

Datos, patrones y narrativas: nuevas perspectivas sobre la Guerra Civil y la represión franquista a partir de la visualización de datos abiertos

Data, patterns and narratives: New perspectives on the Civil War and Franco's repression from the visualization of open data

Xavier Rubio Campillo

Institut de Recerca en Educació.

Universitat de Barcelona

University of Edinburgh

✉xrubio@ub.edu

Maria Feliu Torruella

Grup de recerca Didpatri

Universitat de Barcelona

✉mfeliu@ub.edu

Tania González Cantera

Master

Universitat de Barcelona

✉tania.gonzalez999@gmail.com

Rebut: 15/09/2020

Acceptat: 17/10/2020

Resumen

En los últimos años ha tenido lugar un incremento sustancial de bases de datos y portales en los que consultar información sobre la Guerra Civil Española y la represión franquista. Estos recursos son útiles para encontrar información sobre personas específicas, pero hasta ahora sus usos por parte de investigaciones centradas en dinámicas históricas de mayor escala ha sido escaso.

Este artículo presenta nuevas aproximaciones a la información contenida en estas nuevas bases de datos a partir de las Humanidades Digitales. Se analizan los desafíos presentados por portales de consulta en relación a las prácticas científicas actuales y, en especial, en relación a la adopción de políticas Open Data. En segundo lugar se plantea cómo el uso de metodologías cuantitativas de análisis proporciona nuevas perspectivas a la disciplina histórica. Este planteamiento se ejemplifica a partir de dos casos de estudio al que se han aplicado herramientas digitales distintas: a) la distribución espacial de fosas comunes mediante Sistemas de Información Geográfica y b) la magnitud de la represión franquista en Catalunya a partir de la visualización de datos.

Palabras clave: Humanidades Digitales, Sistemas de Información Geográfica, ciencia de datos, Open Data, represión franquista.

Abstract

Recent years have seen a substantial increase on the number of databases and web portals providing access to historical records related to the Spanish Civil War and Fran-

coist repression. These resources are useful to find information about specific individuals, but they have been seldom used to explore large-scale historical dynamics.

This work presents new Digital Humanities-based approaches to these newly accessible historical resources. We identify the challenges posed by the use of these web portals within the context of current scientific data practices and particularly the adoption of Open Data policies. Second, the paper explores how these resources facilitate the use of quantitative analytical methods able to create new perspectives for historical research. These new perspectives are exemplified through two case studies where different computational tools were applied: a) Geographical Information Systems to identify the spatial distributions of mass graves and b) data visualization to explore the scale of Francoist repression in Catalonia.

Keywords: *Digital Humanities, Geographical Information Systems, Data Science, Open Data, Francoist repression.*

INTRODUCCIÓN

Las actividades de recuperación de la memoria desarrolladas en las últimas dos décadas han comportado una mayor facilidad de acceso a la información generada por la Guerra Civil y a la represión franquista. Estas mejoras en el acceso a los datos históricos se ha centrado en la creación de distintos portales de información que permiten consultas sobre individuos concretos, con la intención principal que los familiares puedan conocer la experiencia y destino de las víctimas del conflicto y la dictadura. El ingente trabajo necesario para publicar estos datos ha facilitado estas consultas individuales, pero hasta qué punto ha contribuido a una mayor comprensión global de los hechos históricos?

Este incremento en el volumen de datos históricos accesibles no se da únicamente en el conflicto español, ya que es parte del gradual incremento de datos vinculado al desarrollo de las llamadas Humanidades Digitales y, de manera más general, a la explosión de datos generada por la sociedad actual¹. Dicha explosión ha creado nuevas necesidades en la gestión y acceso a grandes volúmenes de datos y nuevos campos de investigación como son la ciencia de datos, el Big Data o la visualización de información que, por desgracia, no parecen haberse implementado aún dentro de la disciplina histórica. Así, mientras que el volumen de datos sobre nuestro pasado ha aumentado de manera sustancial, los métodos de análisis usados en investigación histórica siguen siendo los mismos, y simplemente no son aptos para tratar semejantes volúmenes de información.

El presente trabajo explora los desafíos que la nueva realidad que las Humanidades Digitales generan para la investigación sobre la Guerra Civil y la represión franquista. La

1. BORGMAN, C. L. (2010). The digital future is now: A call to action for the humanities. *Digital Humanities Quarterly*, 3(4).

siguiente sección presenta un análisis crítico de la situación actual en cuanto a bases de datos existentes vinculadas a la guerra civil, con especial relación a las nuevas tendencias de investigación vinculadas al movimiento de ciencia abierta. A continuación las secciones tercera y cuarta ejemplifican el potencial de las Humanidades Digitales para analizar estos datos abiertos a través de una herramienta específica: la visualización de datos. Esta exploración del potencial analítica de las tecnologías digitales es analizada en base a dos casos de estudio: a) la localización de fosas comunes y b) las dinámicas de represión en Cataluña durante el franquismo. El trabajo finaliza con un resumen del potencial y posibles limitaciones de las metodologías cuantitativas brindadas por las Humanidades Digitales para mejorar la comprensión del impacto de la Guerra Civil y la represión franquista en la sociedad española.

DATOS, GUERRA CIVIL Y FRANQUISMO

La Ley de Memoria Histórica de 2007 motivó la progresiva creación de plataformas digitales de consulta sobre la Guerra Civil y el franquismo, con el objetivo de localizar y conocer el destino de familiares represaliados por el régimen franquista. Este objetivo primario es distinto al que tienen la mayor parte de bases de datos digitales vinculadas al pasado, cuyas finalidades están inspiradas en facilitar el acceso a la información para una mejor divulgación e investigación histórica, arqueológica y patrimonial. Esta divergencia en los objetivos genera una serie de limitaciones y desafíos que los portales de información sobre la Guerra Civil y la represión franquista tienen en cuanto a herramientas de investigación.

El portal más importante dentro de este contexto es el Centro Documental de la Memoria Histórica gestionado por el Gobierno de España². Este portal permite buscar información sobre numerosas bases de datos relacionadas con la memoria histórica, como serían listados de víctimas del franquismo, de militares de la Segunda República o de deportados a los campos nazis (pese a no incluir otros campos como los controlados por el régimen colaboracionista de Vichy). La utilidad de los recursos proporcionados por el Centro en cuanto a la búsqueda de información de individuos concretos es sustancial, pero al mismo tiempo el portal plantea serios desafíos al investigador en cuanto a a) entender la creación de la colección digital, b) acceder a los datos brutos y c) vincular información entre las distintas fuentes de datos.

En primer lugar, el acceso a datos históricos debería ir acompañado de extensa documentación sobre el origen de los datos y su casuística dentro de los metadatos, es decir, los datos que proporcionan información sobre la base de datos. Esta documentación

2. <https://www.culturaydeporte.gob.es/cultura/areas/archivos/mc/archivos/cdmh/bases-de-datos.html>

debería tener una rigurosidad similar a la de cualquier trabajo académico, y responder a preguntas como de qué archivos se ha recogido la documentación, qué limitaciones o sesgos tiene o quiénes son sus autores. Las colecciones disponibles en la plataforma mencionada están documentadas con desigual detalle, y no permiten en ningún caso un acceso sistemático a sus metadatos. El ejemplo más representativo de los problemas causados por la falta de metadatos es que se desconoce exactamente cómo deberían referenciarse las colecciones digitales dentro de un trabajo académico, uso que parece no haberse contemplado en su creación; algo tan simple como un DOI (Digital Object Identifier), identificador básico de cualquier entidad publicada en Internet, facilitaría cuanto menos la referencia a las fuentes de datos del Centro.

