

Importancia de visualizar el mensaje extraído con minería y ciencia de datos. Storytelling con enfoque mixto para comunicación efectiva

Importance of visualizing the message extracted with data mining and science. Storytelling with a mixed approach for effective communication

Alfonso Vázquez-Atochero; Alberto Ledo-Díaz

Cómo citar este artículo:

Vázquez-Atochero, Alfonso; Ledo-Díaz, Alberto (2024). "La importancia de visualizar el mensaje extraído con minería y ciencia de datos [Storytelling con enfoque mixto para comunicación efectiva]". *Infonomy*, 2(3) e24036. <https://doi.org/10.3145/infonomy.24.036>



Alfonso Vázquez-Atochero

<https://orcid.org/0000-0002-1657-8275>

<https://directorioexit.info/ficha6958>

<https://www.alfonsovazquez.com>

Universidad de Extremadura

Facultad de Formación del Profesorado

Departamento de Ciencias de la Educación

Avda. de la Universidad, s/n. 10071 Cáceres, España

alfonso@unex.es



Alberto Ledo-Díaz

<https://directorioexit.info/ficha6989>

Universidad de Extremadura

Facultad de Ciencias de la Documentación y la Comunicación

Plazuela Ibn Marwan, s/n

06071 Badajoz, España

albertoledo@unex.es

Resumen

La visualización del mensaje extraído mediante minería y ciencia de datos es fundamental en la comunicación efectiva de hallazgos. La combinación de técnicas de storytelling ofrece un enfoque mixto poderoso para transmitir información compleja de manera accesible. Esta estrategia no solo resulta familiar, sino que también se adapta a diversas audiencias. Al reproducir gráficos y narrativas visuales, el objetivo no es simplemente presentar datos concretos, sino proporcionar ejemplos de representaciones visuales divergentes. Estas narrativas visuales no solo permiten una comprensión más profunda de los datos, sino que también facilitan su interpretación y difusión. Adaptadas para diferentes contextos y audiencias, estas técnicas se convierten en aliadas clave en la comunicación de resultados de investigación y análisis de datos. En resumen, la importancia de visualizar el mensaje extraído mediante minería y ciencia de datos radica en su capacidad para transmitir de manera efectiva información compleja y relevante a través de una variedad de medios visuales y narrativos.

Palabras clave

Minería de datos; Ciencia de datos; Narrativas visuales; Estudios de percepción; Comprensión holística; Individualidad; Colectividad; Glocalidad (global/local); Entrevista en Profundidad; Grupo de discusión; Método mixto; Cualitativo; Cuantitativo.

Abstract

The visualization of the message extracted through data mining and science is fundamental in the effective communication of findings. The combination of storytelling techniques offers a powerful mixed approach to convey complex information accessibly. This strategy is not only familiar but also adaptable to diverse audiences. By reproducing graphics and visual narratives, the goal is not simply to present concrete data, but to provide examples of divergent visual representations. These visual narratives not only allow for a deeper understanding of the data but also facilitate its interpretation and dissemination. Adapted for different contexts and audiences, these tools become key allies in communicating research results and data analysis. In summary, the importance of visualizing the message extracted through data mining and science lies in its ability to effectively convey complex and relevant information through a variety of visual and narrative means.

Keywords

Data mining; Data science; Visual narratives; Storytelling; Perceptual studies; Holistic understanding; Individuality; Collectivity; Glocality (global/local); In-depth interview; Focus group; Mixed method; Qualitative; Quantitative.

1. Aclaración de intenciones y precisiones epistemológicas

Es importante destacar que este artículo es esencialmente metodológico. Presentamos un proceso de trabajo que se ha puesto en práctica para analizar grandes volúmenes de información (cientos de cuestionarios y horas de grabaciones de audio). Sin embargo, no se incluyen resultados específicos de investigaciones, ya que el objetivo es compartir el enfoque interdisciplinar de trabajo

empleado y algunas formas de visualización de datos, no los resultados de estudios particulares. Por ello se han incorporado al final del texto diversas formas de expresión gráfica a modo de aclaración, donde lo relevante no son los datos en sí, sino la presentación visual de la información. Por lo tanto abarcamos el proceso metodológico y labor computacional y también la representación gráfica de hallazgos, abarcando por tanto las fases de metodología y resultados de la estructura **IMRD** de un artículo científico.

2. Comprensión holística del fenómeno social: data science y minería de datos

En un mundo incierto como el que estamos viviendo, dinámico y cambiante, medir la opinión pública se convierte en una tarea ardua y compleja. Los avances en tecnologías de la comunicación e información están reconfigurando sustancialmente muchos aspectos de la vida humana (**Sidorenko-Bautista, 2024**). Los métodos clásicos parecen no ser lo suficientemente explícitos cómo para entender la realidad y las potencialidades de la IA pueden ser grandes aliadas a la hora de reinterpretar y repensar la contemporaneidad (**Cervi, 2024**). No es que los preceptos de Gallup dejen de ser válidos o el dilema de Goldstein resulte caduco (**Vázquez-Atochero, 2012**). Pero tenemos que asumir que la certeza absoluta se convierte en una ficción inalcanzable (**Grima, 2011**) y debemos avanzar en la búsqueda de una verdad profunda que subyace en el ideario colectivo a pesar de los mensajes con los que bombardea la esfera mediática (**Sampedro, 2014**). Delaney y Kaspin (**2017**) integran esa esfera colectiva en el pensamiento individual, retando al individuo a reflexionar cómo los conceptos culturales moldean a la persona. Lévi-Strauss (**1958**), también desde un enfoque antropológico en el que la cercanía a la fuente primaria se vuelve prioritaria y prima la profundidad de la información frente a la amplitud de la muestra, afirma que

“la necesidad de proceder a un estudio detallado de caso entraña como consecuencia una restricción en el número de ejemplos que es posible considerar”.

Y continúa preguntándose si

“¿no se corre entonces el riesgo de limitarse a casos que son demasiado especiales para poder formular, sobre una base tan estrecha, conclusiones válidas para los restantes?”.

Es decir, ¿cantidad o calidad? Y permítasenos redundar en esta idea pues es sobre la que vamos pivotar en nuestro planteamiento metodológico, salvando el bagaje epistemológico acumulado y la amplitud de útiles aparecidos en los últimos tiempos y disponibles en la caja de herramientas del investigador social (**Carminotti; Topi, 2020**) de la segunda década del siglo XXI.

