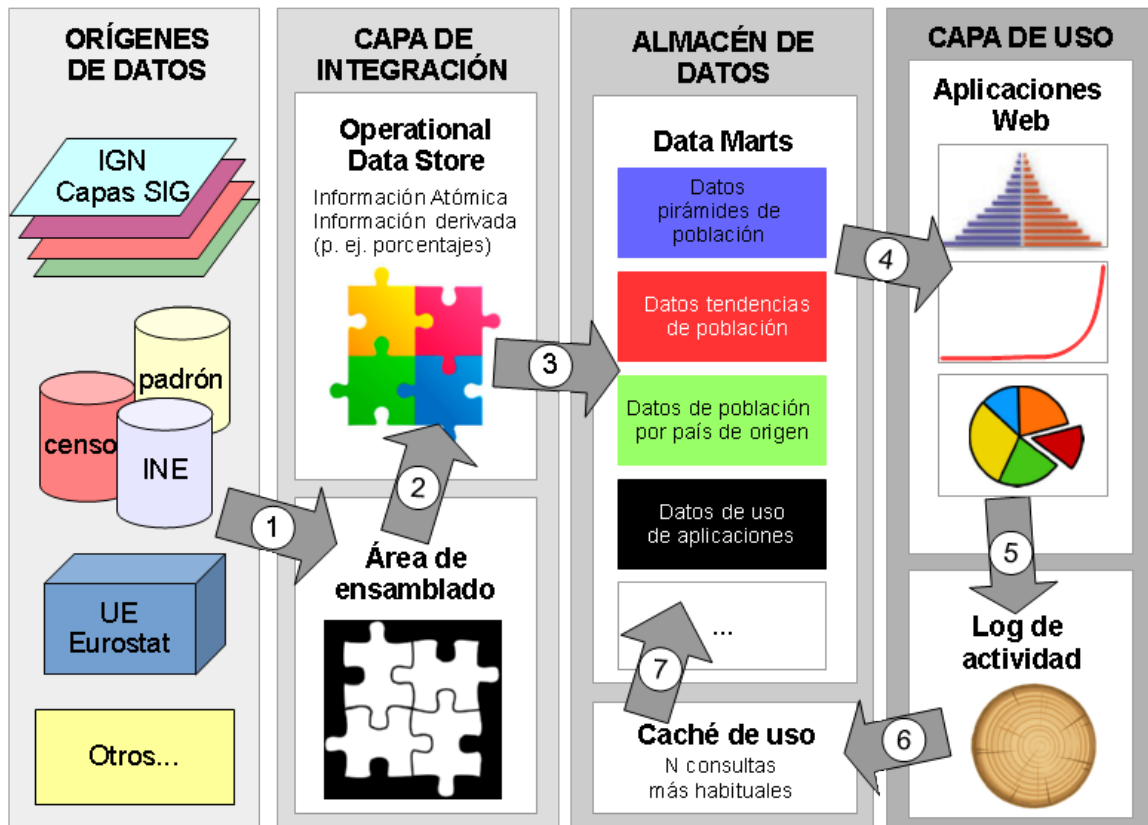
2.2 Data Warehouse: acceso e integración de datos abiertos

Un almacén de datos o Data Warehouse (DW) es un repositorio central de datos integrados de una o más fuentes dispares. Los datos almacenados en el DW provienen de distintos orígenes, pero también se pueden derivar en el mismo proceso en el que se generan las pirámides de población (p. ej. porcentajes, categorización por formas, etc). En la Figura 2 se describe de manera general el sistema de Data Warehouse implementado en este trabajo. El Data Warehouse se compone de varios subconjuntos o Data Marts, que son repositorios de datos donde se han integrado datos de distintos orígenes con una finalidad concreta, en este caso la representación de pirámides de población. Los otros componentes esenciales del Data Warehouse son el log de uso y la caché de actividad, que se integran dentro del DW para estudiar el uso que se hace de las visualizaciones, para así poder contestar las preguntas planteadas en la introducción.