En segundo lugar, la funcionalidad disponible para acceder a las distintas bases de datos se centra en consultas específicas, y no permite acceder a los datos en bruto. Tampoco es fácil discernir bajo qué licencia de usos se pueden explotar o publicar estos datos por terceras partes como pueden ser historiadores, periodistas o ciudadanos interesados en la temática. Esta opacidad contrasta con el auge de las iniciativas Open Data, un movimiento que propugna la publicación en abierto con licencias de tipo libre de las bases de datos, especialmente las gestionadas por instituciones públicas³; si estas bases de datos surgen de los impuestos públicos entonces todos los ciudadanos deberían poder acceder a ellas, reproducirlas y usarlas como crean convenientes, dentro de los límites establecidos por las licencias abiertas (p.e. siempre reconociendo la autoría de los datos originales). Todos los gobiernos, incluyendo el español y las comunidades autónomas, han creado portales de transparencia dónde se pueden encontrar los datos abiertos que gestionan pero, excepto algunos casos concretos, estos portales de Open Data gestionados por las administraciones públicas no contienen información sobre bases de datos de la Guerra Civil como puedan ser las del Centro Documental de la Memoria Histórica.

Más allá de la gestión abierta de los datos públicos, la adopción de políticas Open Data es hoy en día esencial para una ciencia de calidad⁴ gracias a motivos diversos: la publicación de datos con licencias abiertas permite fomentar la cooperación entre investigadores, el acceso a nuevas fuentes de datos, una mayor capacidad efectiva para comunicar los resultados y, especialmente, una mejor capacidad para replicar y evaluar el trabajo de otros investigadores. Así, en la actualidad la publicación en abierto de las

3. KITCHIN, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.
4. LOWNDES, J. S. S., BEST, B. D., SCARBOROUGH, C., AFFLERBACH, J. C., FRAZIER, M. R., O'HARA, C. C., JIANG, N., & HALPERN, B. S. (2017). Our path to better science in less time using open data science tools. *Nature Ecology & Evolution*, 1(6), 0160. <https://doi.org/10.1038/s41559-017-0160>; Molloy, J. C. (2011). The Open Knowledge Foundation: Open Data Means Better Science. *PLoS Biology*, 9(12), e1001195.

bases de datos relacionadas con la investigación se revela como una pieza clave para mejorar la actividad de cualquier disciplina del conocimiento que se autodenomine científica, incluyendo la investigación histórica. La razón es simple: sin acceso a los datos que he analizado y que a priori sustentan mis hipótesis... cómo pueden otros investigadores comprobar que mi análisis es correcto?

Una ventaja crítica de las políticas de acceso abierto es la libertad que brinda a los ciudadanos para explorar y explotar los datos del pasado. Un ejemplo dentro de la investigación del conflicto histórico sería la publicación bajo licencia abierta de los «Theater History of Operations Reports» o THOR por parte de la iniciativa data.mil diseñada por el Departamento de Defensa de los Estados Unidos. La base de datos THOR reúne información sobre decenas de miles de operaciones de bombardeo llevados a cabo por la Fuerza Aérea Americana durante el siglo XX. Pese a que el proyecto ha finalizado, en la actualidad se pueden encontrar decenas de usos de esta base de datos que van desde la visualización de los bombardeos aliados durante la Segunda Guerra Mundial⁵ y Vietnam⁶ a la confección de tutoriales específicamente diseñados para que los historiadores aprendan a usar las herramientas digitales en bases de datos históricas como THOR⁷.

Finalmente, el incremento exponencial en el número de bases de datos abiertas ha tenido como consecuencia un nuevo reto tecnológico ciertamente ambicioso: sería posible consultar a la vez la información contenida en múltiples bases de datos? Esta pregunta es el punto de partida del protocolo Linked Open Data que, promovido por el creador de la World Wide Web, Tim Berners-Lee, permite conectar información de múltiples bases de datos⁸. Esta revolución en el acceso a los datos se ha desarrollado a través de la tecnología de bases de datos semánticas y el concepto de las ontologías: modelos de conocimiento que, a través de una nueva capa común, permiten relacionar múltiples colecciones digitales con solapamiento temático.

Linked Open Data es el corazón de iniciativas tan ambiciosas como Europeana⁹, un portal que permite búsquedas avanzadas sobre patrimonio cultural europeo a través de una ontología común a la que se han acoplado centenares de bases de datos abiertas publicadas por instituciones europeas como museos, archivos y galerías, contando en la actualidad con más de 50 millones de registros. A una escala menor, y centrándonos en la

5. <https://dlozeve.github.io/ww2-bombings/>

6. <https://www.kaggle.com/usaf/vietnam-war-bombing-operations>

7. HARPER, C. (2018). Visualizing Data with Bokeh and Pandas. *The Programming Historian*, 7. <https://doi.org/10.46430/phen0081>

8. BIZER, C., HEATH, T., & BERNERS-LEE, T. (2011). Linked data: The story so far. In *Semantic services, interoperability and web applications: Emerging concepts* (pp. 205–227). IGI Global.

9. <https://www.europeana.eu/>. Para su creación ver Purday, J. (2009). Think culture: Europeana. eu from concept to construction. *BIBLIOTHEK Forschung Und Praxis*, 33(2).

temática del conflicto histórico, el proyecto WarSampo¹⁰ permite acceder a información vinculada a la Segunda Guerra Mundial en Finlandia, incluyendo individuos, lugares, hechos históricos y cementerios de guerra.

La aproximación de Linked Open Data no está contemplada en las bases de datos del Centro Documental de la Memoria Histórica; como consecuencia, su utilidad se ve limitada por la incapacidad de realizar búsquedas cruzadas a través de múltiples bases de datos. Sería relativamente sencillo implementar una ontología de la Guerra Civil que definiera conceptos como «fosa común», «víctima» o «batalla», y usar los beneficios del Linked Open Data para enlazar todas las bases de datos del Centro Documental. Esto permitiría consultas que incluyeran información agregada de personas que, por ejemplo, hubieran servido en el ejército de la Segunda República y posteriormente hubieran sido represaliadas, fuera con años de prisión o bien se hubieran localizado sus restos en una fosa común.

Algunas fuentes oficiales y iniciativas de tipo voluntario corrigen parcialmente algunos de estos desafíos del acceso a los datos de la Guerra Civil y el franquismo. El proyecto SIDBRINT (Sistema d'Informació Digital sobre les Brigades Internacionals)¹¹ ofrece un portal de investigación sobre los brigadistas internacionales en base a tres bases de datos (brigadistas, unidades militares y fuentes documentales). La iniciativa fue creada a partir del proyecto de investigación HAR2010-20983 desarrollado entre 2010 y 2013, para posteriormente ser gestionada como portal institucional de la Universidad de Barcelona. En este caso el portal ofrece amplia información sobre la autoría y contribuyentes de la base de datos y, además, el portal establece claramente los usos de la base de datos según una licencia Creative Commons. De manera paralela el Memorial Democràtic ha publicado un portal de metabúsquedas que, bajo el nombre de «Banc de la Memòria Democràtica»¹², permite hacer una búsqueda general en diversas bases de datos aunque a través de una interfaz de consulta limitada y sin aplicar en realidad el protocolo Linked Open Data.

