



Revista de Educación Matemática

Editada por la Sociedad Andaluza de Educación Matemática Thales

Visualización de datos
EN EDUCACIÓN:
herramientas para el
desarrollo de la
competencia digital

118
2024

Equipo Editorial

Burgos Navarro, María José
Carrillo de Albornoz Torres, Agustín
Cecilia Gámiz, Lina María
Conde Fernández, Silvia
Contreras García, José Miguel
Cruz, María Florencia
España Pérez, Francisco
Fernández Plaza, Jose Antonio
Flores Lamolda, Lucía
Flores Martínez, Pablo
Gallardo Jiménez, Sandra
Gámez Valero, Carmen
García Schiaffino, Margarita
Garzón Guerrero, José Antonio
López Centella, Esperanza
Lupiáñez Gómez, José Luis
Molina Muñoz, David
Molina Portillo, Elena
Montejo Gámez, Jesús
Moreno Verdejo, Antonio
Partal García, Daniel
Pérez Martos, María del Carmen
Ramírez Uclés, Rafael
Rivas Olivo, Mauro Alfredo
Rodríguez González, Miguel
Roquette Rodríguez, Esther
Ruiz Hidalgo, Juan Francisco
Tizón Escamilla, Nicolás
Valero Terrón, Iván
Villegas Escobar, Adela María

118
2024

Edita
Sociedad Andaluza de
Educación Matemática "Thales"
Universidad de Cádiz
C.A.S.E.M.
Facultad de Ciencias
Departamento de Matemáticas
Campus del Río San Pedro
Torre Central, 4ª planta
11510 Puerto Real (Cádiz)
Teléfono: 956012833
Email: thales.matematicas@uca.es

Depósito Legal
SE-421-1984

ISSN
2340-714X

Período
2024

Suscripción
Anual

ÍNDICE / CONTENTS

7

Prólogo

APORTES DE LA INVESTIGACIÓN EN EDUCACIÓN MATEMÁTICA

9

Explorando la visualización de datos

Elena Molina Portillo, José Miguel Contreras García, David Molina Muñoz y Rubén Cabrera Lozano

21

Visualización de datos con CODAP

Elena Molina Portillo y Zuly Restrepo Patiño

31

Visualización de datos con Tuva

Elena Molina Portillo y Carlos Santiago Sánchez Muñoz

41

Visualización gráfica con jamovi

David Molina Muñoz y Sara Venegas Sepúlveda

53

Visualización de datos con PSPP

Inmaculada Serrano García y Ana Enríquez Molina

69

Visualización de datos con Power BI

José Miguel Contreras García y Rubén Cabrera Lozano

81

Exploración y visualización de datos en Educación Matemática con Jupyter y Python

José Miguel Contreras García y Cristina Hidalgo Castillo

91

Visualización de datos con R

Carlos Santiago Sánchez Muñoz y Manuel Alejandro Jiménez Morales

ACTIVIDADES EN LA SAEM THALES

101

Tercer anuncio del 19 CEAM. Huelva, del 12 al 14 de abril de 2025

Comités Local de Organización y Científico

PRÓLOGO

La sociedad actual se caracteriza por una creciente dependencia de los datos para comprender fenómenos complejos y tomar decisiones informadas. Este contexto exige que las instituciones educativas promuevan el desarrollo de competencias que permitan a los estudiantes no solo analizar, sino también comunicar la información contenida en los datos de forma clara, efectiva y crítica. En este sentido, la alfabetización informacional y en datos, constituye un eje fundamental dentro de la sexta área del Marco de Referencia de la Competencia Digital Docente, orientada a la facilitación de la competencia digital de los estudiantes.

El presente monográfico se inscribe en esta necesidad formativa en alfabetización en datos, entendida como la capacidad para interpretar, manipular, analizar y visualizar datos, ofreciendo una perspectiva práctica basada en el uso de distintas herramientas para la representación de la información basada en datos. Consta de ocho artículos que abordan, en primera instancia, una introducción que sirve de eje vertebrador, y posteriormente, el estudio de diversas plataformas, programas y herramientas que facilitan la representación gráfica y el análisis de datos en el ámbito educativo.

El primer artículo introduce el marco conceptual y metodológico que articula el monográfico. Mediante éste se explora la relación entre la visualización de datos y la competencia digital ciudadana, una de las competencias clave en el actual currículo educativo. Se introduce una metodología común de trabajo basada en la utilización de una misma base de datos extraída de la Encuesta sobre Equipamiento y Uso de las TIC en los Hogares (INE, 2024), que constituirá el nexo común de las actividades y visualizaciones presentadas en los diferentes trabajos. Con dicha fundamentación, los siguientes siete artículos se centran en el análisis de herramientas específicas, evaluando sus potencialidades y limitaciones en distintos niveles educativos, desde la Educación Primaria hasta el Bachillerato.

Cada uno de estos artículos analiza una herramienta en particular, como CODAP, Tuva, Jamovi, PSPP, Power BI, Jupyter-Python y R, todas ellas seleccionadas por su relevancia y accesibilidad para el ámbito educativo. Se exponen ejemplos prácticos, destacando cómo estas herramientas pueden facilitar la visualización e interpretación de los datos, al tiempo que promueven un aprendizaje activo y significativo. Además, se evalúan aspectos clave como la facilidad de uso, el grado de interactividad y los niveles educativos en los que se considera adecuado su uso.

Este monográfico tiene como propósito principal dotar al profesorado de recursos prácticos y accesibles que les permiten incorporar la visualización de datos en sus prácticas pedagógicas, fomentando así la alfabetización en datos desde etapas tempranas. Se considera que el uso adecuado de estas herramientas no solo mejora la comprensión de los datos, sino que también contribuye al desarrollo del pensamiento

crítico y la competencia digital, habilidades esenciales para participar plenamente en una sociedad basada en el conocimiento.

Con esta obra, esperamos contribuir al fortalecimiento de una educación orientada a los desafíos del siglo XXI, proporcionando a los docentes herramientas innovadoras que contribuyan a un proceso de enseñanza-aprendizaje dinámico, interactivo y relevante para los estudiantes.

Elena Molina Portillo

Explorando la visualización de datos

Elena Molina Portillo
Universidad de Granada, elemo@ugr.es

José Miguel Contreras García
Universidad de Granada, jmcontreras@ugr.es

David Molina Muñoz
Universidad de Granada, dmolinam@ugr.es

Rubén Cabrera Lozano
IES José Luis López Aranguren, rcabreralozano@educa.madrid.org

Resumen: Los cambios sociales, tecnológicos y curriculares requieren competencias digitales para analizar y comunicar la información contenida en los datos. Este monográfico, constituido por 8 estudios, ofrece al profesorado de matemáticas recursos para la exploración de datos mediante su visualización, apropiados para diferentes niveles de demanda cognitiva y habilidades digitales. Este primer trabajo establece las bases metodológicas presentando la base de datos del INE sobre el uso de tecnología en menores que se utilizará en todos los trabajos y analizando algunas representaciones tabulares de los datos. Los estudios posteriores profundizan en la utilización de diversos softwares proporcionando distintas formas de visualización. Se concluye que un uso adecuado de las herramientas tecnológicas facilita la comprensión de los datos.

Palabras clave: visualización de datos, análisis de datos, competencia digital, gráficos estadísticos, representación tabular.

Exploring data visualization

Abstract: Social, technological, and curricular advances require digital competences to analyze and communicate information embedded in data. This monograph, consisting of 8 studies, provides mathematics educators with resources for data exploration and visualization adapted to different levels of cognitive demand and digital skills. This first study establishes the methodological foundations presenting an INE database on technology use among minors that will be used by all the studies and analyzing some tabular representation. Subsequent studies will explore different visualization methods facilitated by specialized software. The findings highlight that the appropriate use of technological tools significantly enhances data comprehension.

Key words: data visualization, data analysis, digital competence, statistical graphs, tabular representation.

1. INTRODUCCIÓN

La cantidad de datos disponibles en cualquier ámbito de conocimiento, incluso en el contexto de nuestra vida personal, hace necesario disponer de las habilidades suficientes para extraer

información de ellos. El proceso de alfabetización digital se inicia en la etapa de Infantil y se concreta para cada una del resto de las etapas de educación preuniversitaria (Primaria, Secundaria y Bachillerato) en cinco descriptores operativos dentro del perfil de salida en sus distintos desarrollos curriculares (Real Decreto [RD] 157/2022, RD 217/2022 y RD 243/2022). Dichos descriptores operativos están conceptualizados en correspondencia con las áreas competenciales del Marco de Referencia de la Competencia Digital Ciudadana, #DigComp 2.2 (Vuorikari et al., 2022), y, por ende, con las competencias de la sexta área del Marco de Referencia de la Competencia Digital Docente, MRCDD (Ministerio de Educación y Formación Profesional y Administraciones Educativas de las Comunidades Autónomas [MEFP y AECA], 2022).

En este contexto, la competencia digital que se pretende desarrollar en el alumnado tiene un enfoque multidisciplinar, promoviéndose habilidades y destrezas aplicables a cualquier ámbito de conocimiento. La actual normativa se ha impregnado de conceptos referidos a la competencia digital tales como pensamiento computacional, análisis crítico de la información, alfabetización en información y datos, y educación mediática, entre otros. En dicha competencia digital ciudadana, la alfabetización mediática y el tratamiento de la información y los datos es la primera destreza a desarrollar en el alumnado. Por tanto, se hace necesario promover la educación en el tratamiento de datos, entendiendo por este su recolección, almacenamiento y organización, estructura y conservación, modificación y consulta, uso y comunicación, restricción y supresión, así como su evaluación y destrucción.

En particular este monográfico se enfoca en presentar al profesorado de matemáticas de los distintos niveles educativos diversas herramientas para la exploración de datos mediante su visualización, atendiendo a los diferentes niveles de demanda cognitiva y habilidades digitales. Para este propósito, el monográfico se estructura en 8 artículos. Este primer estudio introduce el planteamiento y objetivo general, centrado en analizar el uso de las tecnologías de la información y comunicación (TIC) en niños de 10 a 15 años en España. Asimismo, se describe el marco teórico, la metodología aplicada y la base de datos seleccionada ofreciendo, finalmente, una representación tabular como referencia común para la interpretación de resultados. Los siguientes 7 estudios exploran la visualización de datos mediante gráficos estadísticos a través de diversos softwares tales como CODAP, Tuva, jamovi, PSPP, Power BI, Python y R, tomando como punto de partida las cuestiones planteadas en este primer trabajo.

2. ANTECEDENTES

En las últimas décadas una parte importante del tratamiento de datos se ha abordado en el bloque de contenidos de Estadística y Probabilidad mediante el análisis de datos (recolección, limpieza, transformación, exploración -mediante visualización y/o técnicas de análisis descriptivo-, modelado, e interpretación y comunicación de resultados). Actualmente, el análisis de datos en el ámbito escolar sigue presentándose en el desarrollo del sentido estocástico del área de matemáticas, como muestran las siguientes referencias:

“El sentido estocástico se orienta hacia el razonamiento y la interpretación de datos y la valoración crítica, así como la toma de decisiones a partir de información estadística. También comprende los saberes vinculados con la comprensión y la comunicación de fenómenos aleatorios en situaciones de la vida cotidiana” (RD 157/2022, p. 24486)

“El sentido estocástico comprende el análisis y la interpretación de datos, la elaboración de conjeturas y la toma de decisiones a partir de la información estadística, su valoración

crítica y la comprensión y comunicación de fenómenos aleatorios en una amplia variedad de situaciones cotidianas” (RD 217/2022, p. 142 y RD 243/2022, p. 46334)

Entre los elementos curriculares del sentido estocástico asociados al análisis de datos se destacan los saberes básicos de organización y análisis de datos (Tabla 1).