Pero además de ese acercamiento a la persona como elemento atómico de la realidad cultural, no podemos perder de vista que la individualidad se diluye en un complejo entramado social, si queremos dar respuestas al ser contemporáneo dentro de su mundo. Si bien el pasado siempre ha sido el mejor informante del presente, en estos momentos el escenario se complica con acciones elementales como la formación de identidades, la construcción de significados y la creación del sentido, dentro de un paradigma participativo en el que se construye de manera participativa (**Day, 2011**). No podemos comprender al individuo sin su entorno, pero tampoco a una cultura sin los que la conforman. Por lo tanto, junto al conocimiento

profundo de la persona, el pensamiento estadístico se vuelve cada vez más esencial para comprender nuestro complejo mundo y debe redirigirse a analizar datos y tomar decisiones informadas basadas en datos inciertos (Poldrack, 2023). La utilización de softwares específicos se convierte en una opción imprescindible para el científico social, que debe trabajar de manera interdisciplinar y no solo en su ámbito; así la Sociología, la Antropología y la Informática pueden establecer un triángulo que, lubricado desde el sentido común, dé respuestas a los interrogantes y cuestiones que se plantean las sociedades del siglo XXI. Precisamente ese es el tándem que articula esta metodología de prospección social. Tanto el software de data science y minería de datos (Vicente-Cestero; Mateos-Caballero, 2018; Ortega-Candel, 2022) como la propia interpretación y análisis de los mismos consisten un nuevo reto que nos ayuda a entender cómo funciona y se construye la opinión colectiva (Joyanes-Aguilar, 2020).

Para una comprensión holística de la sociedad, es esencial entender al individuo con sus peculiaridades y particularidades, describiendo así el complejo socio-cultural en el que vive. Estas características individuales tienden a repetirse, formando generalidades que, al ser entendidas en su estructura y relaciones, facilitan la comprensión de la sociedad. De esta manera, podemos comprender mejor cómo los protagonistas perciben y experimentan su entorno social.

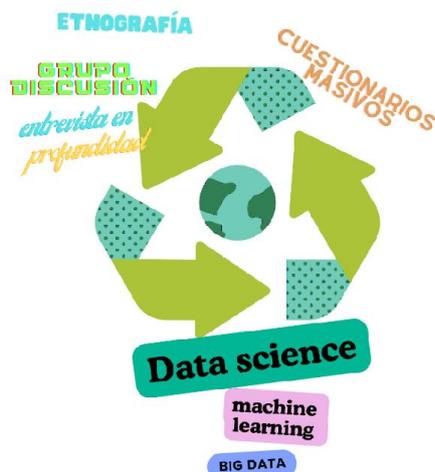


Imagen 1. Es necesario tener en cuenta los parámetros individuales y privados de cada individuo para poder entenderlo en su entorno. También es necesaria la percepción social de un fenómeno para comprender mejor la relación entre lo global y lo personal.

Es fundamental comprender las individualidades para entender el funcionamiento de un grupo y viceversa. Entre la individualidad y la grupalidad existe una constante dinámica de flujos, un complejo juego de poderes y contrapoderes que moldea cada elemento -nodo- del grupo -sistema-. El individuo aporta diversidad al grupo, a medida que éste adquiere forma y abre puertas a las peculiaridades individuales. Internet agrega una nueva dimensión, introduciendo una complejidad multicapa al proceso. La red actúa como un mecanismo metabólico dentro del organismo sociocultural sujeto de estudio. Por lo tanto, es esencial profundizar en el análisis del discurso y las narrativas individuales, mientras se investiga la estructura vital del grupo. Para lograrlo, necesitamos instrumentos que midan aspectos cuantitativos para establecer un marco de referencia, y estrategias cualitativas para comprender cómo el individuo se nutre de la esfera social, y viceversa, en este constante intercambio de influencias y significados.

3. Importancia de un análisis mixto

La importancia de combinar fuentes primarias y estadísticas en ciencias sociales radica en la necesidad de obtener una comprensión más completa y precisa de los fenómenos sociales (De-Waal *et al.*, 2017). Las fuentes primarias, como entrevistas, encuestas, observaciones directas y documentos históricos, proporcionan datos detallados y contextuales sobre experiencias individuales y eventos específicos. Estas fuentes son valiosas para capturar la riqueza y la diversidad de las experiencias humanas, así como para explorar procesos sociales complejos desde una perspectiva cualitativa (Daines *et al.*, 2017). Por otro lado, las estadísticas ofrecen una visión más amplia y generalizada de los patrones y tendencias en la sociedad. Al analizar datos numéricos recopilados a través de encuestas de población, censos, estudios longitudinales y otros instrumentos cuantitativos, los investigadores pueden identificar correlaciones, realizar comparaciones y realizar inferencias sobre poblaciones más grandes. Las estadísticas son fundamentales para detectar patrones, hacer predicciones y evaluar la magnitud y la dirección de las relaciones entre variables en un estudio (Maravelakis, 2019).

Combinando fuentes primarias y estadísticas se enriquece el análisis y se obtiene una comprensión más holística de los fenómenos sociales: Las fuentes primarias proporcionan detalles contextuales y perspectivas individuales que pueden ayudar a contextualizar. Las estadísticas pueden respaldar y generalizar los hallazgos derivados de fuentes primarias, brindando una visión más amplia y representativa.

Al combinar fuentes primarias y estadísticas, los investigadores pueden enriquecer su análisis y obtener una comprensión más holística de los fenómenos sociales. Las fuentes primarias proporcionan detalles contextuales y perspectivas individuales que pueden ayudar a contextualizar y enriquecer los hallazgos estadísticos (Aguirre-Baztán, 1995). Por otro lado, las estadísticas pueden respaldar y generalizar los hallazgos derivados de fuentes primarias, brindando una visión más amplia y representativa de los fenómenos sociales estudiados (Sagaró-del-Campo, 2019). Por lo tanto, la combinación de fuentes primarias y estadísticas en ciencias sociales permite a los investigadores abordar preguntas complejas desde múltiples perspectivas, enriqueciendo así la comprensión y el análisis de los fenómenos sociales. Esta integración de enfoques cualitativos y cuantitativos es fundamental para avanzar en el conocimiento en ciencias sociales y para abordar los desafíos y complejidades de las sociedades contemporáneas. De esta colaboración metodológica entre la investigación cuantitativa y la cualitativa surge el enfoque o método mixto.

La metodología mixta, también conocida como enfoque mixto (Flick, 2007), se refiere a la combinación y utilización de métodos de investigación tanto cualitativos como cuantitativos en un mismo estudio o investigación. Es un enfoque que busca

aprovechar las fortalezas de ambos tipos de métodos para obtener una comprensión más completa y profunda de un fenómeno o problema de investigación. En la metodología mixta, los investigadores pueden recopilar y analizar datos tanto cualitativos como cuantitativos de manera simultánea o secuencial, y luego integrarlos para obtener una visión más holística del fenómeno estudiado. Esta integración puede realizarse en diferentes etapas del proceso de investigación, como la recolección de datos, el análisis o la interpretación de los resultados ([Hernández-Sampieri, 2018](#)).

Los métodos cualitativos se centran en comprender y explorar las experiencias, percepciones, significados y contextos sociales de las personas mediante técnicas como entrevistas, observaciones participantes o análisis de contenido. Estos métodos proporcionan una comprensión profunda y detallada de los fenómenos estudiados, permitiendo capturar la complejidad y riqueza de las interacciones humanas ([Ñáupas-Paitán et al., 2014](#)).

Por otro lado, los métodos cuantitativos se basan en la recopilación y análisis de datos numéricos y estadísticos para identificar patrones, relaciones y regularidades en un fenómeno. Estos métodos utilizan técnicas como encuestas, experimentos, análisis estadístico u otros procedimientos cuantitativos para medir variables, realizar generalizaciones y establecer relaciones causales ([Ramírez-Mercado, 2023](#)).