Diversos gobiernos autonómicos también han publicado datos de la Guerra Civil en formatos abiertos, que ayudan a mitigar la limitación general de acceso. El sesgo causado por las demarcaciones autonómicas estará ciertamente presente, pero la información recogida puede ser muy interesante siempre y cuando el investigador sea consciente de las limitaciones causadas por esta fragmentación. Así, en el portal de open data de

10. <https://www.sotasampo.fi/en/>. Más información en HYVÖNEN, E. (2020). 'Sampo' Model and Semantic Portals for Digital Humanities on the Semantic Web. *DHN*, 373–378.

11. Portal disponible en <http://sidbrint.ub.edu/>. Para más información PRADES ARTIGAS, L. (2012). *Sistema de información digital sobre las brigadas internacionales: Brigadistas, fuentes documentales y bases de datos (SIDBRINT)*. Univ de Castilla La Mancha.

12. <https://banc.memoria.gencat.cat>

la Generalitat de Catalunya¹³ se pueden encontrar diversos datasets vinculados a esta temática, como por ejemplo un censo de personas desaparecidas, un listado geográfico de fosas comunes o un listado de represaliados por el franquismo. Los datos brutos son ofrecidos con abundante información sobre su uso en abierto y, pese a que no se incluyen facilidades para transformar el conjunto en Linked Open Data, sería una funcionalidad ciertamente posible.

De manera paralela a estos esfuerzos han surgido varias iniciativas de tipo voluntario, que reclaman una mayor apertura de los datos vinculados a la Guerra Civil. En este contexto destaca el proyecto Innovation and Human Rights (IHR)¹⁴. Esta asociación sin ánimo de lucro, fundada en 2016, tiene como objetivo principal facilitar el acceso a la información de la Guerra Civil a partir del refinamiento y reutilización de datos publicados por otras fuentes gracias, cuando es posible, a licencias de datos abiertas. La tarea de entidades no gubernamentales como IHR plantea otro desafío al acceso a los datos relacionados con la historia del siglo XX: la apertura de información sobre individuos puede entrar en conflicto con las leyes de protección de datos personales¹⁵. La problemática en este caso es que algunas de las personas que aparecen en las bases de datos podrían seguir vivas y, como consecuencia, sería ilegal abrir la información a terceros porque según el espíritu de la ley no se pueden publicar datos personales que puedan dañar a individuos. Dicho espíritu es ciertamente vital para proteger los datos personales que generamos a diario, pero nada tiene que ver con la investigación histórica, y en ningún caso debería justificar la limitación al acceso de información sobre la represión franquista. Por otra parte, por otra parte, no está claro por qué la ley aplica sólo cuando preguntamos por muchas personas y, en cambio, sí que es posible acceder a la información de un solo individuo a través de los portales gubernamentales. Finalmente, no es este un problema vinculado tan sólo a la investigación histórica, y una posible solución sería buscar vías para la anonimización de los datos en caso de registros de personas aún vivas que, en realidad, son tan sólo un pequeño porcentaje del total.

Como vemos, los tímidos avances en la digitalización y publicación de datos relacionados con la Guerra Civil y el franquismo van en la dirección correcta, pero aún queda un largo camino para recorrer hasta un escenario alineado con los preceptos de la ciencia abierta en los que se reconozca la autoría de las bases de datos, éstas sean accesibles de manera libre y, finalmente, sea posible agregar información de todas las fuentes disponibles. Pese a ello, existen herramientas de análisis que permiten una exploración cuantitativa a gran escala de los registros existentes, con el objetivo de complementar las

13. <https://analisi.transparenciacatalunya.cat>

14. <https://ihr.world/>

15. Ver por ejemplo la negativa de la Generalitat a ofrecer un censo de desaparecidos a IHR en <https://ihr.world/ca/2018/10/05/no-a-lacces-als-noms-de-victimes-de-desaparicio-forcada-de-la-guerra-civil/>

tradicionales aproximaciones cualitativas típicas de la investigación histórica y mejorar así nuestra comprensión del conflicto y la posterior represión franquista; a continuación proponemos dos aproximaciones a este desafío a través de los Sistemas de Información Geográfica y la visualización de datos.

CARTOGRAFÍA DE LAS FOSAS COMUNES

La mayor parte de la información usada en el estudio del pasado contiene algún tipo de dimensión espacial. Sean las coordenadas de un yacimiento, un texto mencionando topónimos, o una descripción de rutas de comercio, este tipo de información puede servir para identificar estructuras espaciotemporales; es decir, correlaciones entre dinámicas históricas y las coordenadas espaciales en las que tuvieron lugar.

La Guerra Civil y el franquismo no son ajenos a esta característica espacial de la información histórica, y han generado un vasto volumen de datos espaciales. Un ejemplo es el llamado «Mapa de Fosas» publicado por el Ministerio de la Presidencia del Gobierno de España en 2008¹⁶. El portal permite consultar los datos a nivel textual a partir de dos formularios (por fosas y por víctimas), y facilita también la exploración espacial a partir de un mapa interactivo como se ve en la Figura 1.

En primer lugar, cabe decir que la información proporcionada por este portal es obsoleta, ya que no parece haber incorporado la intensa labor de exhumación de fosas llevada a cabo a lo largo de todo el territorio español desde el 2008 hasta la actualidad. Es este un punto crítico de cualquier proyecto de bases de datos vinculado a un programa de financiación pública: las iniciativas tienen una fecha de ejecución específica y, al acabarse ésta, no hay mecanismos para continuar el proyecto. La falta de presupuestos para el mantenimiento y actualización de las bases de datos es un problema común en los proyectos de investigación, para el que desafortunadamente no hay una solución fácil. Así se genera un conflicto entre la necesidad de proveer de presupuesto estable estas iniciativas y la estructura misma de la financiación en investigación pública.

Más allá de la actualización de los datos, el mapa está diseñado para permitir el acceso a información básica de las fosas por localidad, pero sin funcionalidad adicional que permita explorar los datos con una mayor profundidad espacial, pese a que es posible conseguir abundante información adicional a nivel textual. A nivel práctico el portal no permite descargar directamente los datos, ni acceder a ellos a través de una API (Application Programming Interface) que permita a otras aplicaciones obtener los datos mediante algún formato compatible. Finalmente, no se explicita licencia de uso ni hay una autoría clara, así que el ciudadano que desee explorar estos datos tendrá una difícil tarea para acceder a ellos o contactar a sus creadores.