En España, los datos necesarios para preparar pirámides de población son los proporcionados por el Instituto Nacional de Estadística (INE). Es posible componer pirámides de población municipales utilizando los datos de los Censos de Población (1981, 1991, 2001 y 2011) y los Padrones de Población (1996-2014). Además, a partir de 2004 se pueden crear pirámides a escalas inframunicipales utilizando datos de secciones censales. El INE facilita el acceso a estos datos a través de formularios web para consultas personalizadas o también mediante la descarga de ficheros de intercambio PC-Axis (*.px). PC-Axis es un software estadístico desarrollado por el gobierno sueco. En Europa se utiliza en Croacia, Dinamarca, Estonia, Finlandia, Irlanda, Islandia, Letonia, Lituania, Noruega, Groenlandia, Eslovaquia, Eslovenia, España y Suecia. Otros países y organizaciones que utilizan este formato son: Taiwán, Filipinas, Kuwait, Argelia, Mozambique, Namibia, Uganda, Sudáfrica, Tanzania, Bolivia, Brasil y organizaciones como la Comisión Europea, la ONU, la Comunidad del África Oriental y la FAO. Este programa no es un software de código abierto, sino que se trata de un freeware disponible para Windows. Entre las muchas otras opciones, este software permite crear pirámides de población, si se siguen unas instrucciones sobre la preparación de los datos. Sin embargo, otras tareas como por ejemplo comparar dos pirámides de población resultarán más trabajosas. El formato de PC-Axis (*.px) es básicamente un fichero CSV con una cabecera de metadatos y pueden ser leídos desde distintos lenguajes de programación (Javascript, Python o R.), sin necesidad de escribir código adicional. Los usuarios también pueden utilizar una extensión de PC-Axis para exportar los datos a Excel.

La existencia de múltiples formatos de intercambio débilmente estandarizados dificulta aún más la tarea de los docentes para su integración en bases de datos y el diseño de visualizaciones. Esto es aún más complicado cuando los datos publicados por una misma fuente cambian de unos ejercicios a otros (p. ej. municipios que se disgregan, desaparecen o cambios de regionalización). También se dan casos de datos homólogos que son publicados por organizaciones a distintos niveles jerárquicos (ayuntamientos, diputaciones, comunidades autónomas, INE, Eurostat, etc). Todos estos aspectos son los que se solucionan en las etapas de integración (ver Figura 2).

Figura 2. Plataforma propuesta como servicio de consulta de pirámides de población



En la Figura 2 se describe la arquitectura de la plataforma tecnológica propuesta. PostgreSQL es el Sistema Gestor de Bases de Datos escogido para la integración de los datos del INE, el mantenimiento de un registro de uso de la aplicación y un sistema de cacheado de las consultas más habituales. R GNU es el entorno para diseñar la visualización y R Shiny se ha utilizado como servidor de aplicaciones. Finalmente, Leaflet.js es el visor utilizado para filtrar las pirámides de población mostradas mediante una selección espacial (ver Figura 3).

2.3 Primeros resultados

En esta sección se describe la aplicación Web desarrollada para explorar las pirámides de población de España, y se presenta su interfaz de usuario. La aplicación puede ser consultada en pyramids.chichinabo.org. Efectivamente, en esta plataforma se publican servicios minimalistas y sencillos de utilizar, pero que cobrarán importancia con la incorporación de nuevos datos y con el análisis que se haga de las estadísticas de uso.

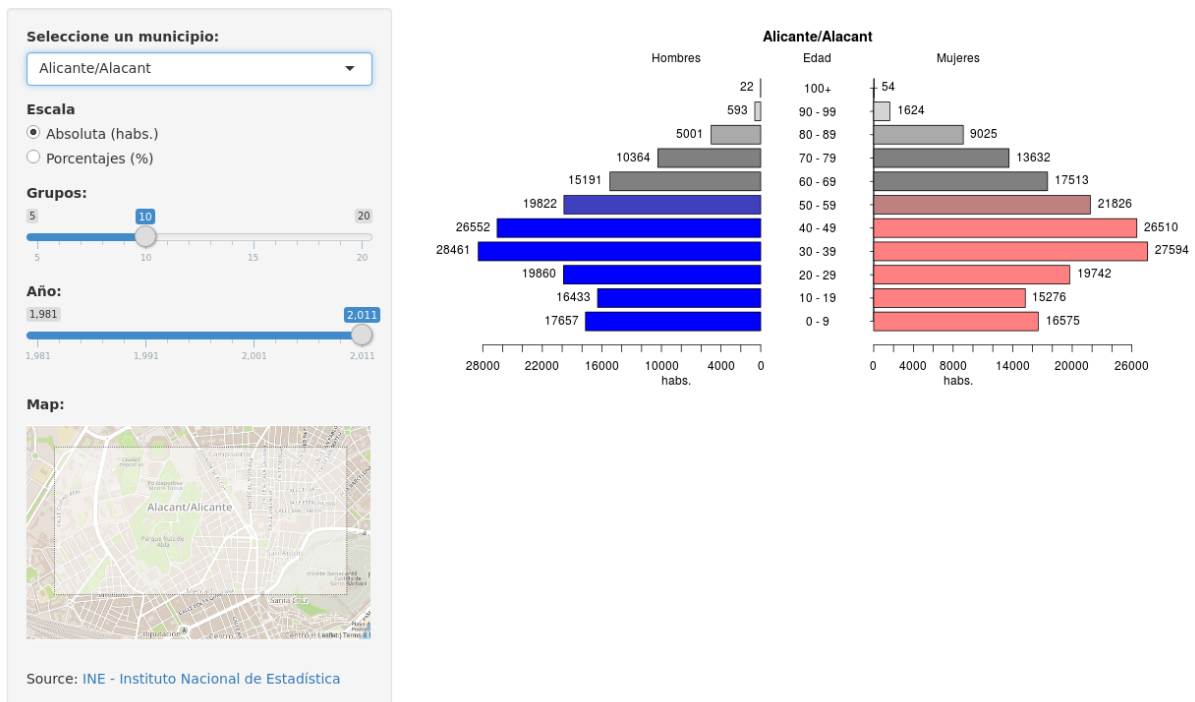
Al desarrollar esta visualización se ha tenido en cuenta su aplicación docente. El desarrollo de una aplicación web interactiva y minimalista disminuye la curva de aprendizaje