Tabla 1

Saberes básicos de organización y análisis de datos según la etapa educativa

	1er Ciclo	2º Ciclo	3er Ciclo
Educación Primaria	– Estrategias de reconocimiento de los principales elementos y extracción de la información relevante de gráficos estadísticos sencillos de la vida cotidiana (pictogramas, gráficas de barras...). – Estrategias sencillas para la recogida, clasificación y recuento de datos cualitativos y cuantitativos en muestras pequeñas. – Representación de datos obtenidos a través de recuentos mediante gráficos estadísticos sencillos y recursos manipulables y tecnológicos.	– Gráficos estadísticos de la vida cotidiana (pictogramas, gráficas de barras, histogramas...): lectura e interpretación. – Estrategias sencillas para la recogida, clasificación y organización de datos cualitativos o cuantitativos discretos en muestras pequeñas mediante calculadora y aplicaciones informáticas sencillas. Frecuencia absoluta: interpretación. – Gráficos estadísticos sencillos (diagrama de barras y pictogramas) para representar datos, seleccionando el más conveniente, mediante recursos tradicionales y aplicaciones informáticas sencillas. – La moda: interpretación como el dato más frecuente. – Comparación gráfica de dos conjuntos de datos para establecer relaciones y extraer conclusiones.	– Conjuntos de datos y gráficos estadísticos de la vida cotidiana: descripción, interpretación y análisis crítico. – Estrategias para la realización de un estudio estadístico sencillo: formulación de preguntas, y recogida, registro y organización de datos cualitativos y cuantitativos procedentes de diferentes experimentos (encuestas, mediciones, observaciones...). – Tablas de frecuencias absolutas y relativas: interpretación. – Gráficos estadísticos sencillos (diagrama de barras, diagrama de sectores, histograma, etc.): representación de datos mediante recursos tradicionales y tecnológicos y selección del más conveniente. – Medidas de centralización (media y moda): interpretación, cálculo y aplicación. – Medidas de dispersión (rango): cálculo e interpretación. – Calculadora y otros recursos digitales, como la hoja de cálculo, para organizar la información estadística y realizar diferentes visualizaciones de los datos. – Relación y comparación de dos conjuntos de datos a partir de su representación gráfica: formulación de conjeturas, análisis de la dispersión y obtención de conclusiones.
	1er - 3er curso	Matemáticas A // Matemáticas B	
Educación Secundaria	– Estrategias de recogida y organización de datos de situaciones de la vida cotidiana que involucren una sola variable. Diferencia entre variable y valores individuales. – Análisis e interpretación de tablas y gráficos estadísticos de variables cualitativas, cuantitativas discretas y cuantitativas continuas en contextos reales. – Gráficos estadísticos: representación mediante diferentes tecnologías (calculadora, hoja de cálculo, aplicaciones, ...) y elección del más adecuado. – Medidas de localización: interpretación y cálculo con apoyo tecnológico en situaciones reales. – Variabilidad: interpretación y cálculo, con apoyo tecnológico, de medidas de dispersión en situaciones reales. – Comparación de dos conjuntos de datos atendiendo a las medidas de localización y dispersión.	– Estrategias de recogida y organización de datos de situaciones de la vida cotidiana que involucren una variable bidimensional. Tablas de contingencia. – Análisis e interpretación de tablas y gráficos estadísticos de una y dos variables cualitativas, cuantitativas discretas y cuantitativas continuas en contextos reales. – Medidas de localización y dispersión: interpretación y análisis de la variabilidad. – Gráficos estadísticos de una y dos variables: representación mediante diferentes tecnologías (calculadora, hoja de cálculo, aplicaciones, ...), análisis, interpretación y obtención de conclusiones razonadas. – Interpretación de la relación entre dos variables, valorando gráficamente con herramientas tecnológicas la pertinencia de realizar una regresión lineal. Ajuste lineal con herramientas tecnológicas.	
	Matemáticas I // Matemáticas aplicadas a las CCSS // Matemáticas Generales		
Bachillerato	– Organización de los datos procedentes de variables bidimensionales: distribución conjunta y distribuciones marginales y condicionadas. Análisis de la dependencia estadística. – Estudio de la relación entre dos variables mediante la regresión lineal y cuadrática: valoración gráfica de la pertinencia del ajuste. Diferencia entre correlación y causalidad. – Coeficientes de correlación lineal y de determinación: cuantificación de la relación lineal, predicción y valoración de su fiabilidad en contextos: ▪ científicos y tecnológicos (Matemáticas I) ▪ de las ciencias sociales (Mat. aplicadas a las CCSS) ▪ científicos, económicos, sociales, etc. (Mat. Generales) – Calculadora, hoja de cálculo o software específico en el análisis de datos estadísticos.		

Fuente: RD 157/2022, RD 217/2022 y RD 243/2022

No obstante, también se pueden observar en algunos saberes asociados a inferencia y probabilidad en Matemáticas A y B de 4º de la Educación Secundaria Obligatoria (ESO), por ejemplo:

“Estrategias y herramientas de presentación e interpretación de datos relevantes en investigaciones estadísticas mediante herramientas digitales adecuadas (RD 217/2022, pp. 41744 y 41739)”;

“Cálculo aplicando la regla de Laplace y técnicas de recuento en experimentos simples y compuestos (mediante diagramas de árbol, tablas...) y aplicación a la toma de decisiones fundamentadas” (RD 217/2022, pp. 41739 y 41744),

ó en Matemáticas Aplicadas a las Ciencias Sociales II en Bachillerato

“Herramientas digitales en la realización de estudios estadísticos” (RD 243/2022, p. 46358);

“Modelización de fenómenos estocásticos mediante las distribuciones de probabilidad binomial y normal. Cálculo de probabilidades asociadas mediante herramientas tecnológicas” (RD 243/2022, p. 46358).

En todo caso, los saberes básicos de organización y análisis de datos en el sentido estocástico que contiene el actual desarrollo curricular de la actual ley educativa no suponen una ruptura con los contenidos de las anteriores normativas curriculares. La principal divergencia es en el abordaje competencial que confiere la actual ley educativa que, sin ser nuevo, promueve una instrucción en base al diseño de situaciones de aprendizaje y una evaluación competencial.

En este marco ya no es suficiente conocer los conceptos estadísticos, saber organizar datos manualmente y/o construir determinados gráficos estadísticos, como tradicionalmente se ha entendido el proceso de enseñanza-aprendizaje de la estadística. Actualmente para cualquier componente del análisis de datos mencionadas se hace necesario, además, el manejo adecuado de diferentes herramientas digitales que permitan el uso de datos reales y el desarrollo de habilidades para la interpretación, evaluación y comunicación de la información que se produce. De igual manera, en línea con el MRCDD, el docente debe fomentar en el alumnado el uso de las herramientas necesarias para el desarrollo de esa alfabetización en datos. Dichas herramientas se ubicarán como idóneas en distintas etapas educativas atendiendo a los diferentes niveles de demanda cognitiva y habilidades digitales, favoreciendo una progresión en el desarrollo de las competencias digitales y una comprensión más profunda y detallada de los conjuntos de datos en estudio.

3. MARCO TEÓRICO

Los gráficos estadísticos son una de las principales herramientas para la visualización de datos. Estos muestran información compleja de manera concisa y permiten transmitirla de forma atractiva, lo que ha conllevado el desarrollo de numerosas técnicas de visualización y softwares disponibles para su diseño. No obstante, la información transmitida podría ser confusa o/y errónea si no se tienen en cuenta características importantes para su selección, su elaboración y para la interpretación de la información que mediante éstos se representa.

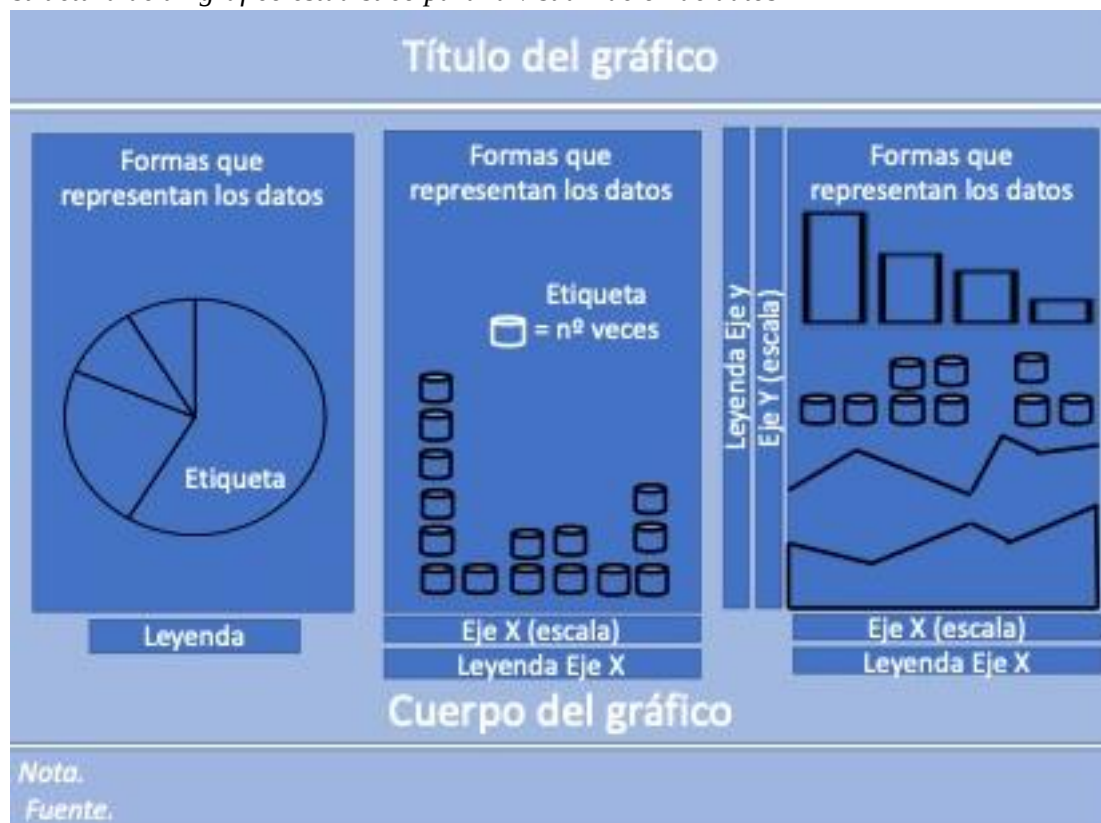
Entre los aspectos clave que se deben tener en cuenta destaca que la representación utilizada debe ser auto-explicativa de la información implícita en los datos, que existen diferentes tipos de gráficos según la naturaleza de la/s variable/s que se representa/n y que debe ser adecuada al propósito final. Entre la gran variedad de gráficos disponibles para la visualización de datos, en las citadas etapas educativas es habitual trabajar con diagrama de barras, sectores, histogramas, pictogramas, diagramas de líneas o de áreas, y diagramas de dispersión.

La Figura 1 ilustra que cualquiera de estas tipologías se debe regir por la siguiente estructura común (Ministerio de Trabajo y Seguridad Social, 2013):

1. *Título del gráfico*, que proporciona el contenido de la información que está representada, atendiendo o describiendo los siguientes interrogantes:
 - ¿qué? Contenido, característica o variable principal
 - ¿cómo? Indica las categorías en las que se organiza la información
 - ¿cuándo? Referencia temporal a los datos
 - ¿dónde? Referencia geográfica de los datos
2. *Cuerpo del gráfico*, es la representación de los datos. En él se distinguen:
 - la figura o forma con la que se representarán los datos, prestando especial interés a la proporcionalidad entre estas
 - ejes y escalas, que sirven de referencia de posición y de comparación entre las distintas categorías, respectivamente
 - leyendas o etiquetas, que aportan información sobre los ejes, escalas o categorías
3. *Notas y fuente*, las primeras son aclaraciones de carácter conceptual o metodológico de los datos, mientras que la segunda, ubicada al final del gráfico, señala la procedencia de los datos o información que se muestra.

Figura 1

Estructura de un gráfico estadístico para la visualización de datos



Fuente: elaboración propia

De igual forma, se debe prestar atención a las distintas habilidades en la interpretación y comunicación de información mediante gráficos estadísticos, desde la lectura literal del gráfico hasta la valoración crítica del método de recopilación, validez y fiabilidad de los datos, incluyendo las posibilidades de generalización de las conclusiones. En este sentido, se considera

el modelo propuesto por Friel et al. (2001). Estos autores, partiendo de la clasificación para la lectura de gráficos propuesta por Curcio (1989):

- “*leer entre los datos*”, lectura literal del gráfico sin interpretar la información contenida en el mismo
- “*leer dentro de los datos*”, interpretación e integración de los datos del gráfico. Esta capacidad requiere la comparación de datos o la realización de operaciones con los datos
- “*leer más allá de los datos*”, realizar predicciones e inferencias a partir de los datos sobre informaciones que no se reflejan directamente en el gráfico

añaden un cuarto nivel a su modelo:

- “*leer detrás de los datos*”, valorar críticamente el método de recogida de datos, su validez y fiabilidad, así como las posibilidades de extensión de las conclusiones

En este primer estudio introductorio, la representación tabular adopta lo anteriormente expuesto en cuanto a las características explicitadas para los gráficos. En este sentido esta visualización en forma de tabla también debe ser auto-explicativa, su estructura debe contemplar título, cuerpo de la tabla y notas, y debe facilitar habilidades en la interpretación y comunicación de la información así representada.

4. METODOLOGÍA

Este monográfico se organiza como un conjunto de estudios exploratorio-descriptivos que utilizan la visualización de datos proporcionada mediante diversos softwares para extraer información de una base de datos. En concreto, cada trabajo realiza un análisis detallado de las funcionalidades y capacidades de un programa o software disponible para la exploración de datos mediante visualización, según la etapa educativa para la que se recomienda su uso. En conjunto se ofrece una visión comprensiva de las opciones tecnológicas actuales en educación, facilitando la selección informada de herramientas para mejorar las habilidades digitales en el manejo y representación visual de datos, así como el desarrollo de habilidades en la interpretación y comunicación de información representada. Los programas utilizados en los diferentes estudios son: CODAP, Tuva, jamovi, PSPP, Power BI, Python y R.

Cada uno de estos softwares se presenta tomando como objetivo común el estudio real sobre el uso de TIC para niños de 10 a 15 años en España. Este tópico además de ser un problema real relacionado con el desarrollo de la competencia digital, que puede estar en el ámbito de interés de profesorado y alumnado, está alineado con los Objetivos de Desarrollo Sostenible (ODS), específicamente con la meta 4.4 (Naciones Unidas, 2015).

Para tal fin se consideran los datos muestrales provenientes de la Encuesta sobre equipamiento y uso de TIC en los hogares españoles para el año 2023, que ofrece en abierto el Instituto Nacional de Estadística (INE, 2024a). De su metodología, que puede ser consultada en la web de dicho organismo, se destaca el muestreo trietápico donde la información de los menores se obtiene a partir de una persona de 16 o más años que haya sido seleccionada en la vivienda familiar.

Por tanto, como preprocesamiento de los datos aportados, se seleccionan los casos que contienen información sobre uso de TIC para niños de 10 a 15 años, considerando a los menores de una misma unidad familiar como sujetos independientes para nuestro estudio. Para la realización de los distintos análisis se consideran variables:

- *identificativas*: número de cuestionario, identificador del niño en el hogar

- *sociodemográficas*: comunidad autónoma (INE, 2024b), provincia, sexo, edad, tipo de hogar, total de miembros del hogar, total de menores entre 10 y 15 años en el hogar
- *equipamiento tecnológico en la vivienda*: ordenador (no se incluye Tablet), Tablet (no se incluye ordenador), teléfono fijo (se incluyen inalámbricos), teléfono móvil, acceso a internet y forma de conexión a internet de banda ancha.
- *de interés*: utilización de ordenador desde su hogar (últimos 3 meses), utilización de internet desde casa (últimos 3 meses) y disponibilidad de teléfono móvil personal.