16. <https://mapadefosas.mjusticia.es/>

Información **Mapa de Fosas** Buscador Fosas Buscador Víctimas Aportación Ciudadana de Información

Mapa de fosas : - Ley de la Memoria Histórica

Permite la localización geográfica de las fosas o lugares de enterramiento en el territorio español de las que se dispone información, representadas con símbolos de diferentes colores según la general del territorio Español o de la Península. También se puede realizar zoom manual sobre el mapa a través de la barra diseñada al efecto o haciendo doble clic sobre el mapa. Puede pulsar el mapa. Para más información, pulse [aquí](#)

Mostrar localización en el mapa

Com. Autónoma: - Todas - Provincia: - Todas - [Península](#) [Territorio](#)

[Español](#)

Leyenda

- No Interventida
- Desaparecida
- ▲ Trasladaada al Valle de los Caídos
- Exhumada Total o Parcial
- ★ Valle de los Caídos
- Varias Fosas en la misma Localidad

Fuente de Mapas: IGN - Instituto Geográfico Nacional

Información de la Ubicación seleccionada

Número de Fosas No Interventidas: 1 ■

Número de Registro: 2228/2010 BARC
Denominación Fosa: Fosa en el Cementerio de Sant Joan Despi
Localidad: Sant Joan Despi (Barcelona)
Localidad: Sant Joan Despi
Tipo:
Actuación: NO INTERVENIDA
Estado Actual:
Víctimas: 0
[Más detalle](#)

Número de Fosas Trasladaadas al Valle de los Caídos: 2 ▲

Número de Registro: 948/2009 BARC
Denominación Fosa:
Localidad: Cornellà de Llobregat (Barcelona)
Localidad: Cornellà de Llobregat
Tipo:
Actuación: TRASLADADA AL VALLE DE LOS CAÍDOS
Estado Actual:
Víctimas: 1
[Más detalle](#)

Número de Fosas Desaparecidas: 1 □

Número de Registro: 2296/2011 BARC
Denominación Fosa: Fosa en el monte Sant Pere Martí
Localidad: Sant Just Desvern (Barcelona)
Localidad: Sant Just Desvern
Tipo:
Actuación: DESAPARECIDA
Estado Actual:
Víctimas: 0
[Más detalle](#)

Figura 1: Interfaz gráfica del mapa de fosas. La selección de los símbolos desplegados en el mapa permite acceder a información textual de las fosas según localidad.

En este contexto la única opción posible es hacer uso del llamado *web scraping*: la extracción sistemática de los datos ofrecidos por una web a partir de programación. En el caso que nos ocupa se implementó un pequeño programa o *script* en el lenguaje de programación Python 3, con la intención de trasladar los datos ofrecidos en la web a un formato más fácil de usar, como por ejemplo el formato CSV (Comma Separated Value). Este CSV replicaba la información del portal web en base a una estructura *tidy data*, en la que cada fila es una observación (en este caso una fosa) y cada columna una variable (coordenadas geográficas, nombre, estado, número de víctimas...)¹⁷.

Dicha estructura *tidy data* está estandarizada en la ciencia de datos y, como consecuencia, es posible importar fácilmente el fichero CSV obtenido por *web scraping* a

17. WICKHAM, H. (2014). Tidy data. *Journal of Statistical Software*, 59(10), 1–23.

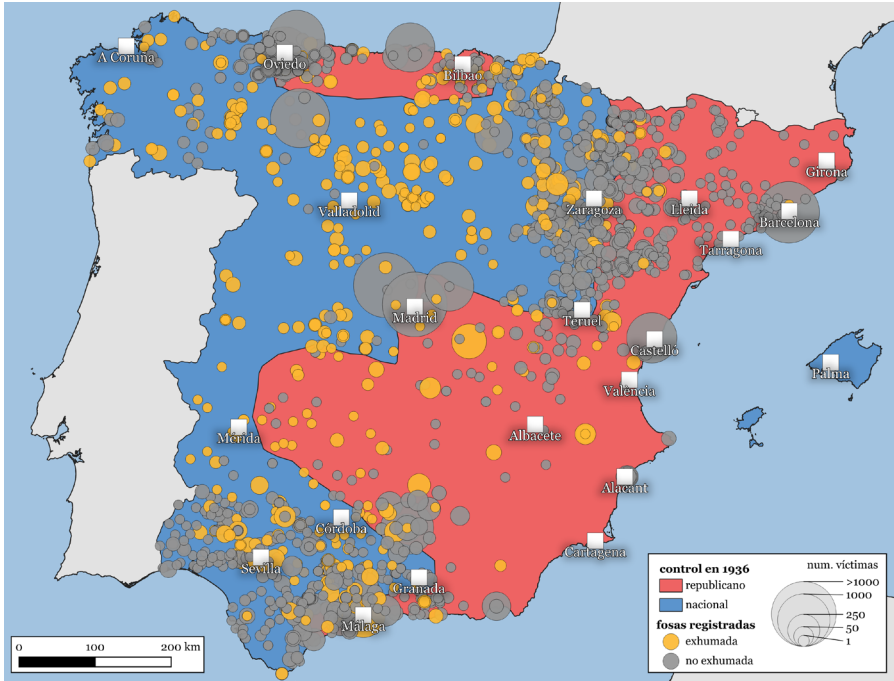


Figura 2: Distribución espacial de las fosas registradas en 2008. Cada punto es una fosa, y su tamaño es proporcional al número estimado de víctimas enterradas.

cualquier programa de análisis como por ejemplo un Sistema de Información Geográfica o GIS (*Geographical Information System*). Los GIS son la herramienta ideal para el análisis espacial de grandes volúmenes de datos conteniendo información geoespacial. En este caso concreto, la importación del fichero CSV a QGIS, un GIS de código abierto, permite generar cartografía más ilustrativa de la escala y tipología de las fosas comunes de la Guerra Civil, como se puede ver en el mapa de la Figura 2.

Este mapa ofrece una visualización alternativa de los mismos datos, donde se puede observar información adicional a la ofrecida por la cartografía del portal oficial. Ciertamente la eficacia de cualquier visualización está limitada por la calidad de los datos que usa, y en este caso la falta de actualización de los datos hace que este mapa no muestre la intensa actividad de la última década; pese a ello el mapa revela diversos patrones que no son identificables en la plataforma original.

Primero, la adición de una capa mostrando los territorios controlados por ambos contendientes en 1936 aporta información temporal vinculada a la distribución espacial de fosas, bajo la asunción que la mayoría de fosas fueron creadas por la ejecución de di-