y no intimidada a los alumnos (Henry & Semple 2012). Aunque los temas relativos a la estructura de la población ocupan una parte importante de la enseñanza de la Geografía, no existe una herramienta Web adaptada para examinar estos temas en detalle y con datos locales (Lee et al. 2008). Por lo tanto, consideramos que continuar extendiendo esta visualización podría ser una solución para hacer frente a esta carencia y proponer nuevas visualizaciones que se puedan ajustar a la infraestructura propuesta. La gestión y consulta del registro de actividad de uso desde un enfoque de Big Data es esencial para responder a las dos preguntas planteadas al final de la introducción. El log de actividad puede utilizarse para identificar perfiles de usuario y patrones espacio-temporales de consultas mediante algoritmos utilizados habitualmente en Minería de Datos. PostgreSQL es compatible con MADlib, que es una librería de Minería de Datos ideal para evaluar las hipótesis señaladas en la introducción.

Por el momento, en este primer prototipo se trabaja exclusivamente con datos públicos del Instituto Nacional de Estadística (INE), agregados a distintas escalas (autonómica, provincial, municipal y censal) y para distintas fechas (censo y padrones). Con todo ello, la base de datos contiene más de 100.000 registros distintos, con los que poder crear gráficas interactivas únicas.

Figura 3. Captura de pantalla del interfaz de usuario

Estructura de la población, por edad y género



En la Figura 3 se puede ver los distintos componentes presentes en el interfaz de la aplicación web. El usuario puede filtrar las pirámides de población mediante sencillos widgets y después puede personalizar las gráficas con varias opciones (escala, colores, etc). Sin embargo, por motivos de escalabilidad, únicamente se muestran las primeras 10 gráficas que cumplen los parámetros especificados y el número total de resultados. Los distintos widgets que componen la web app son:

- Una lista de las 10 pirámides más importantes que cumplen los parámetros especificados por los usuarios, mostrando el número total de resultados coincidentes.
- Un panel gráfico donde se mostrará la pirámide y varios componentes para interactuar con ésta.
- Un mapa interactivo para seleccionar espacialmente los límites de las áreas administrativas.
- Una línea temporal para filtrar los años de las pirámides (fecha de referencia de los datos).
- Varios desplegables para limitar la búsqueda según las fuentes y otros metadatos.

3. CONCLUSIONES

La plataforma tecnológica desarrollada ha servido para comprobar dos aspectos importantes: (1) que hay datos abiertos en abundancia para el estudio de la estructura de la población en España y (2) que la solución tecnológica escogida permite crear un interfaz de usuario dinámico y escalable. Esto será decisivo en cuanto se incorporen nuevos orígenes de datos y aumente el número de usuarios simultáneos.