El análisis realizado en cada uno de los estudios se corresponderá con la visualización de datos y las habilidades estadísticas correspondientes al nivel educativo, según la normativa curricular actual de cada una de las etapas educativas, en el que se considera apropiado la utilización de dicho software.

Como punto de partida y nexo común entre los estudios, este trabajo presenta la representación tabular de datos, destacando la información que dicha visualización proporciona. Inicialmente, se muestra una tabla de frecuencias de la variable *total de menores entre 10 y 15 años en el hogar* para el total de España y para la comunidad autónoma de Andalucía, así como la representación tabular de los resúmenes estadísticos para la variable *edad*. Ambas representaciones son adecuadas desde la etapa de Educación Primaria. Adicionalmente, se introduce la tabla de contingencia que proporciona información sobre la relación entre algunas de las *variables sociodemográficas* (género y edad) y *variables de interés*, cuya representación puede ser indicada para la etapa de Secundaria. Finalmente, de interés para Bachillerato y estudios superiores, se muestra una matriz de correlaciones entre las *variables de equipamiento tecnológico en la vivienda* y las *variables de interés*. Para esta última representación tabular se destaca que las variables utilizadas son cualitativas dicotómicas (SI/NO) con valores perdidos, por lo que se emplea el coeficiente de correlación de phi, aunque su uso en Bachillerato puede ser limitado. No obstante, se considera relevante su presentación porque ejemplifica la interpretación de datos a partir de la matriz de correlaciones, similar a la correlación de Pearson en variables cuantitativas.

Esta representación tabular de los datos puede ser obtenida mediante cualquiera de las herramientas que se presentarán posteriormente. No obstante, en algún software como CODAP, ideado principalmente para la visualización gráfica desde edades tempranas, no se facilita directamente dicha representación. De igual forma, la base de datos de trabajo en el INE está disponible en diferentes formatos, facilitando la importación de dichos datos a cualesquiera de los diferentes programas. En este sentido, tanto para el tratamiento inicial de la base de datos como para su representación tabular, se utiliza el software R.

5. RESULTADOS

Este estudio introductorio aborda la representación tabular de los datos sobre uso de TIC para niños de 10 a 15 años en España y en la comunidad autónoma de Andalucía. Mediante la interpretación de los datos mostrados se pretende ejemplificar la cantidad de información que es posible albergar en esta visualización.

En primer lugar, atendiendo a un nivel básico de descripción necesario y suficiente para la etapa de Educación Primaria, se puede obtener la tabla de frecuencias de cualquier variable. Como ejemplo se muestra la distribución de frecuencias para la variable *total de menores de 10 – 15 años en el hogar* para España y para la comunidad autónoma de Andalucía (Tabla 2).

Tabla 2*Distribución del total de menores de 10-15 años en hogares andaluces y españoles encuestados.*

Menores de 10-15 años en el hogar	Andalucía			España		
	N	FR (%)	FR acum. (%)	N	FR (%)	FR acum. (%)
1	163	66,3	66,3	1609	62,7	62,7
2	74	30,1	96,4	862	33,6	96,3
3	9	3,7	100	90	3,5	99,8
4	0	0	100	4	0,2	100

Nota: FR = frecuencia relativa.

Fuente: *Encuesta sobre Equipamiento y Uso de TIC (año 2023)*. INE (2024a). Elaboración propia.

Esta representación tabular permite observar que el número de menores de 10 a 15 años en el hogar sigue una distribución similar en los hogares españoles y andaluces. En ambos contextos, más de la mitad de los menores son los únicos menores que residen en el hogar, mientras que un tercio de los menores comparten residencia con otro menor en esa franja etaria. El porcentaje de casos en el que conviven tres o más menores con estas edades en una misma vivienda es muy pequeño.

Adicionalmente, desde el tercer ciclo de dicha etapa educativa y para variables cuantitativas, se pueden mostrar los resúmenes estadísticos mediante su representación tabular. Como ejemplo se muestran los resúmenes estadísticos para la variable *edad* para España y para la comunidad autónoma de Andalucía (Tabla 3). En ella se muestra que, en ambos casos, el rango de edades oscila entre los 10 y 15 años, el 50% de los menores tienen edad menor o igual a 13 años y que poseen una distribución casi simétrica, con un leve sesgo hacia la izquierda.

Tabla 3*Medidas de dispersión y tendencia central sobre la edad de los menores en hogares andaluces y españoles encuestados*

Región	Min.	1er cuartil	Mediana	Media	3er cuartil	Max.
Andalucía	10	11	13	12,6	14	15
España	10	11	13	12,6	14	15

Fuente: *Encuesta sobre Equipamiento y Uso de TIC (año 2023)*. INE (2024a). Elaboración propia.

De igual forma, la representación tabular puede ser utilizada para obtener información sobre la relación de dos variables mediante tabla de contingencia, también conocida como tabla de doble entrada o tabla cruzada. Este tipo de visualización se introduce en el currículo para 4º de ESO, permitiendo explorar descriptivamente unas variables en función de otras. Para ilustrarlo se utiliza una tabla de doble entrada para mostrar la información sobre el *total de menores entre 10 y 15 años en el hogar* y su distribución respecto a las *variables sociodemográficas género y edad*, para el total de España y para la comunidad autónoma de Andalucía. De igual forma se presenta la relación entre dichas *variables sociodemográficas y variables de interés (usuarios de ordenador en casa durante los tres últimos meses, usuarios de internet en casa en los tres últimos meses y menores que disponen de teléfono móvil propio)*. En ella se muestra el porcentaje respecto a los totales para Andalucía y España (porcentaje horizontal) de los menores que son usuarios de tecnología (*ordenador, internet y/o móvil propio*), según dichas variables sociodemográficas.

En base a la información que muestra la Tabla 4 se puede calcular que el número de niños residentes en hogares andaluces entrevistados representa solo el 9,6%, y que la distribución por

sexo y edad es uniforme. Se destaca también que aproximadamente 1 de cada 4 menores de 10 a 15 años no dispone de teléfono móvil personal, tanto en Andalucía como en España.

Tabla 4

Número de menores y porcentaje horizontal de menores usuarios de tecnología (PC, internet y/o móvil) según variables sociodemográficas, en los hogares andaluces y españoles encuestados

2023	Nº de menores de 10 a 15 años		% usuarios de ordenador en casa últimos 3 meses		% usuarios de internet en casa últimos 3 meses		% menores que disponen de teléfono móvil	
	Andalucía	España	Andalucía	España	Andalucía	España	Andalucía	España
Total	246	2565	87,0	93,8	92,7	94,3	73,2	70,8
Sexo								
Niños	117	1332	82,9	93,1	91,5	94,4	73,5	70,9
Niñas	129	1233	90,7	94,6	93,8	94,1	72,9	70,6
Edad								
10 años	36	385	83,3	89,4	91,7	90,6	33,3	22,1
11 años	48	430	81,2	92,1	91,7	90,5	45,8	46,0
12 años	38	433	86,8	94,5	94,7	95,2	84,2	72,7
13 años	43	426	90,7	94,6	93,0	94,1	86,0	88,5
14 años	36	438	91,7	95,7	94,4	97,0	100	93,4
15 años	45	453	88,9	96,2	91,1	97,6	91,1	95,1

Fuente: *Encuesta sobre Equipamiento y Uso de TIC (año 2023)*. INE (2024a). Elaboración propia.

De igual forma, se observan porcentajes similares por género del uso de ordenador, internet y disponibilidad de teléfono móvil propio, salvo para el caso de uso de ordenador en Andalucía. No obstante, el uso de ordenador en Andalucía y España, así como el uso de internet en Andalucía, presenta un porcentaje superior en chicas. De igual forma se muestra un porcentaje ligeramente superior en chicos que disponen de teléfono móvil personal en ambas regiones y en chicos que son usuarios de internet en España. Alrededor del 90% de los menores andaluces y del 94% de los menores españoles son usuarios de ordenador e internet. Este porcentaje es ligeramente menor para los andaluces de 10-12 años que usan ordenador y para el uso de ordenador e internet en menores de 11 años españoles. Menos de la mitad de los jóvenes menores de 12 años tienen teléfono móvil personal, edad en la que aproximadamente 3 de cada 4 menores ya disponen de dicho dispositivo, siendo mayor el porcentaje en Andalucía.

Finalmente, a partir de la etapa educativa de Bachillerato podría ser apropiado el estudio de la asociación de distintas variables mediante la matriz de correlaciones. En este sentido sería interesante estudiar si *las características tecnológicas de la vivienda* se asocian con el uso de ordenador, internet y disposición de móvil en menores de 10 a 15 años (Tabla 5). Teniendo en cuenta que en este caso las variables son cualitativas dicotómicas (SI/NO) y que tienen valores perdidos, se calcula el coeficiente phi (ϕ) para cada par de variables

Tabla 5

Correlación entre variables de equipamiento tecnológico del hogar y uso de ordenador, internet y disposición de teléfono móvil en menores de 10 a 15 años andaluces y españoles.

	Usuarios de ordenador en casa últimos 3 meses (%)	Usuarios de Internet en casa últimos 3 meses (%)	Disposición de teléfono móvil (%)	Usuarios de ordenador en casa últimos 3 meses (%)	Usuarios de Internet en casa últimos 3 meses (%)	Disposición de teléfono móvil (%)
	Andalucía			España		
Ordenador	0,242*	0,007	0,050	0,308*	0,139*	0,107*
Tablet	0,320*	0,085	0,079	0,262*	0,116*	0,041*
Tlf. fijo	0,101*	0,049	0,088	0,120*	0,060*	0,007
Tlf. móvil	--	--	--	0,007	0,006	0,044*
Internet	--	--	--	0,084*	0,057*	0,066*
Banda ancha	0,132*	0,063	0,057	0,190*	0,105*	0,077*

Nota: * p-valor significativo para H_0 : independencia de las variables involucradas

Fuente: *Encuesta sobre Equipamiento y Uso de TIC (año 2023)*. INE (2024a). Elaboración propia.

Para España se observa asociación para cada par de variables analizadas, excepto para las variables *teléfono fijo en casa y disponibilidad de móvil propio*, así como entre las variables *teléfono móvil en casa y uso de ordenador e internet en los 3 últimos meses*. No obstante, la fuerza de la asociación de las variables es débil o muy débil salvo para la asociación entre el *equipamiento de ordenador o Tablet en la vivienda y el uso del ordenador en los 3 últimos meses*, cuya fuerza es moderada. Para las viviendas andaluzas no es posible calcular el coeficiente de correlación entre las variables de interés y las variables *teléfono móvil e internet*, ya que éstas presentan un único valor constante en todos los casos. Esto implica que no hay variabilidad en la *disponibilidad de teléfono móvil e internet* en los hogares andaluces, un requisito para calcular las medidas de asociación que se presentan. Adicionalmente, solo se ha constatado asociación entre *equipamiento tecnológico del hogar y uso de ordenador en los tres últimos meses*. Dicha asociación es débil para todos los casos salvo para el coeficiente phi entre *equipamiento de Tablet en casa y uso del ordenador* en los tres últimos meses que es moderada.

6. CONCLUSIONES

Este documento inaugura un monográfico que introduce varias herramientas útiles para la exploración de datos que pueden resultar de interés para el profesorado de matemáticas en distintos niveles educativos. Mediante los estudios que lo componen se realiza una descripción detallada sobre el uso de herramientas, plataformas y/o softwares disponibles para la representación gráfica de datos tales como CODAP, Tuva, jamovi, PSPP, Power BI, Python y R. La elección de dichos instrumentos se fundamenta en múltiples factores. En primer lugar, las herramientas seleccionadas disponen de una buena accesibilidad, pues en su mayoría son softwares disponibles de forma gratuita totalmente, o dispone de una versión gratuita para la visualización de datos con características superiores a las necesarias para los niveles educativos estudiados, como es el caso de Power BI. De igual forma, se presentan programas de distinta potencialidad para que, atendiendo a aspectos como la eficacia y la eficiencia, el profesor disponga de las herramientas digitales necesarias para contribuir al desarrollo de la alfabetización mediática y el tratamiento de la información y los datos en los distintos niveles educativos. Finalmente, los instrumentos presentados tienen la propiedad de mantener la integridad de los datos, es decir, que dichos datos permanecen sin cambios accidentales o

cambios debidos a una manipulación inapropiada, asegurando su confiabilidad y precisión a lo largo del tiempo.

Mediante la utilización de las herramientas digitales apropiadas se realiza el análisis y exploración de la información contenida en cualquier base de datos, contribuyendo a la interpretación y comunicación de la información que contiene. En concreto, para ejemplificar el uso de los distintos softwares seleccionados, se analiza, visualiza y describe la información sobre el uso de TIC para niños de 10 a 15 años españoles en el año 2003 (INE, 2024a).

Una de las limitaciones encontradas en este estudio radica en su propia génesis, pues si bien el objetivo es la introducción y familiarización con el uso de diversas herramientas para el análisis y visualización de datos, esto ha hecho imposible incluir exhaustivamente todos los tipos de análisis, visualizaciones y posibles interpretaciones que dichas herramientas ofrecen. Esta limitación es particularmente relevante en el caso de programas que incluyen una gran cantidad de funcionalidades y niveles de profundidad, como Python o R, que permiten un rango muy amplio de técnicas y métodos avanzados. Dicho cometido podría ser abordado en estudios especializados sobre cada uno de estos softwares para la introducción de técnicas de visualización avanzadas y/o de ciencia de datos.