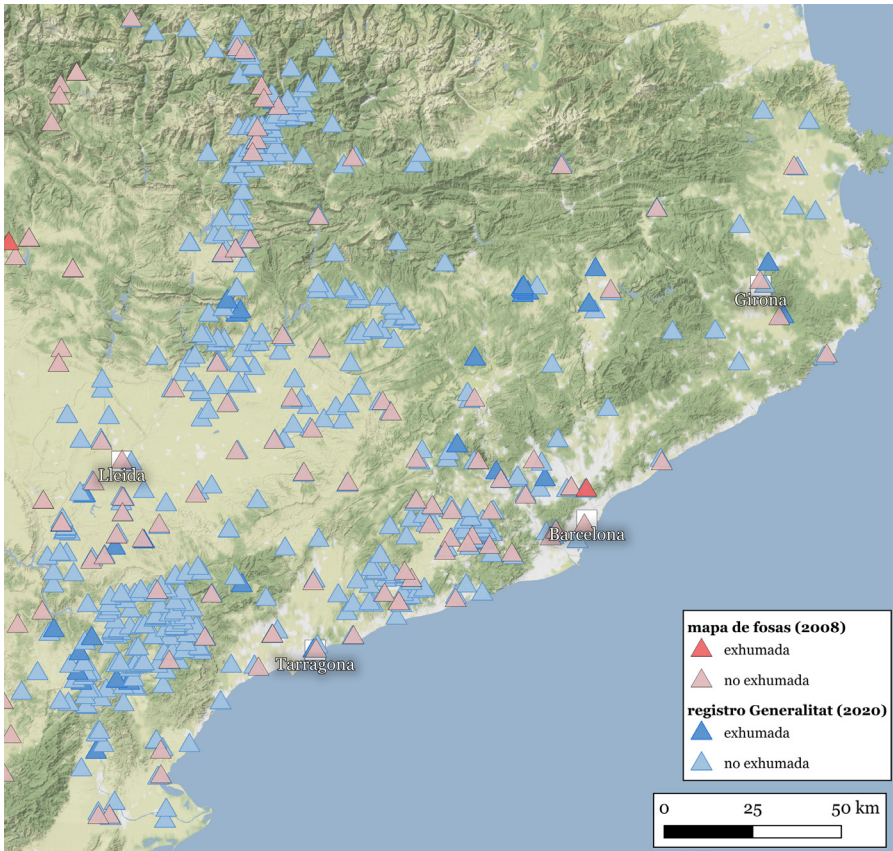


Figura 3: Comparativa de las fosas registradas en Catalunya según el mapa de fosas de 2008 y los datos actualizados ofrecidos en abierto por la Generalitat.

sidentes en la zona ocupada por el ejército nacional. Al mismo tiempo, el símbolo que identifica cada fosa tiene un tamaño proporcional al número estimado de víctimas en ella, por lo que el usuario puede hacerse una mejor idea de la escala de la represión, así como de posibles diferencias regionales. Finalmente, se ha simplificado el esquema de colores a tan solo dos (exhumada o no exhumada) para evidenciar de manera más explícita que la inmensa mayoría de ellas no había sido aún excavada en 2008. Este es un tema clave de cualquier visualización: las decisiones sobre su diseño están tomadas en función de su utilidad, así como de los objetivos de investigación.

En este caso hemos construido una visualización ajustada al objetivo de mostrar cómo el territorio afecta a la tipología y estado de las fosas. Se puede observar que la mayor parte de fosas están localizadas en tres regiones: la conquista inicial de Andalucía por parte de las tropas nacionales, el frente de Aragón y, a cierta distancia, el frente Norte y Cataluña. Por otra parte, las fosas con más víctimas, con la obvia excepción del Valle de los Caídos, están localizadas en ciudades con un control inicial por parte de los republicanos y, por lo tanto, sus víctimas serían ejecutadas después de 1936.

La distribución de los datos no sólo responde a la creación de las fosas, sino también a sesgos actuales en relación al patrimonio de la Guerra Civil y a las intervenciones arqueológicas. Es relevante destacar la inexistencia de fosas comunes exhumadas en 2008 en comunidades como Cataluña y Asturias. Otras regiones, como por ejemplo Andalucía o Castilla y León, mostraban una dinámica distinta, con numerosas fosas exhumadas aunque distribuidas de manera desigual; esta diversidad podría responder a diferencias en la sensibilidad política de cada municipio, así como la actitud del gobierno autonómico, o la mayor actividad de asociaciones para la recuperación de la memoria histórica.

Desde esta perspectiva actual es interesante visualizar el dataset de fosas comunes publicado bajo licencia abierta por la Generalitat de Catalunya en abril del 2017 (frecuentemente actualizado). Como se puede observar en la Figura 3, consta de un número de registros sensiblemente mayor que el conjunto del mapa de fosas para Cataluña (500 versus 150).

Este contraste pone en evidencia la falta de actualización de los datos estatales, ya que otras administraciones autonómicas han llevado a cabo tareas similares de recopilación de datos actualizados sobre las fosas comunes. En este contexto es importante destacar que una plataforma Linked Open Data, capaz de integrar la información recogida y actualizada por las distintas administraciones, resolvería el problema de la sincronización y agregación de datos sobre la excavación de fosas comunes de una vez por todas.

VISUALIZANDO LA REPRESIÓN FRANQUISTA

La visualización de datos es una técnica exploratoria que permite identificar patrones complejos en grandes volúmenes de datos¹⁸. Este marco de trabajo permite al investigador desplegar un catálogo diverso de tipologías de visualización con dos objetivos específicos: a) identificar posibles dinámicas y relaciones de causa-efecto presentes en los datos y b) presentar los patrones identificados al lector de manera didáctica a modo de narrativa visual. La elección de una visualización u otra dependerá de la tipología de los datos (p.e.

18. CAIRO, A. (2012). *The Functional Art: An introduction to information graphics and visualization*. New Riders; Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed). Graphics Press.

valores numéricos, categorías, etc.), así como de las preguntas que el investigador quiera explorar y los puntos clave que se quieran destacar de la colección de datos.

Para este segundo caso de estudio hemos elegido explorar la base de datos de víctimas de la justicia franquista en Catalunya, publicada bajo licencia abierta por el Arxiu Nacional de Catalunya¹⁹. Esta fuente de datos contiene información de las víctimas de la represión franquista a través de las causas legales que se abrieron contra ellas, incluyendo datos personales (nombre y apellidos, lugar de nacimiento, edad, etc.) así como de la causa (tipo de proceso, pena impuesta, posible indulto). Nos hemos centrado en tres preguntas de investigación: a) ¿de dónde proceden los represaliados juzgados en Catalunya?, b) ¿como evolucionó la represión durante el franquismo? y c) ¿qué edad tenían los represaliados?

Inmigración y represión

La base de datos publicada por el ANC contiene información espacial sobre el municipio de nacimiento de cada encausado, así como el municipio de residencia. Una pregunta relevante es hasta qué punto la represión franquista se dirigió al colectivo inmigrante residente en Catalunya, vinculado tradicionalmente a los movimientos sindicalistas radicales o, por el contrario, fue ejercida de manera indiscriminada contra toda la población residente en territorio catalán.

Desde un punto de vista de ciencia de datos, cabe destacar que esta información es de tipo espacial, pero en este caso un mapa no sería la mejor solución para visualizar los datos. Un mapa de calor como el de la Figura 4 permite comparar mejor los números de cada región (en este caso, comunidades autónomas).

Las comunidades autónomas (eje Y) están ordenadas de mayor a menor volumen de encausados en Catalunya y nacidos en ellas. Como es esperable la mayoría de víctimas son de origen catalán: alrededor del 70% en Barcelona y hasta un 90% en las otras tres provincias. En cuanto al origen de las víctimas emigradas, los datos parecen reflejar un balance representativo de la emigración recibida por Cataluña durante la primera oleada migratoria del siglo XX, siendo el grueso de los encausados originarios de Andalucía, Valencia, Aragón y Murcia²⁰.