La metodología mixta permite combinar las fortalezas de ambos enfoques. Por ejemplo, los métodos cualitativos pueden ayudar a comprender los contextos y las perspectivas de los participantes, mientras que los métodos cuantitativos pueden proporcionar una visión generalizada de los fenómenos a través de un muestreo representativo y análisis estadísticos rigurosos. La metodología mixta es un enfoque de investigación que combina métodos cualitativos y cuantitativos para obtener una comprensión más completa, rica y precisa de los fenómenos sociales estudiados. Permite abordar preguntas de investigación complejas, obtener resultados validados y contextualizados, y proporciona flexibilidad en el diseño y la implementación de estudios. La metodología mixta en ciencias sociales es importante por varias razones:

Complementariedad de enfoques: la metodología mixta combina tanto métodos cualitativos como cuantitativos, lo que permite abordar de manera más completa y holística los fenómenos sociales. Los métodos cualitativos proporcionan una comprensión profunda y detallada de los contextos, significados y experiencias humanas, mientras que los métodos cuantitativos permiten analizar y generalizar los resultados en un nivel más amplio.

Validación de resultados: al utilizar diferentes métodos de recolección y análisis de datos, se pueden obtener resultados convergentes y proporcionar mayor validez a las conclusiones. La triangulación de datos provenientes de diferentes fuentes y perspectivas fortalece la confiabilidad de los hallazgos y reduce los sesgos inherentes a un solo método.

Abordaje de preguntas de investigación complejas: algunas preguntas de investigación en ciencias sociales requieren una comprensión profunda de los procesos sociales, así como la identificación de patrones y tendencias a nivel poblacional. La metodología mixta permite explorar tanto los aspectos cualitativos como los cuantitativos de un fenómeno, lo que facilita el análisis de relaciones complejas y la obtención de una imagen más completa y precisa.

Contextualización de resultados: los métodos cualitativos proporcionan información rica sobre los contextos específicos en los que se desarrollan los fenómenos sociales. Esta contextualización es fundamental para comprender las dinámicas sociales, las interacciones y los significados atribuidos por los actores involucrados. Los métodos cuantitativos, por su parte, ayudan a generalizar los resultados a una escala más amplia, lo que permite identificar patrones más generales y evaluar el impacto de políticas o intervenciones a nivel poblacional.

Flexibilidad y adaptabilidad: la metodología mixta brinda flexibilidad para adaptarse a diferentes situaciones de investigación. Dependiendo de la naturaleza de la pregunta de investigación, los investigadores pueden elegir y combinar diferentes métodos de acuerdo con las necesidades del estudio y los recursos disponibles.

La metodología mixta en ciencias sociales es básica porque permite una comprensión más profunda y completa de los fenómenos sociales, facilita la validación de los resultados, aborda preguntas de investigación complejas, contextualiza los hallazgos y proporciona flexibilidad en el diseño y la implementación de estudios

La metodología mixta en ciencias sociales es básica porque permite una comprensión más profunda y completa de los fenómenos sociales, facilita la validación de los resultados, aborda preguntas de investigación complejas, contextualiza los hallazgos y proporciona flexibilidad en el diseño y la implementación de estudios ([Sarramona, 2023](#)).

4. Herramientas para el análisis de datos

En el ámbito del análisis de datos de encuestas, se ha producido un avance significativo gracias a la incorporación de nuevas tecnologías que permiten aprovechar al máximo los conjuntos de datos recopilados. Algunas de las tecnologías más recientes utilizadas en el análisis de datos incluyen *PSPP*, *Atlas.ti* y *Python* con *Quantipy*, *NumPy* y *Pandas*. Además, para la visualización de datos se destaca *Chart.js* y *Sigma.js*. Estas tecnologías combinadas han transformado el proceso de análisis de datos de encuestas, proporcionando una mayor precisión y eficiencia en la interpretación de los resultados. En cuanto al análisis de datos.

PSPP es un software libre que ofrece capacidades avanzadas de análisis estadístico (Sto.-Tomas *et al.*, 2019). Con PSPP, es posible realizar una variedad de análisis descriptivos e inferenciales, incluyendo análisis de frecuencia, correlación, regresión y pruebas de significancia. Este programa nos ha permitido explorar los datos de la encuesta de manera exhaustiva y obtener información valiosa sobre las variables estudiadas (Shepherd, 2024). Poseemos otras opciones también de código abierto, como Jamovi (Şahin, 2019).

Por un lado encontramos la familia de los programas de Análisis Cualitativo Asistido por Computador, o CAQDAS (Qualitative Data Analysis Software), capaces de codificar datos, facilitando que la investigación y estudios ganen en rigor y eficacia. Nos ayudarán para trabajar con estudios de casos, entrevistas en profundidad o historias de vida, análisis narrativo y del discurso. Atlas.ti es un software de análisis cualitativo que facilita el procesamiento y análisis de datos no estructurados en las encuestas, como respuestas abiertas o entrevistas (Paulus, 2016). Con Atlas.ti, los investigadores pueden organizar y codificar los datos cualitativos, identificar patrones temáticos y realizar análisis detallados. Es especialmente útil cuando se busca comprender las percepciones y opiniones de los participantes en la encuesta de manera más profunda. En la misma línea se puede usar NVivo, software que facilita la organización, visualización y acceso a los datos de la investigación, lo que representa una economía de tiempo y trabajo (Dias-do-Nascimento *et al.*, 2016) o Maxqda (Consoli, 2021). Estos softwares propietarios cuentan con alternativas de código abierto, como Ligre o Aquad, que ofrecen servicios similares.



Imagen 2. Pantalla de trabajo en Ligre.

Además del habitual software de escritorio, varias de las aplicaciones descritas en párrafos anteriores están disponibles de manera híbrida, con versiones online. Y otras, puramente online. El trabajo online también se abre paso en tanto en cuanto permite mantener los resultados a salvo y facilita el trabajo cooperativo (Kunicki, 2019).

Por otra parte podemos recurrir a Python, junto con las bibliotecas Quantipy, NumPy y Pandas, nos han proporcionado una combinación poderosa para el análisis de datos de encuestas. Python es un lenguaje de programación versátil y popular que permite el procesamiento y análisis de datos de manera eficiente (Brown, 2023). Quantipy es una metodología basada en Python que ofrece un enfoque estructurado

y coherente para el análisis de datos complejos de encuestas. Con *Quantipy*, hemos gestionado los datos, aplicando ponderaciones y filtros, y generando tablas y gráficos estadísticos de manera automatizada. Por su parte, *NumPy* y *Pandas* son bibliotecas de *Python* que nos han brindado una amplia gama de funciones y utilidades para el procesamiento y análisis de datos numéricos y estructurados. Estas bibliotecas permiten realizar operaciones matemáticas y estadísticas, realizar limpieza y transformación de datos, y explorar relaciones y patrones en los datos de la encuesta ([Matthes, 2023](#)).