REFERENCIAS BIBLIOGRÁFICAS

- Curcio, F. (1989). *Developing Graph Comprehension. Elementary and Middle School Activities*. National Council of Teachers of Mathematics.
- Friel, S., Curcio, F. y Bright, G. (2001). Making sense of graphs: critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124-158.
- Instituto Nacional de Estadística [INE] (2024a). Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares, año 2023. <https://tinyurl.com/2bq2py8r>
- Instituto Nacional de Estadística (2024b). Relación de provincias por comunidades autónomas y sus códigos. https://www.ine.es/daco/daco42/codmun/cod_ccaa_provincia.htm
- Ministerio de Educación y Formación Profesional y Administraciones Educativas de las Comunidades Autónomas (2022). *Marco de Referencia de la Competencia Digital Docente*. https://intef.es/wp-content/uploads/2023/05/MRCDD_GTTA_2022.pdf
- Ministerio de Trabajo y Seguridad Social. (2013). *Manual para la elaboración de cuadros y gráficos estadísticos*. Dirección General de Planificación del Trabajo, San José, Costa Rica
- Naciones Unidas (2015). Objetivo 4. Educación de calidad. En *Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible*. Recuperado de <https://www.un.org/sustainabledevelopment/es/sustainable-development-goals/>
- Real Decreto 157/2022, de 1 de marzo, por el que se establecen la ordenación y las enseñanzas mínimas de la Educación Primaria. *Boletín Oficial del Estado*, Madrid, 2 de marzo de 2022, núm. 52 <https://www.boe.es/buscar/pdf/2022/BOE-A-2022-3296-consolidado.pdf>
- Real Decreto 217/2022, de 29 de marzo, por el que se establece la ordenación y las enseñanzas mínimas de la Educación Secundaria Obligatoria. *Boletín Oficial del Estado*, Madrid, 30 de marzo de 2022, núm. 76 <https://www.boe.es/buscar/pdf/2022/BOE-A-2022-4975-consolidado.pdf>

- Real Decreto 243/2022, de 5 de abril, por el que se establecen la ordenación y las enseñanzas mínimas del Bachillerato. *Boletín Oficial del Estado*, Madrid, 6 de abril de 2022, núm. 82 <https://www.boe.es/buscar/pdf/2022/BOE-A-2022-5521-consolidado.pdf>
- Vuorikari, R., Kluzer, S. y Punie, Y. (2022). *DigComp 2.2: The Digital Competence Framework for Citizens*, EUR 31006 EN, Publications Office of the European Union. ISBN 978-92-76-48882-8, <https://doi.org/10.2760/115376>

Visualización de datos con CODAP

Elena Molina Portillo

Universidad de Granada, elemo@ugr.es

Zuly Restrepo Patiño

Praxis für Heilpädagogik und systemische Therapie Ulrich Zimmermann (Alemania),
z.hagenloch@az-bruchsal.de

Resumen: La visualización gráfica de datos ha experimentado un auge debido, en parte, a la cantidad de softwares disponibles para tal fin. En este trabajo se analiza la exploración gráfica de datos mediante la plataforma web CODAP. Para ilustrarlo, se documentan los primeros pasos en el uso de dicha herramienta, así como algunos ejemplos con datos procedentes de la Encuesta sobre Equipamiento y Uso de las TIC para el año 2023 (INE, 2024). Dicho software, debido a su interactividad y simplicidad de uso, se recomienda desde etapas tempranas y para la iniciación en la visualización de datos.

Palabras clave: visualización de datos, gráficos, CODAP, exploración de datos, educación en ciencia de datos.

Data Visualization with CODAP

Abstract: Data visualization has increased, partly due to the amount of software available for this purpose. This paper analyzes data exploration through the CODAP web platform. To illustrate this, the first steps in using the software are documented, as well as some examples referring to the Survey on Equipment and Use of ICT for the year 2023 (INE, 2024). This software is recommended from the early stages and for initiating data exploration due to its interactivity and ease of use.

Key words: data visualization, graphs, CODAP, data exploration, data science education.

1. INTRODUCCIÓN

La visualización gráfica de datos consiste en presentar la información que albergan de manera que se pueda facilitar la comprensión de estos, su análisis y una comunicación efectiva. Una de las principales formas de representación es mediante gráficos, específicamente gráficos estadísticos. Actualmente existen infinidad de representaciones gráficas promovidas en parte por la proliferación de las técnicas e instrumentos disponibles para realizarlas.

Una de las herramientas de las que se dispone para la visualización de datos es CODAP (Common Online Data Analysis Platform), que se define como un software para el análisis de datos creado para educación, de código abierto y gratuito, financiado por el National Science Foundation. Este programa de ciencia de datos basado en la web está diseñado como una plataforma para educadores, investigadores y desarrolladores y como una aplicación para estudiantes a partir del tercer ciclo de Educación Primaria (The Concord Consortium, 2024). Esta herramienta en línea facilita la exploración y análisis de conjuntos de datos, permitiendo la creación de gráficos interactivos.

La plataforma web ofrece múltiples recursos desde tutoriales y ejemplos, hasta foros, reuniones y webinars, para educadores, desarrolladores e investigadores. Además, en su página inicial, en la sección ‘Conceptos básicos de CODAP’ encontramos un enlace que nos facilita la información inicial sobre cómo abrir un documento de muestra, la introducción/importación de datos, así como información introductoria a la creación de gráficos, mapas y tablas, entre otros.

El manejo de este instrumento es muy simple, en el que para introducir los datos solo se necesita arrastrar el fichero hasta la plataforma y para realizar los análisis y gráficos se seleccionan los distintos elementos de navegación del software o menús. Por tanto, no se necesitan habilidades digitales significativas para su utilización, ni en el manejo de archivos ni en programación o codificación. Esto incide en el hecho de que esta herramienta pueda ser utilizada desde edades muy tempranas, en consonancia con las destrezas básicas presupuestas en el primer grupo etario a partir del cual se recomienda su uso.

En línea con el documento de referencia del monográfico de la primera sección titulado ‘*Explorando la visualización de datos*’ (Molina-Portillo et al., 2024), este trabajo tiene como propósito ofrecer al profesorado de matemáticas de los distintos niveles educativos una muestra de la utilización del software CODAP para la exploración de datos mediante su visualización, atendiendo a los diferentes niveles de demanda cognitiva, y habilidades digitales del alumnado.

2. ANTECEDENTES

En el ámbito educativo la investigación en torno a CODAP ha sido abordada desde múltiples perspectivas. Entre los estudios que exploran el software, Mojica et al. (2019) documentan detalladamente cómo utilizarlo para la exploración de datos, Erikson y Chen (2020) destacan la forma innata de realizar el filtrado, agrupamiento y medidas de resumen de los datos, y Erikson (2016) realiza un estudio comparativo con otra herramienta web de análisis de datos, *Tuva* (Tuva Labs Inc., 2024). El enfoque de dichos estudios se asemeja en cuanto al uso de datos reales para su propósito, si bien el primero es más descriptivo y el resto se basa en la aplicación práctica por parte de los estudiantes del uso de dicho software.

Este análisis de experiencias en sesiones de aula con el alumnado también se ha utilizado para introducir conceptos de ciencia de datos en Secundaria, centrándose en el razonamiento estadístico y la exploración de datos (Budde et al., 2020; Frischemeier et al., 2021). De igual forma, se emplea la metodología de estudio de caso para introducir el trabajo con datos multinivel en CODAP, es decir, datos que presentan estructuras jerárquicas subyacentes (Haldar et al., 2018).

Finalmente, el software también ha sido explorado en sesiones de formación del profesorado. Shreiner y Guzdial (2022), a través de un recurso educativo en abierto, presentan diversas herramientas tecnológicas que pueden asistir en la enseñanza de la alfabetización en datos en ciencias sociales, encontrando debilidades del modelo de aceptación de la tecnología por parte del profesorado. Esto se alinea con lo descrito por Peters et al. (2024) quienes abordan la necesidad de involucrar y preparar a los educadores para la enseñanza de estadística y ciencia de datos, disciplinas cada vez más importantes en el currículo educativo, enfatizando la importancia del desarrollo profesional continuo y el apoyo institucional para que los educadores se mantengan actualizados en estos campos.

Otro tipo de estudios muestran varias aplicaciones educativas, incluida la enseñanza de árboles de decisión para problemas de clasificación (Engel et al., 2019; Erikson y Engel, 2023),

o el examen de desigualdades sociales y espaciales (Poling y Weiland, 2020), con intención de hacer que la exploración de datos sea más accesible y atractiva para los estudiantes más nóveles.

Para esta investigación tomamos en consideración la fundamentación teórica descrita en Molina-Portillo et al. (2024)

3. METODOLOGÍA

Para ilustrar la utilización del software CODAP en la visualización y análisis de datos se va a realizar un estudio exploratorio-descriptivo tomando como tópico de interés el estudio real sobre el uso de las tecnologías de la información y la comunicación (TIC) en niños de 10 a 15 años, a partir de la *Encuesta sobre Equipamiento y Uso de las TIC para el año 2023* (Instituto Nacional de Estadística [INE], 2024). Este ha sido ampliamente descrito en el primer estudio del monográfico por lo que, teniendo en cuenta la fuente, la selección de casos, variables y resultados allí documentados, se focaliza este apartado en indicar el análisis que se va a realizar, describiendo cómo se lleva a cabo.

En un primer lugar, se realiza un acercamiento al software describiendo su interfaz y las primeras acciones necesarias para trabajar con él. Posteriormente, se presentan visualizaciones para el conjunto de datos en estudio que mostrarán información sobre la distribución de menores de 10 a 15 años en la *Encuesta sobre Equipamiento y Uso de las TIC para el año 2023*, para España y por comunidad/ciudad autónoma. De igual forma se presentan las medidas de resumen estadístico junto con las opciones gráficas del diagrama de cajas y bigotes o la curva normal. Finalmente se mostrará gráficamente la relación entre dos variables, bien mediante la distribución de casos en las categorías (si las variables son cualitativas) o bien mediante su diagrama de dispersión (con el ajuste de mínimos cuadrados, si las variables son cuantitativas).

4. RESULTADOS

4.1. Primeros pasos

En la web CODAP (The Concord Consortium, 2024), seleccionando ‘lanzamiento de CODAP’ y posteriormente ‘crear un nuevo documento’, se muestra la interfaz de la aplicación. En la esquina superior izquierda se puede leer ‘Documento sin título’ y a su lado aparece un primer menú colapsado, que denominamos *menú 1* para poder hacer referencia a él. Si se selecciona, se muestra un desplegable con opciones de *abrir, guardar, etc.*, similar al de ‘archivo’ de cualquier programa. Justo debajo se encuentra un segundo elemento de navegación del software (*menú 2*) con las opciones para *tablas, gráficos, mapa, deslizador, calculadora, texto, y plugins*; y a la derecha se muestra el que denominaremos *menú 3* con las opciones *deshacer, rehacer, elementos, opciones y ayuda* que, aunque obvias para el lector, procederemos a describir a medida que vaya siendo requerido (Figura 1).

Para abrir un fichero basta con arrastrar el archivo en formato .csv o .txt, al ‘área de trabajo’, aunque se puede hacer también desde el *menú 1* la opción *importar*. Si se desee introducir los datos manualmente se ha de seleccionar en el *menú 2* la opción *tabla → nuevo*. Este menú *tabla* puede resultar engañoso por su nombre dado que no se refiere a la representación de datos mediante tablas estadísticas ni de resumen, sino que se refiere a la presentación tabular de la recolección de los datos.

Figura 1

Ilustración de la vista de usuario de la plataforma CODAP



Fuente: The Concord Consortium (2024)

Manteniendo seleccionada la cabecera del conjunto de datos, se presenta el icono para *añadir una nueva variable* junto a otros dos nuevos elementos de navegación del software colapsados,

- *menú izquierda*, debajo del *menú 2*, que permite cambiar entre la vista de datos (que muestra el contenido de la base de datos organizada en forma de tabla) y la vista de variables (que proporciona información sobre las variables), y
- *menú derecha*, que permite desde *obtener información del conjunto de datos* hasta *importar, exportar, seleccionar y/o eliminar casos* y *crear nuevos atributos*.

Adicionalmente, pinchando en el nombre de cualquier variable aparece un menú contextual que permite modificar sus atributos mediante las opciones *renombrar*, *ajustar ancho al contenido*, *editar propiedades del atributo*, *editar fórmula*, *borrar fórmula*, *re-aleatorizar*, *orden ascendente* o *descendente*, *ocultar* y *eliminar atributo*, si bien algunas de ellas se muestran no disponibles.

Todos estos menús permiten un preprocesamiento de los datos adaptado al nivel cognitivo de la etapa educativa en la que su uso comienza a ser indicado, pudiendo ampliar las opciones disponibles para ello en el *menú 2* seleccionado *plugins* → *transformadores*. Dicha opción permite, entre otras cuestiones, *construir y transformar atributos*, *filtrar y ordenar datos*, *realizar medidas de recuento y centralización*, *agregar utilizando diferentes medidas* y reestructurar el fichero *agrupando o combinando casos*, *seleccionando atributos* o *uniendo conjuntos de datos*.

Para construir gráficos se selecciona *menú 2* → *gráfico* y se arrastra la variable a estudiar al eje X. Se muestra entonces un gráfico de 'puntos' que, en su *menú lateral derecho* → *opción 4ª* (gráfico), se puede *'agrupar en bins'* o mostrar *'barras para cada punto'*. Para la opción *'agrupar en bins'*, sin seleccionar la opción *'fusionar puntos en barras'*, el menú *'medida'* solo ofrecerá las opciones *'conteo'* y *'porcentaje'*. Para el resto de las opciones gráficas (incluida *'agrupado en bins'* → *fusionar puntos en barras*), en el menú *medida* se puede elegir entre las opciones *conteo*, *promedio* y *media*, *desvío estándar*, el gráfico de *cajas y bigotes* y *curva normal*, y la opción de *valores trazados y móviles*. Estas opciones permitirán la visualización

gráfica, no solo de los resúmenes numéricos tales como *mediana*, *cuartiles* o *promedio*, sino también el *gráfico de cajas y bigotes* o *la curva normal*.

Los gráficos creados son interactivos permitiendo modificar ejes, colores, etc., e incluso, simplemente pinchando en el nombre de la variable numérica arrastrada al *eje X* y seleccionando la opción '*tratar como categoría*', permite visualizarla en el gráfico como categórica (o viceversa -tratar como numérica una variable categórica, si está codificada con valores numéricos-). Este cambio conlleva las modificaciones correspondientes en los menús *gráfico* y *medida del menú derecha*.