19. La base de datos está disponible en formato CSV en <https://analisi.transparenciacatalunya.cat/Mem-ria/Llista-de-reparaci-jur-dica-de-v-ctimes-del-franqu/3bjt-k7vu>; los datos originales fueron publicados en 2018 pero son frecuentemente actualizados.
20. Porcentajes parecidos en el origen de los inmigrantes se pueden observar para casos concretos como el Hospitalet de Llobregat en CAMÓS i CABECERAN, J. (2009). L'Hospitalet i la immigració: Catalanistes i anarquistes als anys trenta. *Quaderns d'estudi*, 69–84.

Catalunya	23495	5646	8283	8334	
Andalucía	2843	142	133	116	
Comunitat Valenciana	2609	127	96	293	
Aragón	2413	74	324	214	
Región de Murcia	2428	79	63	65	
Castilla y León	1141	84	149	73	
Castilla-La Mancha	779	61	63	45	
Comunidad de Madrid	646	29	26	33	
Galicia	422	29	24	22	
Extremadura	286	43	39	35	
Illes Balears	310	24	15	17	
País Vasco	232	16	13	10	
Navarra	202	11	14	7	
Asturias	150	2	24	11	
Cantabria	144	10	7	12	
La Rioja	150	3	14	1	
Ceuta y Melilla	77	1	4	1	
Canarias	40	1	4		
		Barcelona	Girona	Lleida	Tarragona

Figura 4: mapa de calor de causas abiertas dividida por provincia de residencia (eje X) y comunidad autónoma de nacimiento de la víctima (eje Y). El número en cada celda es el número de causas abiertas, y está vinculado al color de la celda (azul para volúmenes bajos y rojo para los más altos).

Esta preeminencia de encausados de origen catalán, pese a ser lógica, da pie a algunas reflexiones sobre la relación entre emigración y represión. En otras obras se ha destacado que los movimientos sindicalistas anarquistas eran mayoritariamente compuestos por inmigrantes de primera generación²¹; si asumimos que buena parte de los represaliados por la justicia franquista pertenecieron a dichos movimientos obreros y, en particular, a FAI y CNT, entonces la hipótesis no parece estar especialmente sustentada por este conjunto de evidencia, ya que se esperaría una mayor proporción de encausados nacidos fuera de Cataluña. Así, el patrón identificado en estos datos daría mayor

21. Ver por ejemplo OYÓN, J. L. (2015). Mundo obrero, inmigración y radicalismo cenetista en la Barcelona de la década de 1930. *Cercles: Revista d'història Cultural*, 9–20.

plausibilidad a hipótesis alternativas, en las que se defiende una mayor integración social e ideológica entre población inmigrante y nativa²².

Esta aparente contradicción entre hipótesis y datos es una de los beneficios más atractivos de herramientas cuantitativas como la visualización de datos: permiten identificar patrones complejos en los que la evidencia no se ajusta a lo que esperaríamos a partir de nuestras ideas preconcebidas. Así, un marco de trabajo cuantitativo permite plantear nuevas preguntas de investigación capaces de hacer avanzar nuestra comprensión de los hechos históricos, además de dotar al investigador de herramientas mucho más efectivas que los modelos verbales en cuanto a la evaluación de hipótesis se refiere²³. En este caso la divergencia se podría dar por múltiples causas, incluyendo una mayor presencia de sindicalistas de origen catalán de lo anteriormente esperado (entre ellos, segunda generación de inmigrante), un mayor grado de exilio de anarquistas inmigrantes después de la derrota o hasta una priorización de la justicia franquista para perseguir al movimiento catalanista por encima del sindicalista, por nombrar tan sólo algunas posibles hipótesis.

La dinámica temporal de la represión franquista

La segunda pregunta planteada tiene relación con las fechas en las que dichas causas fueron procesadas. Se intuye que buena parte de las mismas estaría vinculada a los años inmediatamente posteriores a la derrota republicana y al proceso de «depuración ideológica» realizado por la dictadura franquista. Pero hasta qué punto la represión fue continuada?

La Figura 5 articula los procesos como una serie temporal dividida en dos gráficos: las causas abiertas y las que acabaron en pena de muerte para las víctimas sin indulto alguno.

22. Algunos estudios locales sobre esta temática desarrollados en el área metropolitana de Barcelona apoyan esta visión integradora. Destaca el caso de L'Hospitalet de Llobregat, donde la CNT «tenía una incidencia importante en los tres barrios de la ciudad», los cuales tenían características poblacionales radicalmente distintas. Por tanto, se trata de un «tópico simplista» equiparar la «FAI y La Torrassa» (cuyo 85% de la población en 1930 era inmigrante). En CAMÓS i CABECERAN, J. (2009). L'Hospitalet i la immigració: Catalanistes i anarquistes als anys trenta. *Quaderns d'estudi*, 69-84 y CAMÓS i CABECERAN, J. (1986). El mite de la Torrassa. *L'Hospitalet 1931-1936. Plecs d'història local* nº 1, 6-13.
23. Para una discusión sobre la importancia de cuantificar la plausibilidad de las hipótesis históricas ver por ejemplo RUBIO-CAMPILLO, X., Lorenzo, J. B., MONTANIER, J., DIAZ, J. M., GONZÁLEZ, J. P., RULL, G., & REMESAL, J. (2018). Provincias, sellos e hipótesis nulas: La identificación de rutas de comercio a través de medidas de distancia cultural. In *Cuantificar: Qué, cómo y para qué. Quantification in Classical Archaeology: Objects, methodologies and aims* (pp. 237-249). Universitat de Barcelona.;

El patrón más evidente en estas series temporales es que la gran mayoría de causas fueron abiertas durante los primeros años de posguerra, y que buena parte de ellas condujeron a penas de muerte. El año de 1939 observa más del triple de causas que ningún otro año y, por tanto, el gráfico evidencia la rapidez e intensidad con la que el nuevo régimen reprimió a sus enemigos para asegurar el control de Cataluña y castigar al bando perdedor. Los dos gráficos muestran claramente este gran volumen de causas iniciales, pero precisamente por ese motivo no es posible usarla para explorar las dinámicas posteriores. Una posible solución es usar una escala logarítmica para el eje Y, que permiten estudiar series de datos con distribuciones muy sesgadas como esta (ver Figura 6). Este segundo gráfico muestra cómo en los años posteriores el volumen de causas disminuye, especialmente a partir de 1952, debido posiblemente a la firma de los Pactos de Madrid y el fin del período de aislamiento internacional del régimen, pero vuelve a subir hacia el final del franquismo, a causa de los múltiples movimientos que amenazan al régimen.

Si nos fijamos en los casos que acabaron en penas de muerte el patrón es aún más aparente. La Figura 7 explicita el alcance y dureza de la represión a partir de un diagrama de dispersión, en el que cada punto es una persona ejecutada por el franquismo.

En esta visualización se ha cruzado la información temporal con una variable adicional: la edad de las víctimas. La mayoría de los represaliados tenían edades entre los 20 y los 40 años, aunque la variancia de la muestra es ciertamente elevada, con un rango (i.e. valor mínimo y máximo) que va de los 17 a los 78 años.