Las distintas opciones de registro de actividad permiten plantear varios tipos de experiencias docentes como las mencionadas en la introducción. Las principales preguntas que se puede plantear pasan por (1) evaluar el uso de visualizaciones como apoyo para metodologías constructivistas de aprendizaje y (2) comprobar los patrones geográficos de las consultas de pirámides de población. La opción más directa para explorar estas cuestiones es analizar el log de la base de datos mediante un algoritmo de reglas de asociación, encontrando así los patrones de las consultas realizadas por los usuarios. El elemento clave de esta plataforma es el log de actividad. El registro de actividad se convierte así en un repositorio para la aplicación de Big Data espacio-temporal (ver Figura 2). Es posible encontrar patrones temporales y geográficos de las consultas, identificar perfiles e incluso usuarios distintos. Sin

embargo, esta información cobraría mayor importancia si hubiese un interfaz de consulta disponible para los profesores, es decir, usuarios registrados en la plataforma que pudiesen consultar el log libremente, para conocer lo que ha interesado a sus estudiantes. De este modo, la Minería de Datos se establecería a dos niveles: (1) uso de la plataforma y (2) consultas realizadas por los profesores. Siguiendo las consultas de los profesores, estos ayudarían a identificar sesiones de trabajo en el aula y harían posible distinguir los grupos de alumnos (muchas consultas desde una misma dirección IP) de otros alumnos trabajando en casa, o incluso de otros usuarios.

Una vez desarrollada esta plataforma, el trabajo futuro de esta iniciativa pasa por incorporar nuevos orígenes de datos que cumplan los requisitos planteados en las secciones anteriores y mantener un listado de recursos. Simultáneamente, es posible derivar nuevos datos a partir de los datos de estructura de la población (p. ej. tendencias de población, clasificación de las pirámides según sus formas, sex-ratio, entre otras). Estos nuevos datos se podrían compartir con cada gráfica consultada, o utilizarse para introducir nuevos widgets de filtrado para consultar las pirámides de población.

4. REFERENCIAS BIBLIOGRÁFICAS

- Beringer, J. (2007). Application of Problem Based Learning through Research Investigation. *Journal of Geography in Higher Education*, 31(3), 445–457. doi:10.1080/03098260701514033
- Day, T. (2012). Undergraduate teaching and learning in physical geography. *Progress in Physical Geography*, 36(3), 305–332. doi:10.1177/0309133312442521
- Henry, P., & Semple, H. (2012). Integrating Online GIS into the K–12 Curricula: Lessons from the Development of a Collaborative GIS in Michigan. *Journal of Geography*, 111(1), 3–14. doi:10.1080/00221341.2011.549237
- Kim, M., Kim, K., & Lee, S.-I. S.-I. (2013). Pedagogical Potential of a Web-Based GIS Application for Migration Data: A Preliminary Investigation in the Context of South Korea. *Journal of Geography*, 112(3), 97–107. doi:10.1080/00221341.2012.709261
- Lin, H. (2012). Design and implementation of a mobile application for personal learning analytics, 15, 58–76. Retrieved from http://etd.lib.nsysu.edu.tw/ETD-db/ETD-search/view_etd?URN=etd-0118112-120714

- Liu, S., & Zhu, X. (2008). Designing a Structured and Interactive Learning Environment Based on GIS for Secondary Geography Education. *Journal of Geography*, 107(1), 12–19. doi:10.1080/00221340801944425
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook* (2nd ed.). New York: Springer New York. doi:10.1007/978-0-387-09823-4
- Naps, T. L., Rodger, S., Velázquez-Iturbide, J. Á., Rößling, G., Almstrum, V., Dann, W., ... McNally, M. (2003). Exploring the role of visualization and engagement in computer science education. *ACM SIGCSE Bulletin*, 35(2), 131. doi:10.1145/782941.782998
- Naps, T., Ross, R. J., Anderson, J., Fleischer, R., Kuittinen, M., McNally, M., ... Rantakokko, J. (2003). Evaluating the educational impact of visualization. *ACM SIGCSE Bulletin*, 35(4), 124. doi:10.1145/960492.960540
- Shaffer, C. a, Akbar, M., Alon, A. J. D., Stewart, M., & Edwards, S. H. (2011). Getting algorithm visualizations into the classroom. *Proceedings of the 42nd ACM Technical Symposium on Computer Science Education - SIGCSE '11*, 129. doi:10.1145/1953163.1953204

NOTAS:

ⁱ <http://public.tableau.com/profile/r.i.team.leicestershire.county.council#!/vizhome/WardPopulationPyramid-QuinaryAge/Dashboard>

ⁱⁱ <http://www.excelcharts.com/blog/beautiful-but-terrible-pyramids-tableau-edition/>

ⁱⁱⁱ <http://spark.rstudio.com/nsturaro/pyramid0/>