CODAP destaca por hacer igualmente intuitivo el filtrado, agrupamiento y medidas de resumen de los datos. Esto es, para agrupar los datos por una variable basta arrastrar dicha variable a la izquierda (debajo del *menú izquierda*) y se ilustrará mediante un icono de control de expansión o colapso (ícono de "+" o "-") que los datos han sido agrupados. Cuando se clique sobre uno de estos valores agrupados, se despliega/colapsan todos los casos relacionados con él. Además, si se desea realizar medidas de resumen que tengan en cuenta dicha agrupación, bastará con crear la variable correspondiente en ese nivel.

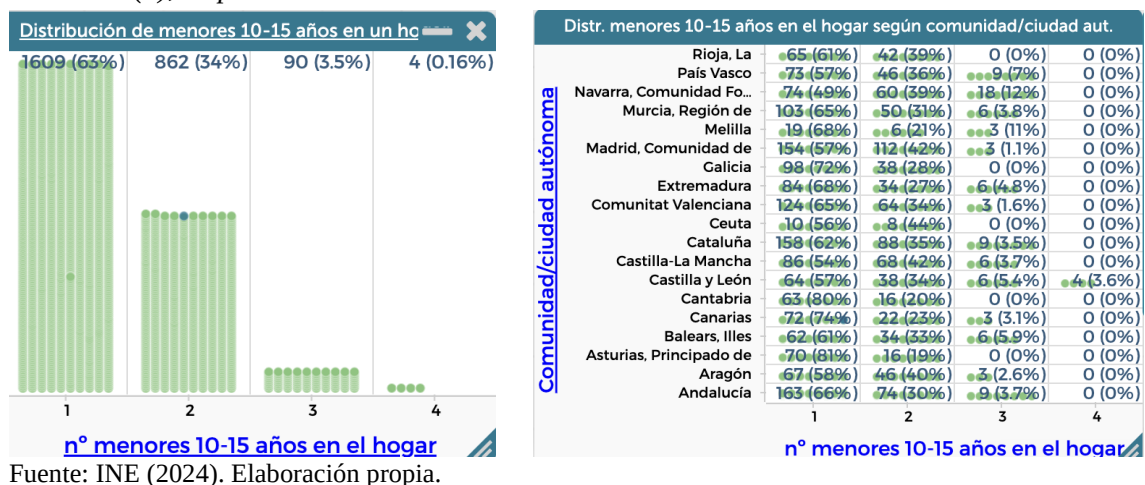
4.2. Exploración de datos

CODAP está diseñado para la visualización gráfica de datos por lo que, para obtener la representación mediante tablas, similar a la mostrada en el primer artículo, se requerirá de transnumeración de gráfico a tabla con los datos que se muestren en el gráfico de salida.

La Figura 2a muestra mediante un *diagrama de puntos* la distribución de frecuencias de la variable '*total menores de 10 a 15 años en el hogar*' en la Encuesta sobre Equipamiento y Uso de las TIC para el año 2023 (INE, 2024). Para ello, se selecciona *menú 2* → *gráfico*, se arrastra la variable al *eje X* y, pulsando en dicho eje sobre el nombre de la variable, se selecciona la opción de '*tratar como categórica*'. Si seleccionamos en *medida* → *conteo y porcentaje* dispondremos de la información necesaria para la transnumeración a la tabla de frecuencias correspondiente. Si además incorporamos al *eje Y* la variable *comunidad/ciudad autónoma* se puede obtener esta información para todas las regiones de forma muy visual (Figura 2b).

Figura 2

Distribución del total de menores de 10-15 años en el hogar en la Encuesta sobre Equipamiento y Uso de las TIC del año 2023, para España (a) y por comunidad/ciudad autónoma (b), respectivamente

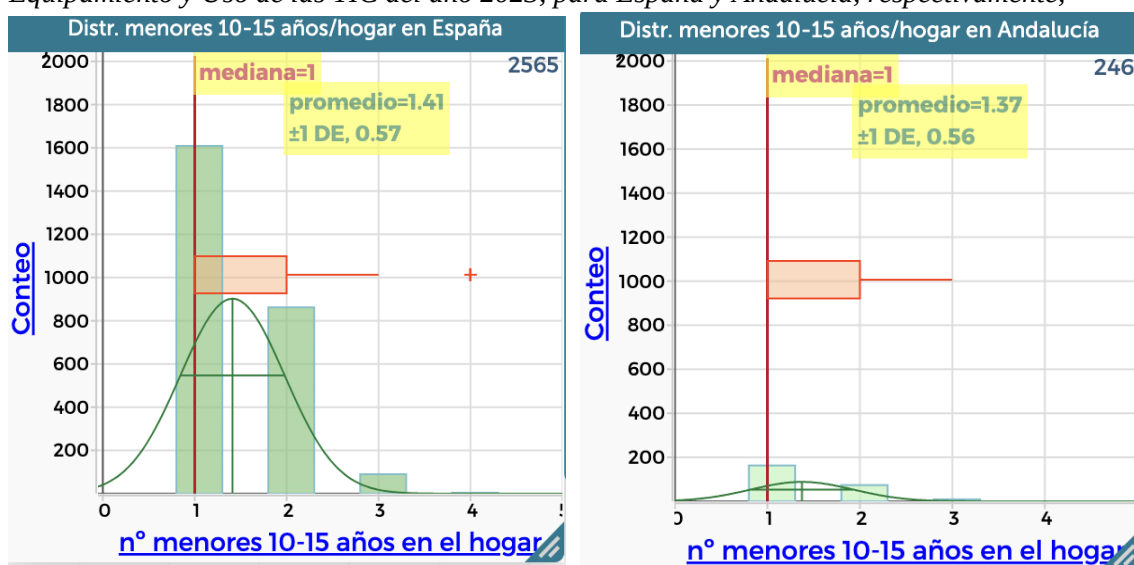


La distribución de la variable para España muestra el número de casos y porcentajes para cada valor de la variable. Además, si se hace click cualquier punto representado, por ejemplo en uno de los 4 casos que aparecen cuando el *nº de menores de 10-15 años en el hogar* es 4, nos lleva hasta el caso en el fichero de entrada, resaltándolo para nuestra inspección. De la distribución según comunidad/ciudad autónoma destaca que esos 4 casos pertenecen a la de Castilla y León o que Navarra es la comunidad con mayor porcentaje de hogares con tres miembros de ese grupo de edad en el mismo hogar.

Para explorar los resúmenes estadísticos tales como *mediana*, *cuartiles* o *promedio*, así como las representaciones del *gráfico de cajas y bigotes* o *la curva normal*, la variable tiene que estar tratada como numérica. Estas medidas se pueden obtener para cualesquiera de los gráficos que ofrece, seleccionando en este caso el *gráfico* → *agrupar en bins* → *fusionar puntos en barras* y las opciones mencionadas del menú *medida*. Esto permite igualmente la transnumeración a una tabla de resúmenes estadísticos para la variable de estudio.

Figura 3

Análisis de la variable ‘total de menores 10-15 años en el hogar’ en la Encuesta sobre Equipamiento y Uso de las TIC del año 2023, para España y Andalucía, respectivamente,



Fuente: INE (2024). Elaboración propia.

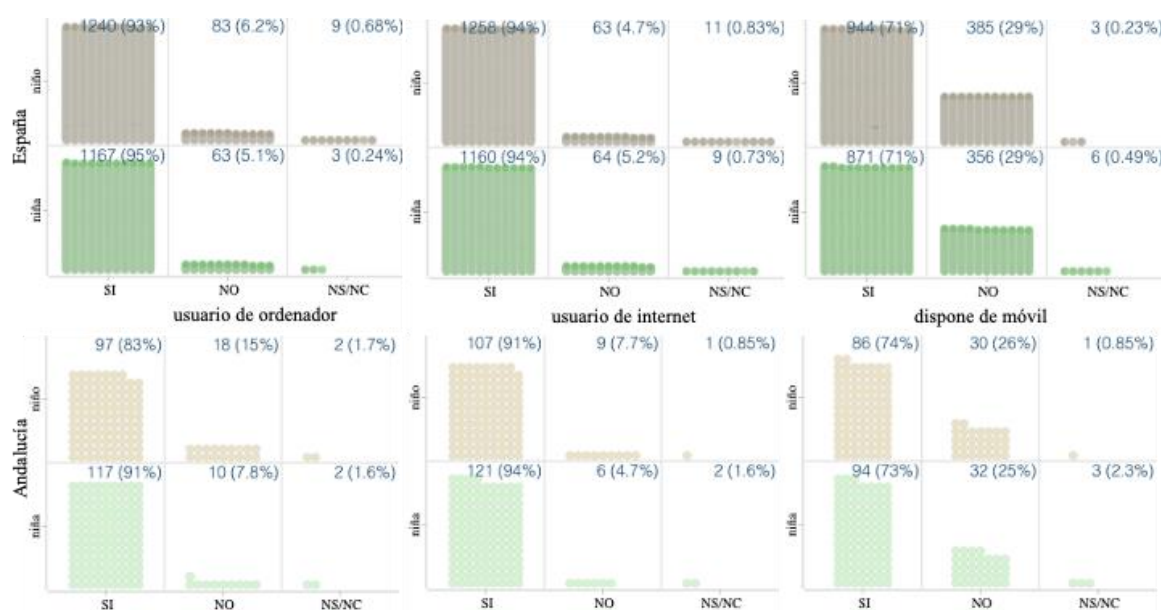
Nótese que se ha modificado el *eje Y*, simplemente pinchando sobre él y estirándolo, para hacer ambas representaciones comparables manteniendo la escala. La distribución en España y Andalucía tiene similares características, si bien en España se aprecia una asimetría mucho mayor a la derecha. Esto es, el 25% de los hogares tienen 2 o más miembros en ese grupo etario, comprendiendo a aquellos que tienen 4, mientras que en Andalucía ese 25% corresponde únicamente a hogares con 2 o 3 menores en ese rango de edad. Como se ha seleccionado la opción *mostrar outliers*, el valor para 4 miembros en el hogar en España se ilustra como tal en el gráfico de la izquierda ya que cumple con las características requeridas para ello. Además, se ha seleccionado la opción *mostrar etiquetas* por lo que se muestran los valores para la mediana y el promedio, junto con la desviación estándar. En este caso se había seleccionado la *mediana* como aparece representado en la línea roja sobre el valor 1.

Para explorar la relación entre dos variables bastaría con arrastrar una variable al *eje X* y la otra al *eje Y*, como se ha mostrado en la imagen derecha de la Figura 2. De este modo, es

posible también estudiar el uso de ordenador e internet en casa, así como la disponibilidad de teléfono móvil en menores de 10 a 15 años andaluces y españoles según cualesquiera de las variables sociodemográficas en la Encuesta sobre Equipamiento y Uso de las TIC para el año 2023 (Figura 4). Para este ejemplo las variables se han recodificado en los valores de sus etiquetas mediante la opción *plugins* → *transformadores*, tratándose siempre como cualitativas.

Figura 4

Uso de ordenador e internet en casa y disponibilidad de teléfono móvil en menores de 10 a 15 años andaluces y españoles según género en la Encuesta sobre Equipamiento y Uso de las TIC para el año 2023



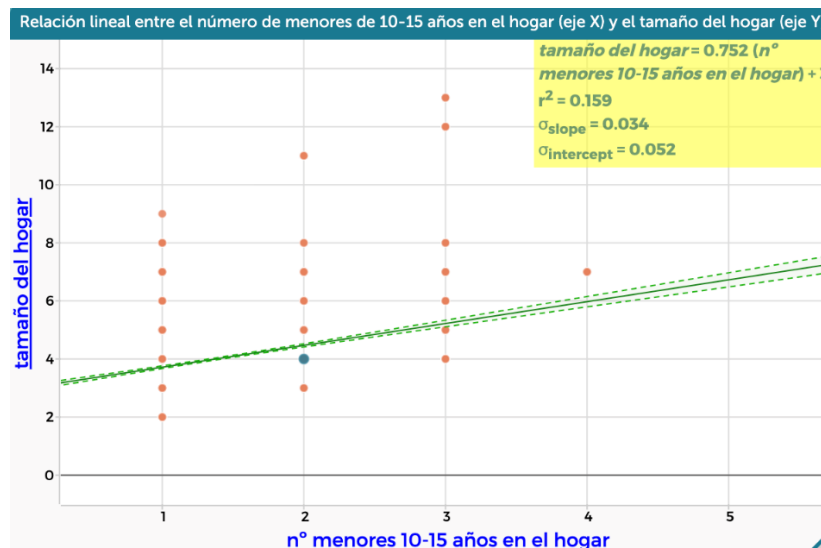
Fuente: INE (2024). Elaboración propia.

Nótese que la simplicidad del software redundaba en la sencillez de las representaciones que ofrece, haciendo la edición posterior más elaborada si se quiere presentar información multidimensional y no permitiendo, por ejemplo, en este caso, cambiar la escala en los gráficos para hacerlos comparables. Esto último podría conllevar errores de interpretación al ver punteadas áreas similares y no reparar en los valores numéricos.

Finalmente, la relación entre dos variables numéricas se puede visualizar mediante el gráfico de dispersión, mostrando la relación entre el *total de menores de 10-15 años en el hogar* (eje X) y el *tamaño del hogar* (eje Y) como se ilustra a continuación en la que se ha seleccionado la opción '*recta de mínimos cuadrados*'.

Figura 5

Gráfico de dispersión y recta de regresión lineal para el estudio de la relación entre el total de menores de 10-15 años en el hogar y el tamaño del hogar



Fuente: INE (2024). Elaboración propia.