Una dinámica evidente en el conjunto de las tres gráficas es el descenso súbito del número de causas después de 1943. Así, por ejemplo, de las 145 ejecuciones de 1943 se pasaron a 13 en el año siguiente; la pena de muerte seguiría estando presente hasta mediados de los años 50, momento en el que la pena desaparece durante 2 décadas hasta virtualmente el final del régimen. Este descenso acaecido en 1944 podría deberse a la finalización de la mayoría de causas contra individuos vinculados a la Segunda República, pero también al temor de una intervención aliada debido al alineamiento de la causa nacional con el Eje y su derrota, que en 1944 ya parece inevitable.

La edad de las víctimas

Las visualizaciones de datos complejos, como son las figuras mostradas en este trabajo, promueven nuevas miradas a los fenómenos históricos, que serían difícilmente generadas por los métodos convencionales de la investigación histórica. En este último gráfico, por ejemplo, se percibe un cierto descenso de la edad media de las víctimas ejecutadas por el franquismo; este patrón requiere un análisis más pormenorizado, así como una comparativa con la edad general de los encausados para poder comprobar si la dinámica es exclusiva de las penas capitales o, por el contrario, es general para todos los represaliados.

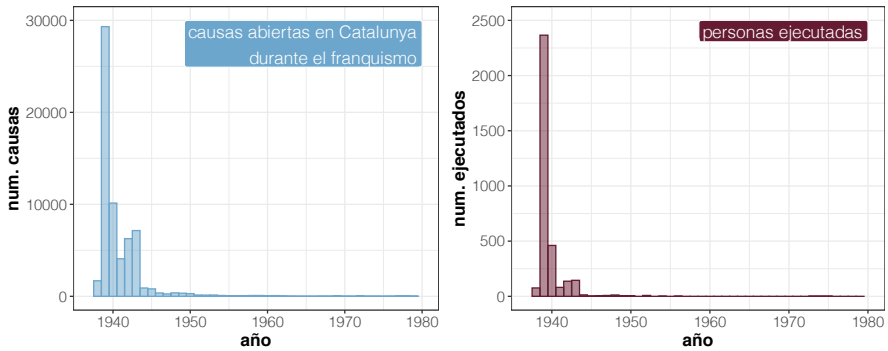


Figura 5: Histograma mostrando el total de causas (izquierda) y penas de muerte ejecutadas (derecha) durante el franquismo. El histograma es un gráfico de barras diseñado para explorar la distribución de valores de dos variables numéricas (en este caso año y número de causas).

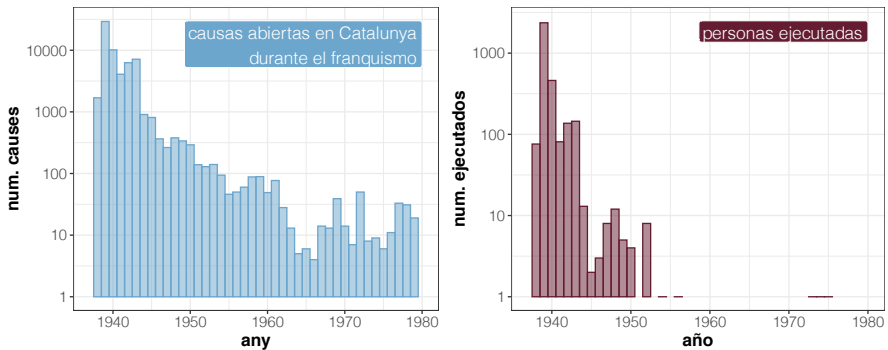


Figura 6: Histograma mostrando el total de causas (izquierda) y penas de muerte ejecutadas (derecha) durante el franquismo a partir de una escala logarítmica (eje Y). Los datos son los mismos que los de la Figura 5 pero en una escala logarítmica el intervalo entre órdenes de magnitud se mantiene constante (i.e. hay la misma distancia entre 1-10 casos que entre 10-100 y 100-1000).

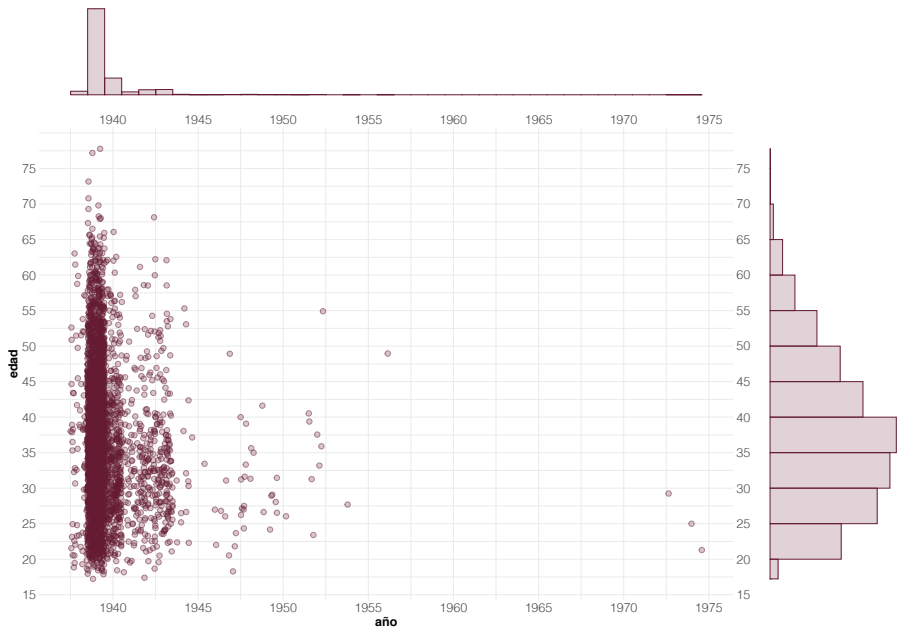


Figura 7: Gráfico de dispersión mostrando el número anual de personas ejecutadas por la dictadura franquista (eje X), así como su edad (eje Y). Cada observación tiene un 75% de transparencia, por lo que las zonas especialmente coloreadas, como el año 1939, hacen muy visible la magnitud de la represión. El gráfico de dispersión va acompañado por distribuciones marginales que, a modo de histogramas, explicitan la distribución de las víctimas por año de ejecución de la condena y edad.

La Figura 8 permite dicha comparativa como una serie temporal comparativa entre las edades medias de encausados (en general) y las personas que fueron finalmente ejecutadas.

Cabe destacar que el objetivo de este gráfico es la comparación de las edades medias y, por ese motivo, no muestra en ningún caso valores absolutos en el número de encausados (al contrario que los gráficos anteriores). Las visualizaciones deben responder a preguntas de investigación concretas y, en este caso, el uso de una estadística resumen como es la media aritmética, en lugar de trabajar con los datos en bruto, permite una mayor claridad en la exploración de la pregunta concreta referida a la edad de las víctimas.