En cuanto a la visualización de datos, *Chart.js* y *Sigma.js* son dos librerías populares y versátiles. *Chart.js* es una biblioteca de JavaScript que permite crear visualizaciones interactivas y atractivas, como gráficos de barras, líneas, tortas y más. Además facilita la representación clara y efectiva de los resultados de la encuesta, permitiendo una comprensión visual rápida de los datos. Por su parte, *Sigma.js* es una biblioteca de visualización en red que se utiliza para representar y explorar redes complejas de datos. Con *Sigma.js*, es posible crear visualizaciones interactivas y dinámicas de las relaciones entre elementos de la encuesta.

Tanto para el análisis cuantitativo como el cualitativo este estudio se ha apoyado en el uso de diagramas n-gram. Los diagramas n-gram se utilizan en el análisis de texto y en el procesamiento del lenguaje natural. Si bien su uso principal está asociado con la metodología cuantitativa, también pueden tener cierta utilidad en la metodología cualitativa ([Guerrero-Carmona, 2014](#)). Los diagramas n-gram pueden ser útiles en un enfoque cualitativo en aspectos como:

Identificación de patrones de lenguaje: los diagramas n-gram permiten visualizar la frecuencia y co-ocurrencia de palabras o secuencias de palabras en un texto. Esto puede ayudar a identificar patrones de lenguaje, frases comunes o términos recurrentes en un conjunto de datos cualitativos. Al observar los diagramas n-gram, los investigadores pueden identificar las palabras clave y las asociaciones entre ellas, lo que puede proporcionar información sobre las ideas y temas que surgen en el texto.

Análisis de conceptos emergentes: los diagramas n-gram pueden ayudar a identificar conceptos emergentes en un conjunto de datos cualitativos. Al analizar los n-gramas más frecuentes, los investigadores pueden identificar nuevas combinaciones de palabras o términos que pueden indicar la aparición de ideas o conceptos relevantes. Estos conceptos emergentes pueden ser explorados en mayor profundidad en el análisis cualitativo, permitiendo una comprensión más rica del fenómeno estudiado.

Validación y apoyo en la interpretación: los diagramas n-gram pueden proporcionar evidencia adicional para respaldar las interpretaciones cualitativas. Al visualizar la frecuencia y distribución de las palabras en un conjunto de datos, los investigadores pueden verificar si las interpretaciones iniciales son consistentes con los patrones observados en los diagramas n-gram. Esto puede ayudar a fortalecer los hallazgos y aumentar la confianza en las conclusiones cualitativas.

Exploración de relaciones entre palabras: los diagramas n-gram pueden revelar relaciones entre palabras o términos en un texto cualitativo. Al observar las co-ocurrencias de palabras en los diagramas n-gram, los investigadores pueden identificar asociaciones significativas o conexiones conceptuales entre diferentes elementos del texto. Esto puede facilitar la identificación de temas, subtemas o relaciones implícitas que pueden ser objeto de análisis cualitativo adicional.

Es importante tener en cuenta que, en la metodología cualitativa, los diagramas n-gram no deben considerarse como un elemento independiente, sino como un complemento para el análisis cualitativo tradicional. Los diagramas n-gram pueden proporcionar información útil y ayudar a generar hipótesis o enfoques exploratorios, pero la interpretación y comprensión en última instancia se basan en la experiencia y el conocimiento del investigador en el campo de estudio.

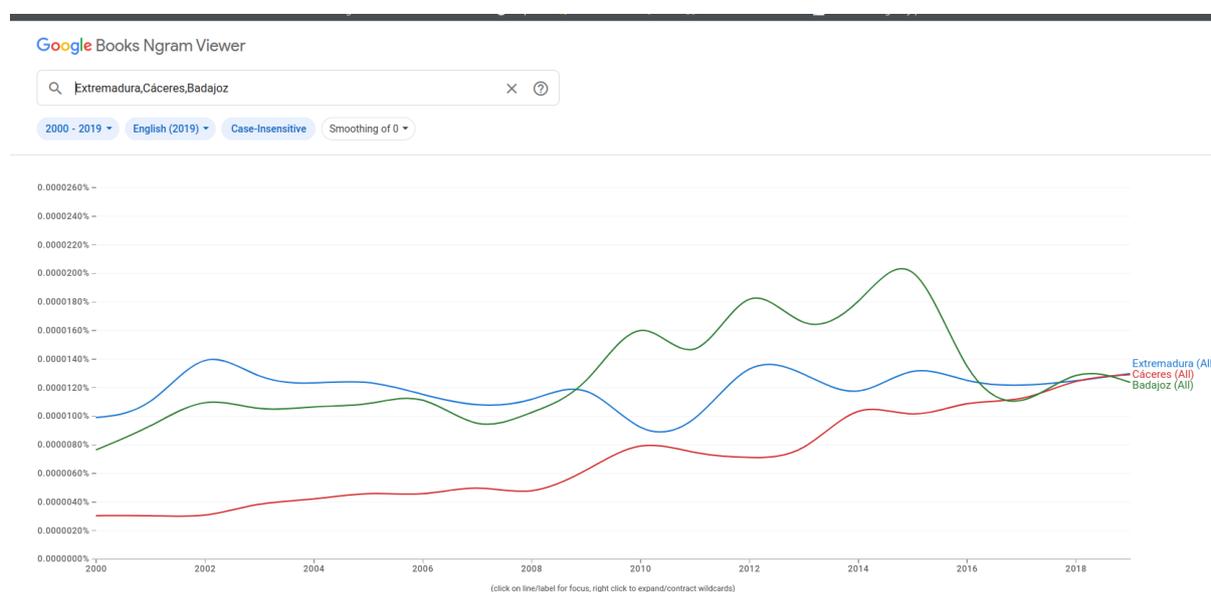


Imagen 3. Pantalla de trabajo en *Ngram viewer*.

Los n-gramas son secuencias de n elementos consecutivos, como palabras o caracteres, que se extraen de un texto. Los patrones n-gram permiten identificar las frecuencias y las relaciones entre los elementos en una encuesta, lo que puede proporcionar conocimientos valiosos sobre las respuestas y las tendencias en los datos recopilados. Veamos algunas formas en las que se pueden utilizar los patrones n-gram en el análisis cuantitativo de encuestas:

Análisis de frecuencia: al aplicar los patrones n-gram a una encuesta, se pueden identificar las frecuencias de las palabras o frases más comunes. Esto ayuda a comprender las respuestas y los temas dominantes en los datos recopilados. Por ejemplo, al extraer los trigramas más frecuentes, se pueden identificar combinaciones de tres palabras que se repiten con mayor frecuencia en las respuestas de la encuesta.

Detección de temas: los patrones n-gram se pueden utilizar para identificar temas recurrentes en las respuestas de la encuesta. Al examinar los n-gramas más frecuentes, es posible agrupar las respuestas en categorías temáticas y comprender los principales problemas, preocupaciones o ideas expresadas por los encuestados.

Análisis de sentimiento: mediante el uso de patrones n-gram, se puede realizar un análisis de sentimiento en las respuestas de la encuesta. Al extraer n-gramas específicos, se pueden identificar combinaciones de palabras que reflejan sentimientos positivos, negativos o neutrales en las respuestas de los encuestados. Esto ayuda a comprender la actitud general de los participantes hacia un tema o pregunta específica.