Esta representación indica que hay una cierta relación positiva, es decir, a medida que aumenta el número de menores en el hogar, también tiende a aumentar el tamaño del hogar. Por cada menor adicional de 10-15 años en el hogar, el tamaño del hogar tiene un aumento moderado con un promedio de 0.752 personas, aunque el coeficiente de determinación $r^2 = 0.159$ nos indica que la relación es relativamente débil pues solo el 15,9% de la variabilidad del tamaño del hogar puede explicarse por el número de menores de 10-15 años en dicho hogar. En conclusión, hay una tendencia a que los hogares con más menores de 10-15 años sean más grandes, pero esta tendencia no es muy fuerte, ya que otros factores también afectan significativamente al tamaño del hogar.

5. CONCLUSIONES

En este estudio se describe el software CODAP, mostrando ejemplos de exploración de datos de la *Encuesta sobre Equipamiento y Uso de las TIC para el año 2023* (INE, 2024). La plataforma es en su mayor parte interactiva por lo que su uso resulta muy sencillo, permitiendo el movimiento de datos de forma natural y la visualización simultánea de varios objetos tales como tablas, gráficos, etc. Esto permite realizar principalmente representaciones gráficas de forma intuitiva mostrando las relaciones entre las variables de estudio, permitiendo discriminar y/o seleccionar casos solo con puntearlos sobre dicha representación y revelando un relato plausible para la información que recogen dichos datos. Por tanto, dado que no se requieren habilidades de programación, únicamente habilidades digitales básicas para su manejo, es un programa recomendado desde las primeras etapas educativas para la iniciación en la exploración de datos (Erikson, 2016; Mojica et al., 2019; The Concord Consortium, 2024, etc.). La única limitación al respecto la constituyen las habilidades cognitivas al respecto que el usuario tiene en cada momento de su aprendizaje

No obstante, su simplicidad redundante en la poca variedad de gráficos disponibles, así como en las limitaciones que esto conlleva respecto a la explotación de datos y la poca versatilidad de las imágenes que muestra, por ejemplo, en la realización de análisis multidimensionales. En este sentido se echa en falta la posibilidad de obtener algunas representaciones básicas como diagrama de sectores o histograma, la posibilidad de obtener directamente una tabla de frecuencias o de medidas de resumen, así como algunas medidas de resumen elementales como por ejemplo el rango. Por el contrario, permite algunas técnicas más avanzadas debido a su movimiento de datos interactivo (Erickson y Chen, 2021) como es el análisis multinivel (Haldar et al., 2018) así como técnicas básicas relacionadas con el aprendizaje automático, como puede ser la elaboración de árboles de clasificación mediante la herramienta Arbor (Engel et al., 2019; Erickson y Engel, 2023).

Por tanto, este instrumento permite la exploración de datos a cualquier ciudadano, independientemente de que luego se conviertan en usuarios de programas basados en código como R o Python que, si bien ofrecen innumerables métodos de exploración y visualización, también requieren de un aprendizaje del lenguaje de programación y mayor desarrollo del pensamiento estadístico y computacional. En este sentido, se considera que CODAP es una herramienta eficiente para la introducción a la exploración de datos. Esto se debe principalmente a la optimización en el tiempo de introducción al software, que hace posible utilizarlo de forma innata a los usuarios nativos digitales, su disponibilidad de forma gratuita y el buen rendimiento mostrado por los estudiantes en las investigaciones en torno a él. Esto último redundante en su consideración como recurso eficaz, pues dichos estudios han demostrado que contribuye a promover la adquisición de competencias en exploración de datos desde los niveles de iniciación junto con un aprendizaje de los conceptos necesarios de forma natural.

Finalmente, se considera que el uso de la plataforma presentada puede tener beneficios recíprocos. Por un lado, su utilización ayudará a estudiantes y profesores en el proceso de enseñanza-aprendizaje relacionado con la exploración de datos en cualquier ámbito de conocimiento y por otro, al ser de código abierto, su uso por la comunidad educativa puede redundar en la completitud del programa adaptándose a las necesidades de los usuarios que lo utilizan.

REFERENCIAS BIBLIOGRÁFICAS

- Budde, L., Frischemeier, D., Biehler, R., Fleischer, Y., Gerstenberger, D., Podworny, S. y Schulte, C. (2020). Data Science Education in Secondary School: How to Develop Statistical Reasoning When Exploring Data Using CODAP. En *IASE Roundtable*.
- Engel, J., Erickson, T. y Martignon, L. (2019). Teaching about decision trees for classification problems. En *IASE Satellite Meeting*.
- Erikson, T. (2016). Practical public online data: introducing Tuva and CODAP. En *IASE Roundtable Workshop*
- Erickson, T. y Chen, E. (2021). Introducing data science with data moves and CODAP. *Teaching Statistic*, 43, S124-S132.
- Erickson, T. y Engel, J. (2023). What goes before the CART? Introducing classification trees with Arbor and CODAP. *Teaching Statistic*, 45, S104-S113.
- Frischemeier, D., Biehler, R., Podworny, S. y Budde, L. (2021). A first introduction to data science education in secondary schools: Teaching and learning about data exploration with CODAP using survey data. *Teaching Statistic*, 43, S182-S189.

- Haldar, L. C., Wong, N., Heller, J. I. y Konold, C. (2018). Students making sense of multi-level data. *Technology Innovations in Statistics Education*, 11(1).
- Instituto Nacional de Estadística [INE] (2024). *Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares, año 2023*. <https://tinyurl.com/2bq2py8r>
- Mojica, G. F., Azmy, C. N. y Lee, H. S. (2019). Exploring data with CODAP. *The Mathematics Teacher*, 112(6), 473-476. <https://doi.org/10.5951/mathteacher.112.6.0473>
- Molina-Portillo, E., Contreras, J.M., Molina-Muñoz, D. y Cabrera, R. (2024). Explorando la visualización de datos. *Epsilon*, 118, 9-20.
- Peters, S. A., Bargagliotti, A. y Franklin, C. (2024). Engaging and preparing educators to teach statistics and data science. *The AMTE Handbook of Mathematics Teacher Education: Reflection on Past, Present and Future—Paving the Way for the Future of Mathematics Teacher Education*, Volume 5, 123.
- Poling, L. y Weiland, T. (2020). Using an interactive platform to recognize the intersection of social and spatial inequalities. *Teaching Statistics*, 42(3), 108-116.
- Shreiner, T. L. y Guzdial, M. (2022). The information won't just sink in: Helping teachers provide technology-assisted data literacy instruction in social studies. *British Journal of Educational Technology*, 53(5), 1134-1158.
- The Concord Consortium, 2024. *Common Online Data Analysis Platform, CODAP*. <https://codap.concord.org/>
- Tuva Labs Inc. (2024). *Tuva: Herramienta de análisis de datos. Tuva*. <https://tuvalabs.com>

Visualización de datos con Tuva

Elena Molina Portillo
 Universidad de Granada, elemo@ugr.es

Carlos Santiago Sánchez Muñoz
 Universidad de Granada, carlossamu7@correo.ugr.es

Resumen: La visualización gráfica de datos se presenta como un enfoque accesible y eficaz para desarrollar las habilidades necesarias para interpretar y comprender información basada en datos. En este estudio se presenta Tuva como una herramienta destinada al profesorado interesado en integrar la visualización de datos y el análisis estadístico en sus prácticas pedagógicas. Como ejemplo se utiliza la representación gráfica de la información proveniente de la Encuesta sobre Equipamiento y Uso de las TIC 2023 (INE, 2024). Este recurso digital interactivo se posiciona como una plataforma educativa innovadora para el desarrollo de la alfabetización informacional y digital necesaria en el contexto actual.

Palabras clave: Visualización de datos, alfabetización digital, análisis estadístico, educación interactiva tecnologías educativas, Tuva.

Data visualization with Tuva

Abstract: Data visualization is an accessible and effective approach to developing the skills required to interpret and understand data-driven information. In this study, Tuva is presented as a tool for teachers interested in integrating data visualization and statistical analysis into their pedagogical practices. As an example, we use the graphical representation of information from the Survey on ICT Equipment and Use 2023 (INE, 2024). This interactive digital resource provides an innovative educational platform for developing the information and digital literacy necessary in the current context.

Key words: Data visualization, digital literacy, statistical analysis, interactive education, educational technologies, Tuva.

1. INTRODUCCIÓN

La visualización de datos es una destreza fundamental en la alfabetización en información y datos, por tanto, es una habilidad requerida para el desarrollo de la competencia digital de los ciudadanos. Asimismo, la visualización de datos a través de representaciones gráficas se reconoce como un enfoque accesible y eficaz para introducir las habilidades esenciales necesarias para interpretar y comprender la información basada en datos.

Para promover desde etapas tempranas la capacidad de interpretar y comunicar dicha información, el software Tuva (Tuva Labs Inc., 2024a) se presenta como una herramienta innovadora que facilita el desarrollo de competencias analíticas y comunicativas mediante la exploración interactiva de datos. Esta plataforma educativa, diseñada para fomentar la alfabetización en datos y el pensamiento crítico, integra un conjunto de herramientas de última generación que son interactivas, gratuitas y fáciles de usar, permitiendo explorar, analizar e

interpretar datos de forma intuitiva. Todo ello contribuye a un aprendizaje significativo y accesible para diversas capacidades cognitivas, basado en la simplificación de la creación de gráficos y estadísticas.

La versión gratuita del software, denominada *Tuva Basic*, permite acceso a la biblioteca de datos de contenidos matemáticos, contenidos científicos y a conjuntos de datos Tuva. De igual manera, tiene la posibilidad de cargar conjuntos de datos propios desde Google Drive, OneDrive u otro dispositivo, además de ofrecer acceso a datos y análisis a nivel de clase y de estudiante. Todo ello está limitado a un número de 10 recursos didácticos de Matemáticas y 10 de Ciencias de forma gratuita para utilizarlo con los estudiantes, así como a la visualización de 5 conjuntos de datos de ejemplo de Tuva y la carga de 5 conjuntos de datos propios. Esta opción se considera suficiente para permitir a los estudiantes aplicar conceptos estadísticos y matemáticos en contextos reales, proporcionando visualizaciones dinámicas y personalizables, tales como diagramas de dispersión, gráficos de barras y tablas, facilitando la comprensión de patrones y tendencias.

Las versiones de suscripción de Tuva, *Tuva Jr.* y *Tuva Premium* aumentan las cantidades indicadas para las anteriores opciones, permitiendo además el acceso a soluciones y guías para docentes, compartir conjuntos de datos entre clases para el aprendizaje colaborativo, sincronización con Google Classroom, Schoology, Clever y Canvas, acceso al generador de actividades para crear lecciones propias de Tuva, acceso a datos y análisis a nivel escolar y acceso a datos y análisis a nivel de distrito. Ambas versiones tienen suscripción anual disponible para profesores y a nivel escuela, si bien *Tuva Premium* aumenta las cantidades en todas las opciones indicadas para *Tuva Jr.*, dando acceso completo a todas las colecciones de Tuva (Tuva Labs Inc., 2024b).

En la información relacionada con las opciones de suscripción en Tuva, destaca la disponibilidad de suscripción a evaluaciones de diagnóstico de alfabetización en datos, así como evaluaciones científicas para los diferentes niveles educativos de los estándares de Estados Unidos.

Adicionalmente, Tuva ofrece una amplia gama de desarrollo profesional para el docente, permitiendo la asistencia a seminarios web o las denominadas ‘horas de oficina’ destinadas a aprender sobre temas específicos, hacer preguntas y reunirse con miembros del equipo de Tuva. También permite trabajar de forma asincrónica revisando seminarios web anteriores, descargando material para estudiantes, obteniendo microcredenciales de Tuva y revisando vídeos que muestran la incorporación de Tuva en sesiones de clase. Por tanto, se presume que esta herramienta puede resultar esencial para docentes interesados en integrar la visualización de datos y el análisis estadístico en sus prácticas pedagógicas.

En línea con lo descrito en el primer trabajo del monográfico, este estudio explora la visualización de datos utilizando *Tuva Basic*.

2. ANTECEDENTES

Tuva Labs, fundada en 2012, es una plataforma educativa diseñada para promover la alfabetización en datos y el pensamiento crítico desde etapas tempranas, ofreciendo soluciones para Primaria y Secundaria. Su propuesta se centra en herramientas interactivas que permiten a estudiantes y docentes explorar, visualizar y analizar datos reales, fomentando la comprensión de patrones y tendencias de manera intuitiva y significativa (Tuva Labs, 2015). Estas son

habilidades esenciales en el desarrollo como ciudadano, así como en las áreas STEM y en el mercado laboral actual.

La alfabetización en datos es una competencia fundamental en el siglo XXI. Thompson (2023) resalta la importancia de diseñar tecnologías web que desarrollen habilidades de visualización de datos, mientras que Erwin (2015) subraya que este tipo de alfabetización permite un aprendizaje basado en problemas reales, fortaleciendo competencias analíticas. Adisa et al. (2024) también destacan que involucrar a estudiantes de Primaria en prácticas de ciencia de datos mediante plataformas como Tuva, mejora significativamente su comprensión de conceptos básicos de análisis de datos.

El uso de Tuva Labs en el aula se alinea con metodologías como el aprendizaje basado en proyectos, promoviendo la participación activa de los estudiantes en actividades relacionadas con datos (Boss, 2014). En la misma línea, se enfatiza el papel de las actividades de alfabetización en datos para integrar conceptos estadísticos y científicos en el aula. Así, actividades como las de ciencia ciudadana, destacadas por Plummer y Van Dis (2019), permiten a los estudiantes liderar proyectos donde recopilan, analizan e interpretan datos reales, fomentando tanto el aprendizaje interactivo como la resolución de problemas.

Erwin (2015) destaca que, a pesar de sus ventajas, Tuva Labs enfrenta retos como la necesidad de adaptarse a contextos educativos internacionales y la falta de estudios longitudinales que evalúen su impacto a largo plazo. No obstante, su enfoque innovador y la capacidad de personalización para diversas necesidades pedagógicas la posicionan como una herramienta valiosa para el desarrollo de competencias esenciales en educación.