La media de edad es diferente para los dos grupos y, aunque hay algunas excepciones, tiende a ser menor para los ejecutados que para el conjunto total de represalia-

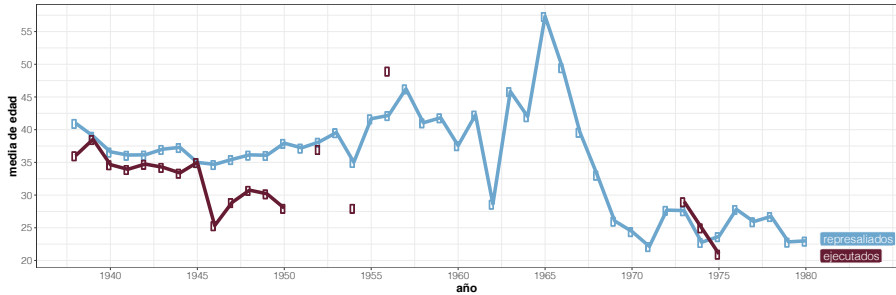


Figura 8: Comparativa de la edad media anual de encausados (azul) y ejecutados (rojo). La serie temporal no está completa para el segundo grupo, ya que no se detectan penas de muerte entre 1951 y 1972 con algunas excepciones

dos. Esta diferencia tiende a aumentar a partir de 1943 y a medida que la intensidad de la represión en Cataluña disminuye, si más no en cuanto al número de causas. A partir de este punto la edad de los encausados sube gradualmente, pasando de los 35 a mediados de los años 40 hasta los 55 en 1965. Esta dinámica sugiere que, hasta mediados de los 60, la represión se centró en el mismo grupo demográfico: personas que tenían entre 20 y 40 años durante la Guerra Civil y, por tanto, pudieron formar parte del bando republicano durante la guerra civil.

La última década de franquismo y la transición posterior ven un cambio radical de tendencia demográfica. En muy pocos años la edad media de las víctimas pasa de los 55 años hasta valores menores de 30 años. Vemos, así, que la represión desencadenada durante los últimos años del régimen se centraron en una nueva generación de combatientes antifranquistas que no vivieron la Guerra Civil, posiblemente a partir de los nuevos movimientos estudiantiles que afectaron a la mayoría de países europeos a finales de los años 60.

CONCLUSIONES

Como cualquier otra herramienta de investigación, la visualización de datos tiene limitaciones que deberían ser comprendidas por sus usuarios. En primer lugar, el objetivo principal de la visualización es identificar relaciones complejas entre las variables recogidas por los datos y, por ese motivo, su eficacia estará determinada por la cualidad de los mismos. Un conjunto de datos demasiado fragmentado o con mucha incertidumbre generaría una serie de desafíos que, pese a no inhabilitar la aproximación, requeri-

rían de una mayor complejidad de herramientas de análisis como son la simulación o la estadística bayesiana²⁴. En segundo lugar, las aproximaciones cuantitativas en investigación histórica son útiles para identificar patrones a gran escala y, por lo tanto, deberían ser complementadas por estudios de carácter más cualitativo con una mayor capacidad para caracterizar el detalle de los procesos históricos. Estos trabajos cualitativos permiten mejorar la interpretación de los análisis cuantitativos, ya que ayudan al investigador a comprender mejor las tendencias generales, así como destacar las excepciones de dichas tendencias. Este punto es especialmente destacable en el tema que nos ocupa, ya que no es posible comprender correctamente el fenómeno de la represión estudiando tan sólo la escala individual o la colectiva; ambas miradas son complementarias y necesarias.

Finalmente, en este análisis nos hemos limitado a analizar los datos mediante una herramienta exploratoria como es la visualización. La identificación de patrones mediante visualización es una fase vital de cualquier investigación cuantitativa, pero frecuentemente es necesario complementarla con el testeo de hipótesis que nos permita cuantificar hasta qué punto nuestras ideas se ajustan a la evidencia y responder, de manera general, a la pregunta clave: ¿cómo podemos saber si estamos equivocados? La visualización de datos es capaz de identificar patrones, pero a veces esos patrones no son estadísticamente significativos, es decir, no son tan claros en los datos como aparentan; el testeo estadístico es vital para distinguir las dinámicas observables en los datos de las que simplemente están en nuestra cabeza. En otras palabras, y como dijo el historiador de la ciencia Michael Shermer: «Los humanos somos animales que explican historias y buscan patrones; por ese motivo somos excelentes en explicar historias sobre patrones, aunque éstos en realidad no existan».

Este trabajo ha explorado las nuevas perspectivas que las Humanidades Digitales aportan al estudio de la Guerra Civil y la represión del régimen franquista. Es ciertamente un contexto de investigación distinto al de otros periodos históricos, y esa singularidad se percibe en el modo en el que se ha hecho accesible la información de modo digital. Desde esta perspectiva se ha priorizado el acceso a la información sobre individuos por

24. En este sentido es relevante ver las aproximaciones propuestas desde la disciplina arqueológica para integrar la incertidumbre temporal en CREMA, E. R. (2012). Modelling Temporal Uncertainty in Archaeological Analysis. *Journal of Archaeological Method and Theory*, 19(3), 440–461; YUBERO-GÓMEZ, M., RUBIO-CAMPILLO, X., & LÓPEZ-CACHERO, J. (2016). The study of spatio-temporal patterns integrating temporal uncertainty in late prehistoric settlements in northeastern Spain. *Archaeological and Anthropological Sciences*, 8(3), 477–490, la fragmentación de los datos en BEVAN, A., CONOLLY, J., HENNIG, C., JOHNSTON, A., QUERCIA, A., SPENCER, L., & VROOM, J. (2013). Measuring chronological uncertainty in intensive survey finds: A case study from Antikythera, Greece. *Archaeometry*, 55(2), 312–328 o la incertidumbre de las hipótesis en RUBIO-CAMPILLO, X., COTO-SARMIENTO, M., PÉREZ-GONZALEZ, J., & REMESAL RODRÍGUEZ, J. (2017). Bayesian analysis and free market trade within the Roman Empire. *Antiquity*, 91(359), 1241–1252.

parte de familiares y, por ese motivo, los portales disponibles no proporcionan la funcionalidad requerida a una aproximación abierta a los datos, acorde a los criterios adoptados por buena parte de la ciencia actual. No creemos que los dos objetivos sean incompatibles, sino al contrario; como es bien sabido, una mayor transparencia en la política de datos permite una comunicación más efectiva no sólo entre los investigadores.

AGRADECIMIENTOS

XRC es financiado por el programa Ramon y Cajal RYC2018-024050-I (Fondo Social Europeo – Agencia Estatal de Investigación). Los gráficos han sido elaborados a partir de QGIS 3.14 «Pi» y el framework estadístico R 3.6.3 (paquete ggplot2)²⁵. El código y los datos para las diferentes visualizaciones están disponibles bajo licencias abiertas en <https://github.com/xrubio/ebre38BDD>. Estamos muy agradecidos a Lourdes Prades, Concha Catalán y Francesc Xavier Hernández por sus reflexiones y comentarios alrededor de las temáticas tratadas en el presente trabajo.

25. QGIS Development Team, *QGIS Geographic Information System* (Open Source Geospatial Foundation, 2009) <<http://qgis.osgeo.org>>; R CORE TEAM, *R: A Language and Environment for Statistical Computing* (Vienna, Austria: R Foundation for Statistical Computing, 2020) <<https://www.R-project.org/>>; WICKHAM, HADLEY, *Ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016) <<https://ggplot2.tidyverse.org>>