Detección de errores o patrones inusuales: al analizar los patrones n-gram en las respuestas de la encuesta, se pueden identificar errores o patrones inusuales en los datos. Por ejemplo, se pueden detectar respuestas repetitivas, respuestas que contienen palabras mal escritas o patrones de palabras que no siguen la estructura esperada. Esto ayuda a identificar posibles problemas en la calidad de los datos recopilados.

Análisis comparativo: los patrones n-gram también se pueden utilizar para realizar comparaciones entre diferentes grupos o segmentos de encuestados. Al aplicar los patrones n-gram a subconjuntos de respuestas, como respuestas de diferentes regiones geográficas o grupos demográficos, es posible identificar diferencias en los temas, las preocupaciones o los patrones lingüísticos utilizados por cada grupo.

Los patrones n-gram son un instrumento eficaz en el análisis cuantitativo de encuestas. Permiten identificar patrones de palabras y frases, extraer información relevante, detectar temas recurrentes, analizar sentimientos y realizar comparaciones entre grupos. Al utilizar los patrones n-gram, los investigadores pueden obtener conocimientos más profundos y significativos a partir de los datos de la encuesta, lo que facilita la toma de decisiones informadas y la comprensión de las opiniones y actitudes de los encuestados.

5. Visualización de datos: ejemplos y casos prácticos

Los autores de este artículo hemos utilizado técnicas de data mining y ciencia de datos para generar storytelling para investigaciones propias y para terceros (administraciones, empresas, fundaciones, grupos de investigación...). Es posible que lo que aquí se cuenta y, sobre todo, la forma de contarlo, resulte familiar. Al reproducir estos gráficos no pretendemos transmitir la información concreta que cada uno de ellos contiene, sino mostrar ejemplos de narrativas visuales divergentes alejadas de la imagería manejada de manera más convencional en informes y papers. Estas narrativas visuales permiten una comprensión más profunda y atractiva de los datos, facilitando su interpretación y difusión entre audiencias diversas. Además, estas representaciones visuales pueden ser adaptadas para diferentes contextos y públicos, lo que las convierte en opción potencial para la comunicación efectiva de resultados de investigación y análisis de datos.

¿Estás de acuerdo o en desacuerdo con la afirmación "Extremadura no tiene relato nacional e internacional y hay que crear uno"?										
	18-25	26-35	36-45	46-55	56-65	66-80				
Totalmente de acuerdo	75	10,22 %	23	3,13 %	36	4,90 %	36	4,90 %	3	0,41 %
De acuerdo	117	15,94 %	31	4,22 %	53	7,22 %	70	9,54 %	7	0,95 %
Ni de acuerdo ni en desacuerdo	53	7,22 %	30	4,09 %	53	7,22 %	68	9,26 %	3	0,41 %
En desacuerdo	9	1,23 %	5	0,68 %	9	1,23 %	27	3,68 %	0	0,00 %
Totalmente en desacuerdo	4	0,54 %	6	0,82 %	5	0,68 %	11	1,50 %	0	0,00 %
Total:	258	35,15 %	95	12,94 %	156	21,25 %	212	28,88 %	13	1,77 %

¿Estás de acuerdo o en desacuerdo con la afirmación "El imaginario colectivo proyecta una valoración negativa de Extremadura"?										
	18-25	26-35	36-45	46-55	56-65	66-80				
Totalmente de acuerdo	132	17,98 %	42	5,72 %	63	8,59 %	90	12,26 %	4	0,54 %
De acuerdo	103	14,03 %	42	5,72 %	58	7,90 %	89	12,13 %	7	0,95 %
Ni de acuerdo ni en desacuerdo	19	2,59 %	9	1,23 %	9	1,23 %	22	3,00 %	2	0,27 %
En desacuerdo	3	0,41 %	1	0,14 %	4	0,54 %	6	0,82 %	0	0,00 %
Totalmente en desacuerdo	1	0,14 %	1	0,14 %	2	0,27 %	5	0,68 %	0	0,00 %
Total:	258	35,15 %	95	12,94 %	156	21,25 %	212	28,88 %	13	1,77 %

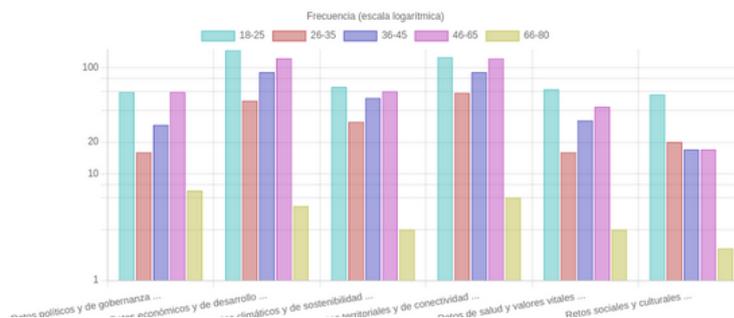
¿Estás de acuerdo o en desacuerdo con la afirmación "Es necesario crear una visión renovada de Extremadura"?										
	18-25	26-35	36-45	46-55	56-65	66-80				
Totalmente de acuerdo	143	19,48 %	40	5,45 %	80	10,90 %	93	12,67 %	4	0,54 %
De acuerdo	87	11,85 %	34	4,63 %	58	7,90 %	85	11,58 %	8	1,09 %
Ni de acuerdo ni en desacuerdo	22	3,00 %	16	2,18 %	12	1,63 %	21	2,86 %	1	0,14 %
En desacuerdo	6	0,82 %	3	0,41 %	4	0,54 %	8	1,09 %	0	0,00 %
Totalmente en desacuerdo	0	0,00 %	2	0,27 %	2	0,27 %	5	0,68 %	0	0,00 %
Total:	258	35,15 %	95	12,94 %	156	21,25 %	212	28,88 %	13	1,77 %

¿Estás de acuerdo o en desacuerdo con la afirmación "Extremadura necesita comunicar su identidad para crear un proyecto común que integre el compromiso...?"										
	18-25	26-35	36-45	46-55	56-65	66-80				
Totalmente de acuerdo	141	19,21 %	42	5,72 %	68	9,26 %	77	10,49 %	4	0,54 %
De acuerdo	92	12,53 %	33	4,50 %	65	8,59 %	72	9,81 %	8	1,09 %
Ni de acuerdo ni en desacuerdo	21	2,86 %	14	1,91 %	16	2,18 %	41	5,59 %	1	0,14 %
En desacuerdo	2	0,27 %	4	0,54 %	4	0,54 %	18	2,45 %	0	0,00 %
Totalmente en desacuerdo	2	0,27 %	2	0,27 %	2	0,27 %	4	0,54 %	0	0,00 %
Total:	258	35,15 %	95	12,94 %	156	21,25 %	212	28,88 %	13	1,77 %

¿Estás de acuerdo o en desacuerdo con la afirmación "Extremadura debe impulsar la imagen de la región"?										
	18-25	26-35	36-45	46-55	56-65	66-80				
Totalmente de acuerdo	137	21,39 %	54	7,36 %	85	11,58 %	106	14,44 %	5	0,68 %
De acuerdo	81	11,04 %	32	4,36 %	60	8,17 %	80	10,90 %	6	0,82 %
Ni de acuerdo ni en desacuerdo	14	1,91 %	8	1,09 %	9	1,23 %	19	2,59 %	1	0,14 %
En desacuerdo	4	0,54 %	0	0,00 %	2	0,27 %	3	0,41 %	1	0,14 %
Totalmente en desacuerdo	2	0,27 %	1	0,14 %	0	0,00 %	4	0,54 %	0	0,00 %
Total:	258	35,15 %	95	12,94 %	156	21,25 %	212	28,88 %	13	1,77 %