3. METODOLOGÍA

Para explorar la visualización de datos mediante la plataforma Tuva, se presenta un estudio exploratorio-descriptivo tomando como tópico de interés el estudio real sobre el uso de las tecnologías de la información y la comunicación (TIC) en niños de 10 a 15 años, a partir de la Encuesta sobre Equipamiento y Uso de las TIC para el año 2023 (Instituto Nacional de Estadística [INE], 2024). Para ello se consideran las características metodológicas descritas en el estudio “*Explorando la visualización de datos*” (Molina-Portillo et al., 2024) en el que se presenta la base de datos considerada, sus variables y selección de casos. Por tanto, se focaliza ahora en describir el análisis que se va a presentar mediante este software.

En primer lugar, se presenta un acercamiento al software describiendo sus características más relevantes, así como las posibilidades que ofrece utilizando cualquiera de los conjuntos de datos a los que dan acceso de forma gratuita. Posteriormente, abordaremos el estudio del uso de TIC en menores de 10-15 años mediante diversas representaciones gráficas que ofrece el programa en estudio. En tal sentido, se realizará un histograma para la variable *edad*, que si bien esta codificada como cuantitativa discreta, en este caso se tratará como una variable cuantitativa continua, en base a su naturaleza. Para ilustrar la representación de varias variables, al histograma anterior se le incorporará la variable *disponibilidad de teléfono móvil*, obteniendo información de a qué edades de las estudiadas es más habitual tener este dispositivo en propiedad. Posteriormente, se presenta un gráfico de sectores para las *provincias* de la comunidad autónoma de Andalucía, al cual también se le realiza la incorporación posterior de la variable *uso de ordenador en los tres últimos meses*.

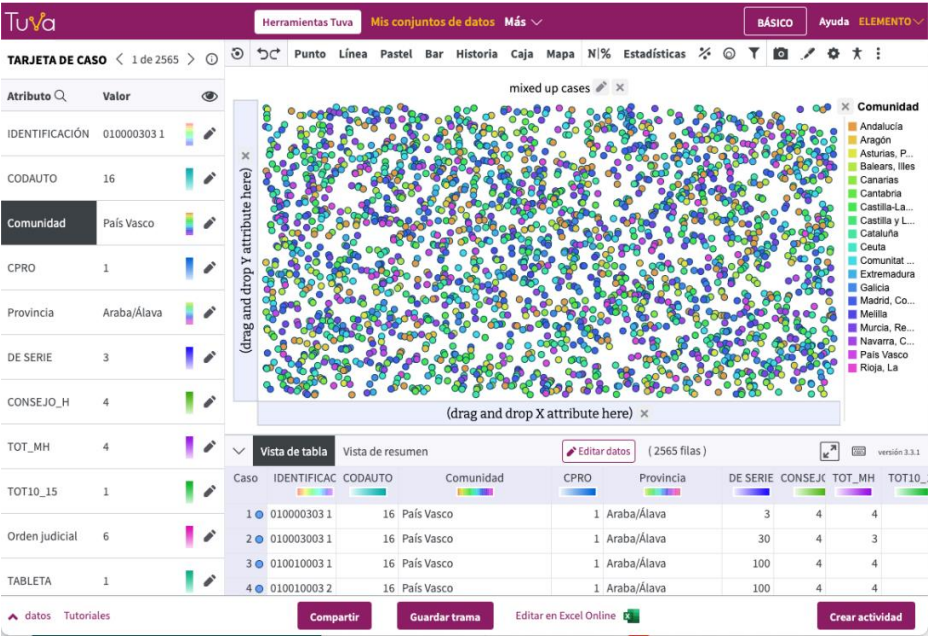
4. RESULTADOS

4.1. Primeros pasos

Al acceder a la plataforma Tuva (Tuva Labs Inc., 2024a) se ofrece un acceso a la *Introducción a Tuva Tools*, así como información, enlaces y pruebas de lecciones a *Tuva Jr. para los grados 2 a 5* (8-11 años), *Matemáticas Tuva para los grados 6 a 12* (12-18 años) y *Ciencias de Tuva para los grados 6 a 12* (12-18 años). Si accedemos a la *Introducción a Tuva Tools* se muestran conjuntos de datos de muestra, los cuales se pueden utilizar para trabajar la alfabetización de datos en clase, así como las opciones *Insertar datos* e *Importar datos*.

Al entrar en cualquier conjunto de datos se muestra una sección que permite nuevamente desplazarnos por los *conjuntos de datos de muestra*, *insertar datos* e *importar datos*. Al abrir un conjunto de datos (Figura 1) se muestra una tabla a la izquierda. En ella se encuentran los valores para todas las variables de uno de los casos, indicando arriba el número total de casos y el número del caso cuya información se está mostrando. Asimismo, ofrece la posibilidad de desplazarse con los cursores laterales para mostrar los valores de las variables de otro registro. A su lado, si pinchamos el botón de información, nos proporcionará el título de la base de datos, el número total de casos y el número de atributos que contiene, cuántos de ellos son categóricos y cuántos continuos, indicando los atributos que pertenecen a cada una de estas tipologías. En la tabla descrita se pueden observar tres columnas: la primera, destinada a ver el nombre de los atributos con los que se trabaja (junto a la cual aparece un buscador que permite filtrar y mostrar el valor de ese único atributo); la segunda, muestra el valor de dichos atributos para el caso indicado arriba; y, la tercera columna contiene un botón de visualización que nos permite cambiar esta tabla por el listado de atributos, para poder deseleccionar aquellos con los que no queremos trabajar.

Figura 1
Vista de la plataforma Tuva con la base de datos sobre el uso de las TIC para el año 2023 (INE, 2024) seleccionando la opción ‘vista de tabla’ y la variable Comunidad Autónoma.



Fuente: Tuva Labs. Inc. (2024) e INE (2024). Elaboración propia.

En la parte central se encuentra una sección para mostrar las gráficas que vayamos realizando. En los ejes se indica *arrastrar y soltar* la variable correspondiente que se quiera representar en cada uno, facilitando la visualización gráfica sin necesidad de comandos complejos y haciéndolo visualmente atractivo y versátil. Si clicamos sobre cualquiera de las variables de la tabla de la izquierda también cambiará esta región central, representando mediante puntos de diferentes colores los valores de dicha variable. Estos colores junto a los valores que representan se indican mediante una tabla a la derecha. Si se selecciona cualquiera de los puntos representados se mostrará en la tabla de la izquierda los valores para todos los atributos de ese caso. Si se selecciona cualquiera de los valores de la variable que se muestran a la derecha, se remarcarán en negrita los puntos que tengan asociados ese valor.

Adicionalmente, en la parte central encontramos dos barras de menú. La que se ubica abajo permite dividir esta sección central mostrando la *vista de tabla* o *Expandir/Contraer* la vista de tabla hasta que ocupe toda la sección central. Mediante la opción *vista de resumen* se divide también esta sección central mostrándose un mensaje para que agreguemos una función desde la barra de herramientas. Y mediante la opción representada por un *teclado* se muestran todos los atajos de teclado disponibles para trabajar en la plataforma. En la barra de herramientas que se ubica arriba tenemos las opciones de *borrar y restablecer*, *deshacer y rehacer*, los distintos tipos de gráficos disponibles (*punto, línea, pastel, barras, histograma, cajas y mapa*, - aunque las opciones aparecen un poco desconfiguradas en los menús en la versión española del software, se intuyen fácilmente), así como el recuento de datos mediante su *número, porcentajes* y la opción de realizar determinadas *estadísticas básicas* o conseguir datos de un *modelo*. Finalmente, en las opciones más a la derecha se permite hacer una *selección de muestra* con distintas opciones, *filtrar datos* (manteniéndolos o eliminándolos), hacer *capturas de pantalla*, *editar* la pantalla, así como ciertas opciones de *configuración, accesibilidad, idioma y ayuda*.

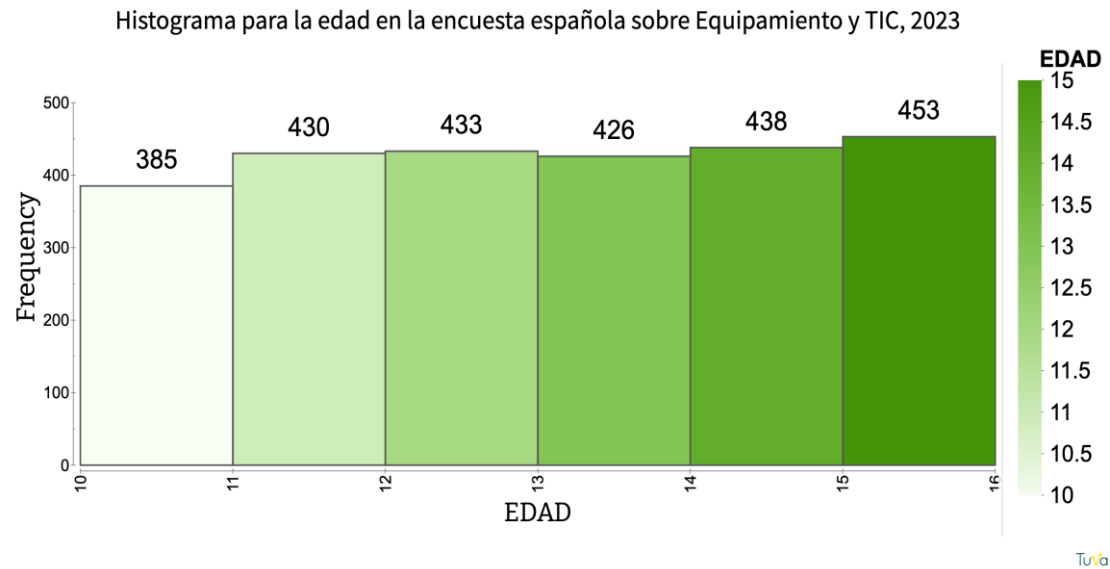
El acceso a tutoriales, a las bibliotecas y a las diferentes herramientas que ofrece Tuva también está disponible desde esta pantalla de inicio en los menús inferior y superior que se muestran en la pantalla.

4.2. Visualización de datos

Para obtener la información relativa a las tablas de frecuencias, indicadas desde la etapa de Primaria, similares a las presentadas en el trabajo introductorio bastaría con arrastrar la variable de estudio al eje X y seleccionar $N\% > Contar > Contar casos$ y $N\% > Por ciento > Porcentaje de casos$. Si la variable de estudio es cuantitativa podríamos obtener además los resúmenes estadísticos seleccionando en la opción *Estadísticas* las medidas deseadas.

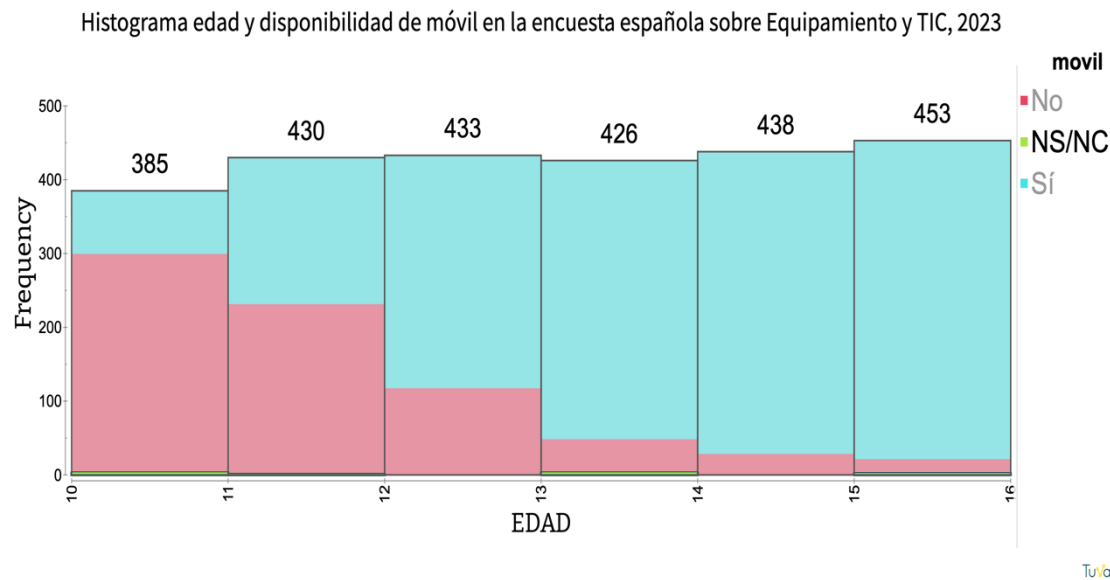
Centrándonos en la visualización gráfica, como se ha comentado previamente, se pueden realizar *gráficos de puntos, líneas, gráfico de sectores, diagrama de barras, histogramas, diagramas de cajas y bigotes*, así como *mapas geográficos*. A modo de ejemplo se presentan dos histogramas, el primero corresponde a la distribución de la variable *edad* en la muestra analizada (Figura 2), mientras que en el segundo gráfico se añade la información de la *disponibilidad de teléfono móvil* a la representación anterior (Figura 3). La primera representación permite observar que el número de participantes según la edad se distribuye de forma cuasi-equitativa, teniendo la máxima representación en la muestra los menores que tienen 15 años y la menor representación corresponde a aquellos con 10 años.