En el ámbito de las ciencias sociales, el uso de tablas llenas de datos ofrece una visión exhaustiva de la información. Sin embargo, en ciertos contextos, esta abundancia de datos puede resultar abrumadora y dificultar la comprensión. En estas situaciones, las narrativas visuales emergen como una metodología esencial. Una tabla densa de datos sobre ingresos

familiares, por ejemplo, puede transformarse en un diagrama de barras que ilustra claramente las disparidades económicas. Esta transición a lo visual simplifica la información, permitiendo a los lectores captar rápidamente las tendencias y patrones. Además, las narrativas visuales pueden ser más accesibles para audiencias con diferentes niveles de familiaridad con los datos. Así, la narrativa visual no reemplaza la tabla de datos, sino que la complementa, ofreciendo una forma más clara y directa de comunicar información compleja. Esta evolución hacia lo visual en las ciencias sociales refleja la necesidad de hacer que los datos sean más digeribles y significativos para un público amplio.



Las nubes de palabras han demostrado ser una opción valiosa para visualizar los términos más representativos de una investigación. Esta técnica transforma un conjunto de palabras en un conglomerado léxico, donde el tamaño de cada palabra indica su frecuencia de aparición.

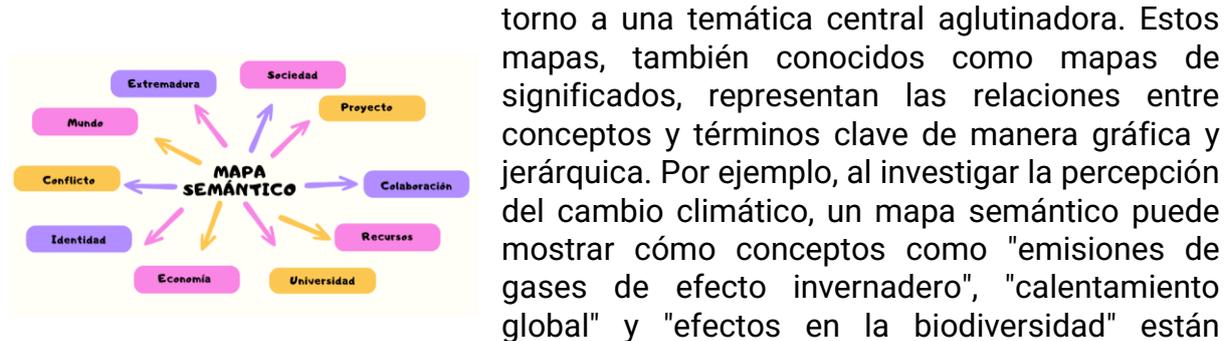


Sin embargo, para una comprensión más rápida y efectiva, podemos ir más allá y darle a esta nube de palabras una forma asociativa que conecte rápidamente con el tema principal. Por ejemplo, al presentar los términos de una investigación sobre consumo y moda en una nube de palabras con forma de bolsa, el



lector se sumerge visualmente en el mundo del consumo y la moda. Esta forma de presentación no solo facilita la identificación de los conceptos clave, sino que también evoca de manera subconsciente la temática central de la investigación. Así, la combinación de la nube de palabras y la forma asociativa potencia la comprensión visual de la investigación, permitiendo al lector captar de un vistazo los temas y conceptos más relevantes. Este enfoque ilustra cómo las narrativas visuales pueden elevar la presentación de datos en las ciencias sociales, ofreciendo una representación visualmente atractiva y significativa de los hallazgos.

Los mapas semánticos nos permiten organizar y visualizar un mapa conceptual en



torno a una temática central aglutinadora. Estos mapas, también conocidos como mapas de significados, representan las relaciones entre conceptos y términos clave de manera gráfica y jerárquica. Por ejemplo, al investigar la percepción del cambio climático, un mapa semántico puede mostrar cómo conceptos como "emisiones de gases de efecto invernadero", "calentamiento global" y "efectos en la biodiversidad" están

interconectados y relacionados. Además, la posición y la proximidad de los términos en el mapa reflejan su nivel de relevancia y su relación con el tema central. Esto permite una comprensión más clara y estructurada de los conceptos clave de la investigación. Los mapas semánticos no sólo ayudan a visualizar la complejidad de las relaciones conceptuales, sino que también sirven como guía para explorar de manera más profunda las interconexiones entre los diferentes aspectos de un tema. Los mapas semánticos son una herramienta esencial en las ciencias sociales para organizar y representar visualmente el entramado conceptual de una investigación, facilitando así la comprensión y el análisis de los datos.



Brand



Influencer



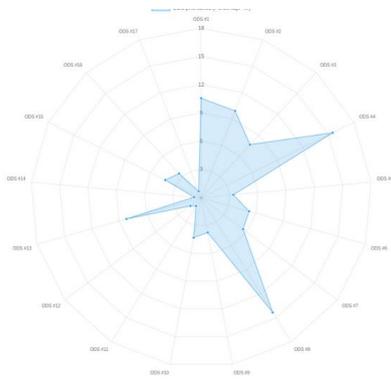
La presentación de palabras clave mediante códigos de colores y tamaños, junto con iconografía asociada, ofrece una manera efectiva de establecer asociaciones entre términos fundamentales. Esta técnica visual permite destacar los conceptos clave y sus relaciones de manera clara y concisa. Por ejemplo, en un estudio sobre la desigualdad de género, los términos "igualdad de oportunidades", "brecha salarial", y "roles de género" pueden presentarse en una visualización con colores distintivos y

tamaños proporcionales a su importancia. La inclusión de iconos relacionados, como una balanza para la igualdad de oportunidades o una figura masculina y femenina para los roles de género, refuerza aún más la asociación de los términos con sus conceptos correspondientes. Es importante evitar sobrecargar la

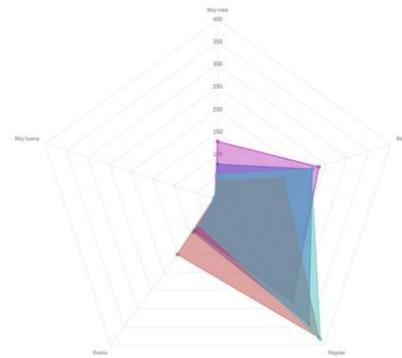


visualización con demasiada información para no perder la claridad y globalidad del mensaje. Esta metodología no solo simplifica la comprensión de los términos clave, sino que también ofrece una representación visualmente atractiva y significativa de los temas abordados en la investigación. Así, la combinación de palabras clave, códigos de colores e iconografía en la visualización de datos en ciencias sociales facilita la comunicación efectiva de conceptos complejos y promueve una comprensión más profunda de los temas estudiados.

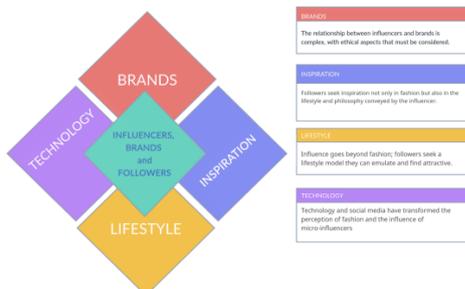
El diagrama circular ofrece una forma visual de identificar la preponderancia de un concepto frente a otros de la misma escala. Por ejemplo, al analizar los Objetivos de Desarrollo Sostenible (ODS) más valorados en un sistema cultural, se pueden colocar los 17 elementos en una circunferencia. El relleno de cada sección del círculo permite identificar



claramente la popularidad de cada uno de los ODS. En el primer caso (elemento circular), el diagrama compara todos los elementos de manera individual, mostrando la proporción de atención que cada ODS recibe. En el segundo caso (elemento pentagonal), el diagrama permite mayor interacción y la superposición de capas de valores. Esto significa que, además de mostrar la popularidad de cada ODS, también se pueden agregar capas adicionales para representar diferentes perspectivas o criterios de evaluación. Por ejemplo, una capa podría representar la percepción de los expertos, otra capa las políticas gubernamentales y otra capa la opinión pública. Esta visualización dinámica y multifacética no solo simplifica la identificación de los ODS más valorados, sino que también proporciona una visión más completa y matizada de cómo se perciben y valoran estos objetivos en un contexto cultural determinado.



claramente la popularidad de cada uno de los ODS. En el primer caso (elemento circular), el diagrama compara todos los elementos de manera individual, mostrando la proporción de atención que cada ODS recibe. En el segundo caso (elemento pentagonal), el diagrama permite mayor interacción y la superposición de capas de valores. Esto significa que, además de mostrar la popularidad de cada ODS, también se pueden agregar capas adicionales para representar diferentes perspectivas o criterios de evaluación. Por ejemplo, una capa podría representar la percepción de los expertos, otra capa las políticas gubernamentales y otra capa la opinión pública. Esta visualización dinámica y multifacética no solo simplifica la identificación de los ODS más valorados, sino que también proporciona una visión más completa y matizada de cómo se perciben y valoran estos objetivos en un contexto cultural determinado.



Organizar los contenidos de forma gráfica y jerárquica, utilizando colores y distribución del espacio, junto con una breve explicación, es una manera efectiva de presentar información de manera sintética y clara. Esta técnica permite crear una

representación visual que resalta la importancia relativa de cada elemento y las relaciones entre ellos. Al jugar con los colores y la distribución del espacio, se pueden enfatizar los conceptos clave y crear una estructura visualmente atractiva y fácil de entender. Además, al agregar una explicación breve a cada elemento, se proporciona contexto y claridad adicional. Este enfoque de presentación gráfica permite a los lectores obtener una visión general rápida y comprensible de los contenidos, ofreciendo una conclusión visual que resume los aspectos más importantes de manera efectiva.



La pirámide es un elemento gráfico efectivo para presentar resultados jerarquizados con una estructura conceptual clara y concisa. De manera intuitiva, el lector interpreta el orden de los conceptos según su posición en la pirámide. Los conceptos principales se colocan en la parte superior, seguidos de los subconceptos en capas inferiores. Esta representación visual permite una comprensión rápida y visualmente clara de la jerarquía de ideas, facilitando la asimilación

de la información. Además, el formato de pirámide enfatiza la importancia relativa de cada concepto en relación con los demás, proporcionando un enfoque visualmente impactante y memorable para la presentación de resultados.



Utilizar el tamaño de una iconografía conocida es una estrategia efectiva para interpretar rangos y escalas de manera intuitiva. En el primer caso, hemos aplicado esta técnica para representar la popularidad de los Objetivos de Desarrollo Sostenible (ODS) entre la población entrevistada. El icono de mayor tamaño representa el ODS considerado más importante, seguido por los iconos de tamaño descendente que representan los ODS

ubicados entre la segunda y sexta posición en popularidad. Esta representación visual permite una comprensión rápida y clara de la jerarquía de importancia de los ODS.

En el segundo caso, aplicamos esta técnica con un enfoque de género. Los iconos de mayor tamaño representan los temas o conceptos más relevantes para un género específico, mientras que los iconos de menor tamaño representan temas menos relevantes para ese género. Esta representación gráfica proporciona una visión visualmente intuitiva de las diferencias y similitudes en las



percepciones o preferencias de género en relación con un tema específico. En ambos casos, el uso del tamaño de la iconografía conocida facilita la interpretación y la comprensión de los rangos y escalas de manera rápida y efectiva.



Para representar el recorrido de un proyecto, como la aplicación de la agenda 2030, podemos recurrir a la metáfora de un trayecto en carretera. En esta analogía, cada meta u objetivo de la agenda 2030 se asemejaría a una parada o etapa en el camino. Al igual que en un viaje por carretera, cada etapa representa un hito que debemos alcanzar para avanzar hacia nuestro destino final. Podemos

visualizar gráficamente esta metáfora utilizando una carretera con marcadores que representan cada uno de los objetivos de la agenda 2030. A medida que avanzamos en el camino, alcanzamos y marcamos cada meta cumplida, creando así un recorrido visual del progreso del proyecto. Esta representación no solo es visualmente efectiva, sino que también permite una comprensión intuitiva del avance y las etapas que se van superando en la implementación de la agenda 2030.

6. Conclusión

Aplicar técnicas de data mining junto con instrumentos cualitativos nos permite obtener un enfoque holístico del fenómeno social. El data mining nos ayuda a extraer patrones y tendencias ocultas en grandes conjuntos de datos, revelando conexiones significativas entre variables. Por otro lado, las técnicas cualitativas nos permiten conocer la visión individual de un fenómeno y hacer un mapeo de conjuntos coherentes, identificando subgrupos dentro de la población estudiada. Al combinar estas dos técnicas, podemos obtener una comprensión más completa y profunda del fenómeno social, identificando patrones de consumo o tendencias de preferencia en ciertos grupos de población. Este enfoque holístico nos ayuda a visualizar el fenómeno social en su totalidad, desde las tendencias generales hasta las particularidades de cada grupo. Además, nos permite tomar decisiones más informadas y diseñar estrategias más efectivas al comprender las necesidades y comportamientos de cada segmento de la población. En última instancia, aplicar estas técnicas nos acerca a una comprensión más completa y detallada de la complejidad del tejido social en el que vivimos.

Aplicar técnicas de data mining junto con instrumentos cualitativos permite obtener un enfoque holístico del fenómeno social. El data mining ayuda a extraer patrones y tendencias ocultas en grandes conjuntos de datos, revelando conexiones significativas entre variables

En cuanto a la presentación de datos a través de narrativas visuales, hemos explorado diversas formas de presentar información de manera efectiva y clara, pudiendo destacar:

Importancia de lo visual: las narrativas visuales son fundamentales para comunicar información compleja de manera accesible y atractiva en ciencias sociales.

Organización jerárquica: elementos como diagramas circulares y pirámides ayudan a organizar jerárquicamente los conceptos, facilitando la comprensión de la información.

Interpretación intuitiva: el uso del tamaño de iconografías conocidas permite una interpretación intuitiva de rangos y escalas, simplificando la comprensión de la información.

Metáforas visuales: las metáforas visuales, como un trayecto en carretera para representar un proyecto, ofrecen una manera efectiva de visualizar y comprender el progreso y las etapas de un proyecto.

En resumen, las narrativas visuales en ciencias sociales son una excelente alternativa para transmitir datos, conceptos y hallazgos de manera clara y efectiva. Así podemos simplificar la información y hacerla más accesible para una amplia audiencia, promoviendo una comprensión más profunda y significativa de los temas tratados.

7. Referencias

Aguirre-Baztán, A. (1995). *Etnografía. Metodología cualitativa en la investigación sociocultural*. Marcombo.

Brown, T. R. (2023). *An introduction to R and Python for data analysis a side-by-side approach*. Chapman and Hall/CRC.

Carminotti, M.; Toppi, P. (2020). *Metodología de la investigación social: Caja de herramientas*. Eudeba.

Cervi, L.; Tejedor, S.; Recoder-Sellarés, M. J. (2024). Qué opinan los periodistas de las herramientas de verificación potenciadas por Inteligencia Artificial. *Infonomy*, 2(2).
<https://doi.org/10.3145/infonomy.24.027>

Daines, J. G.; Nimer, C. L. (2015). In search of primary source literacy: Opportunities and challenges. *RBM: A journal of rare books, manuscripts, and cultural heritage* 16(1), 19-34.
<https://doi.org/10.5860/rbm.16.1.433>

Day, A. (2011). *Changing the social fabric of organisations: the importance of participation*. The Ashridge Journal.
https://eoe.leadershipacademy.nhs.uk/wp-content/uploads/sites/6/dlm_uploads/2019/07/1317116794_wmbF_changing_the_social_fabric_of_organisations.pdf

Delaney, C.; Kaspin, D. (2017). *Investigating culture: An experiential introduction to anthropology*. Wiley-Blackwell.

De-Waal, T.; Van-Delden, A.; Scholtus, S. (2020). Multi-source statistics: Basic situations and methods. *International Statistical Review*, 88(1), 203-228.
<https://doi.org/10.1111/insr.12352>

Dias-do-Nascimento, J.; Meireles-Gomes, I.; Ribeiro-Lacerda, M.; Braga-de-Camargo, T.; Catafesta-Utzumi, F.; Bernardino, E. (2016). Uso del software NVivo® en una investigación con Teoría Fundamentada. *Index de Enfermería*, 25(4).
<https://www.index-f.com/index-enfermeria/v25n4/10639r.php>

Consoli, S. (2021). Uncovering the hidden face of narrative analysis: A reflexive perspective through MAXQDA. *System*, 102.
<https://doi.org/10.1016/j.system.2021.102611>

Flick, U. (2007). *El diseño de la investigación cualitativa*. Morata.

Grima, P. (2011) *La certeza absoluta y otras ficciones (Los secretos de la Estadística)*. RBA.

Guerrero-Carmona, F. M.; Marcos-Aldón, M. (2014), Procesamiento del lenguaje natural. *Revista Instrumental* (monografía).

Hernández-Sampieri, R. (2018). *Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta*. McGraw-Hill Interamericana.

Joyanes-Aguilar, L. (2020) *Inteligencia de negocios y analítica de datos*. Marcombo, Ediciones Técnicas.

Kunicki, Z. J.; Zambrotta, N. S.; Tate, M. C.; Surrusco, A. S.; Risi, M. M.; Harlow, L. L. (2019). Keep your stats in the cloud! Evaluating the use of Google Sheets to teach quantitative methods. *Journal of Statistics Education*, 27, 188-197.
<https://doi.org/10.1080/10691898.2019.1665485>

Lévi-Strauss, C. (1958). *Antropología estructural*. Paidós.

Maravelakis, P. (2019). The use of statistics in social sciences. *Journal of Humanities and Applied Social Sciences*, 1(2), 87-97.
<https://doi.org/10.1108/JHASS-08-2019-0038>

Mathes, E. (2023). *Python crash course, 3rd edition: A hands-on, project-based introduction to programming*. No Starch Press.

Ñaupas-Paitán, H.; Mejía-Mejía, E.; Novoa-Ramírez, E.; Villagómez-Paucar, A. (2014). *Metodología de la investigación cuantitativa - cualitativa y redacción de la tesis*. Ediciones de la U.

Ortega-Candel, J. M. (2022). *Big data, machine learning y data science en Python*. RA-MA S. A. Editorial y Publicaciones.

Paulus, T. M.; Lester, J. N. (2016). *Atlas.ti* for conversation and discourse analysis studies. *International Journal of Social Research Methodology*, 19(4), 405-428.
<https://doi.org/10.1080/13645579.2015.1021949>

Poldrack, R. A. (2023) *Statistical thinking: Analyzing data in an uncertain world*. Princeton University Press.

Ramírez-Mercado, M. (2023). *Metodología cualitativa: uso y aplicación de técnicas para el estudio social*. México: UNAM, Facultad de Estudios Superiores Aragón.

Sagaró-del-Campo, N.; Zamora-Matamoros, L. (2019). Técnicas estadísticas para identificar posibles relaciones bivariadas. *Revista Cubana de Anestesiología y Reanimación*, 19(2).
<https://revanestesia.sld.cu/index.php/anestRean/article/view/603/909>

Şahin, M.; Aybek, E. (2019). *Jamovi*: An easy to use statistical software for the social scientists. *International Journal of Assessment Tools in Education*, 6(4), 670-692.
<https://doi.org/10.21449/ijate.661803>

Sampedro, V. (2014). *El cuarto poder en red. Por un periodismo (de código) libre*. Icaria. ISBN: 8498885906

Sarramona, J. (2023). *La investigación en ciencias sociales: Posibilidades y limitaciones*. Alpha Editorial.

Shepherd, M. A.; Richardson, E. J. (2024). Opting for open-source? A review of free statistical software programs. *Teaching Statistics*, 46(1), pp. 53-63.
<https://doi.org/10.1111/test.12360>

Sidorenko-Bautista, P.; Herranz-de-la-Casa, J. M. (2024). Realidad extendida, computación espacial y metaverso ¿instrumentos de marketing o evidencias de un cambio de paradigma en la comunicación?. *Infonomy*, 2(2).
<https://doi.org/10.3145/infonomy.24.020>

Sto.-Tomas, M.; Tindowen, D. J.; Mendezabal, M. J.; Quilang, P.; Agustin, E. T. (2019). The use of *PSPP* software in learning statistics. *European Journal of Educational Research*, 8(4), 1127-1136.
<https://doi.org/10.12973/eu-jer.8.4.1127>

Vázquez-Atochero, A. (2012). La tiranía de Gauss. Prejuicios y perjuicios de la normalidad en las ciencias sociales. *Revista Caracteres*, 1(2) 62-70.

Vicente-Cestero, E.; Mateos-Caballero, A. (2018). *Data science y redes complejas. Métodos y aplicaciones*. Editorial Universitaria Ramón Areces.