Figura 2
Distribución por edad de los participantes en la Encuesta sobre Equipamiento y Uso de las TIC del año 2023, para España



Fuente: Encuesta sobre Equipamiento y Uso de TIC (año 2023). INE (2024a). Elaboración propia

Figura 3
Distribución por edad y disponibilidad de teléfono móvil de los participantes en la Encuesta sobre Equipamiento y Uso de las TIC del año 2023, para España



Fuente: Encuesta sobre Equipamiento y Uso de TIC (año 2023). INE (2024a). Elaboración propia

Respecto al gráfico mostrado para estudiar la distribución *por edad y disponibilidad de teléfono móvil*, se puede observar que el número de menores que tiene un móvil en propiedad aumenta con la edad, si bien a los 10 años hay un número de ellos que ya lo tienen, siendo la edad de 12 años en la que, por primera vez, es mayor la parte de participantes que disponen de

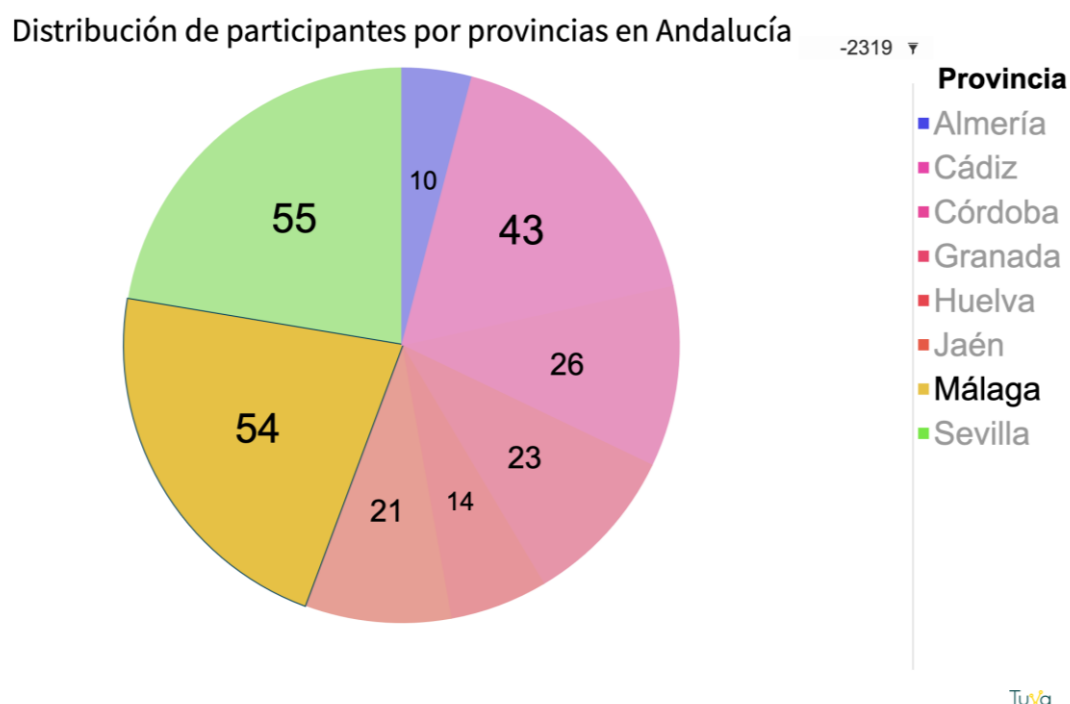
éste, que los que no tienen esa propiedad, observándose esta característica para todas las edades siguientes.

Nótese que en ambos casos se ha modificado el título del gráfico en el propio programa accediendo a la herramienta *edición de título* (situado junto al título por defecto), así como los colores de las figuras, seleccionando la herramienta de *edición de variable* que aparece en la tabla izquierda al lado de los nombres de las variables. Adicionalmente, seleccionando la variable en el menú de la izquierda se pueden seleccionar y deseleccionar categorías, actualizándose la información mostrada a la selección realizada.

A continuación, en la herramienta representada por un embudo de *filtrado > filtrar por Comunidad*, se selecciona la comunidad autónoma de Andalucía. Si aún tenemos en pantalla el gráfico realizado anteriormente mostrará dicho gráfico con la información seleccionada sin necesidad de repetirlo. Para obtener un nuevo gráfico deseado, eliminamos las variables estudiadas anteriormente de los ejes, seleccionamos el *diagrama de sectores* para el número de participantes por provincia de la región estudiada seleccionando la variable *provincia* (Figura 4).

Figura 4

Gráfico circular (o de pastel) que representa la distribución de participantes por provincias en Andalucía.



Fuente: Encuesta sobre Equipamiento y Uso de TIC (año 2023). INE (2024a). Elaboración propia

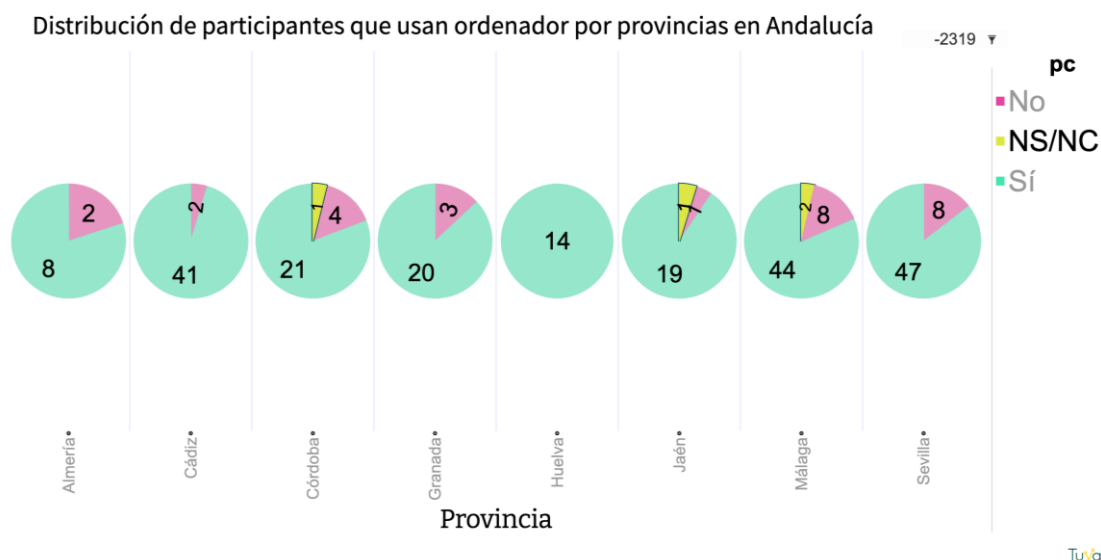
Se observa que la mayor representación corresponde a los participantes sevillanos, malagueños y gaditanos, mientras que los casos provenientes de Almería y Huelva son los menos numerosos.

Si adicionalmente arrastramos la variable *provincia* al eje X, y pinchamos sobre la variable *PC (uso de ordenador en los 3 últimos meses)*, se mostrará un gráfico de sectores para cada una

de las provincias que mostrará la distribución de participantes que han sido usuarios de ordenador en los tres últimos meses en cada una de las provincias (Figura 5).

Figura 5

Distribución de participantes que han usado el ordenador en los 3 últimos meses por provincias en Andalucía.



Fuente: *Encuesta sobre Equipamiento y Uso de TIC (año 2023)*. INE (2024a). Elaboración propia

La plataforma Tuva ofrece una amplia variedad de opciones gráficas, cada una de ellas acompañada de múltiples configuraciones personalizables. Los ejemplos presentados ilustran tan solo una parte del potencial de esta herramienta, destacando su versatilidad y capacidad para adaptarse a diferentes necesidades.

5. CONCLUSIONES

Tuva Labs se posiciona como una plataforma educativa innovadora que aborda de manera efectiva las necesidades emergentes de alfabetización en datos y pensamiento crítico en el contexto educativo. Al proporcionar herramientas interactivas y adaptables, esta plataforma facilita tanto la enseñanza como el aprendizaje de conceptos clave en el análisis de datos, integrándolos en actividades relacionadas con problemas del mundo real (Tuva Labs, 2015). De esta manera, se alinea con enfoques modernos de aprendizaje activo, como el aprendizaje basado en proyectos y la ciencia ciudadana, promoviendo la colaboración y el desarrollo de habilidades críticas (Boss, 2014; Plummer y Van Dis, 2019).

Una de las principales limitaciones del estudio que se presenta ha sido la imposibilidad de acceder a determinadas prestaciones que solo se ofrecen con las versiones de suscripción *Tuva Jr.* Y *Tuva Premium*. Se considera de gran interés explorar las funcionalidades de sincronización con plataformas educativas como Google Classroom y Schoology, así como el acceso a las herramientas del generador de actividades personalizadas. Adicionalmente, en línea con el tópico de interés de este estudio habría resultado clave la consulta de las evaluaciones diagnósticas de alfabetización en datos, ya que constituyen una herramienta fundamental para

identificar el nivel de competencia de los estudiantes en la comprensión, interpretación y uso crítico de la información cuantitativa y gráfica. Esta información sobre las fortalezas y áreas de mejora del alumnado facilita a los docentes la personalización de la enseñanza, resultando clave para diseñar intervenciones didácticas precisas que promuevan de forma progresiva tanto la alfabetización en datos como el pensamiento crítico.

En resumen, Tuva Labs posee un potencial significativo para transformar la enseñanza de la alfabetización en datos. Su capacidad para hacer uso de datos reales en experiencias educativas significativas desde la etapa de Primaria la convierte en una herramienta indispensable para preparar a los estudiantes frente a los retos del siglo XXI.

REFERENCIAS BIBLIOGRÁFICAS

- Adisa, I.O., Herro, D., Abimbade, O. y Arastoopour Irgens, G. (2024). Engaging elementary students in data science practices. *Information and Learning Sciences*, 125(7/8), 513-544.
- Boss, S. (2014). *Implementing project-based learning*. Solution Tree.
- Erwin, R.W. (2015). Data Literacy: real-world learning through problem-solving with data sets. *American Secondary Education*, 43(2), 18-26.
- Molina-Portillo, E., Contreras, J.M., Molina-Muñoz, D. y Cabrera, R. (2024). Explorando la visualización de datos. *Epsilon* 118, 9-20.
- Plummer, A. y Van Dis, J. (2019). Student-driven citizen science. *Science Scope*, 43(3), 65-74.
- Thompson, J. (2023). *Designing web-based technologies for data visualization skill development*. Available from ProQuest Dissertations y Theses Global; ProQuest One Academic. <https://www.proquest.com/dissertations-theses/designing-web-based-technologies-data/docview/2820814554/se-2>
- Tuva Labs (2015). Tuva releases Tuva Premium giving educators interactive, graphical tools and content to teach data literacy. *Business Wire*. <https://www.businesswire.com/news/home/20151005006498/en/Tuva-Releases-Tuva-Premium-Giving-Educators-Interactive-Graphical-Tools-and-Content-to-Teach-Data-Literacy>
- Tuva Labs Inc. (2024a). *Tuva: Herramienta de análisis de datos*. Tuva. <https://tuvalabs.com>
- Tuva Labs Inc. (2024b). *Tuva para profesores, escuelas o distritos*. <https://tuvalabs.com/pricing>

Visualización gráfica con jamovi

David Molina-Muñoz

Departamento de Didáctica de la Matemática, Universidad de Granada,
 dmolinam@ugr.es

Sara Venegas-Sepúlveda

Profesora de Educación General Básica (mención en Matemática) por la Universidad
 del Bío-Bío (Chile), saravenegas@correo.ugr.es

Resumen: Jamovi es un programa gratuito, basado en el software estadístico R, para el análisis y la representación gráfica de datos, muy popular en el ámbito educativo. Con este trabajo se pretende dar a conocer el programa y mostrar las posibilidades que ofrece para la visualización de datos. Para ello, se ilustra cómo utilizar jamovi para crear gráficos de barras y de violín empleando los datos de la edición de 2023 de la Encuesta sobre Equipamiento y Uso de las TIC en los Hogares. Considerando sus beneficios, se recomienda el uso de jamovi en las aulas, especialmente en las de educación secundaria y educación superior.

Palabras clave: jamovi, visualización, software, gráficos

Graphical visualization with jamovi

Abstract: Jamovi is a free program, based on the statistical software R, for the analysis and graphical representation of data, very popular in the educational field. The aim of this paper is to introduce the program and to show the possibilities it offers for data visualization. For this purpose, we illustrate how to use jamovi to create bar charts and violin plots using data from the 2023 edition of the ICT Household Equipment and Use Survey. Considering its benefits, jamovi is recommended for use in classrooms, especially in secondary and university education.

Key words: jamovi, visualization, software, charts.

1. INTRODUCCIÓN

La visualización de datos se refiere a la representación de la información en formato gráfico a través de herramientas visuales tales como tablas, gráficos, mapas, infografías o diagramas. En la sociedad actual, en la que abundan los grandes volúmenes de datos, la visualización desempeña un papel fundamental, ya que facilita la representación, la interpretación y la comunicación de esta información en contextos tan dispares como el económico, el científico, el educativo o el de la salud. Por ello, es fundamental trabajar la visualización de datos en las aulas desde edades tempranas. Para atender esta necesidad, se han desarrollado distintas herramientas tecnológicas que posibilitan la representación gráfica de la información. Entre ellas se encuentra jamovi (The jamovi project, 2024a), que es un programa para el análisis de datos creado por Jonathon Love, Damian Dropmann y Ravi Selker. Tal y como se indica en la página web del programa, jamovi se presenta como una alternativa atractiva a otros productos de pago para el

análisis estadístico (The jamovi project, 2024b). Más concretamente, jamovi ofrece una interfaz gráfica de usuario para el *software* estadístico R, pero con un diseño más moderno que el de otras opciones como R Commander (Abbasnasab Sardareh et al., 2021). En efecto, jamovi proporciona un entorno visual amigable, intuitivo y sencillo de utilizar que permite hacer uso de R sin necesidad de conocer el lenguaje de programación propio de R. Dicho de otra forma, jamovi se encarga de enmascarar a R de cara al usuario. De este modo, cada vez que el usuario realiza un análisis en jamovi, en realidad se ejecuta código de R que es el que lleva a cabo los cálculos correspondientes, aunque el usuario permanece completamente ajeno a este proceso (Berrendero, 2021). Pese a todo ello, existe la posibilidad de ejecutar código de R en jamovi, por lo que un usuario con el conocimiento adecuado sobre el lenguaje de programación de R podría llevar a cabo análisis personalizados sin salir de jamovi con poco esfuerzo.

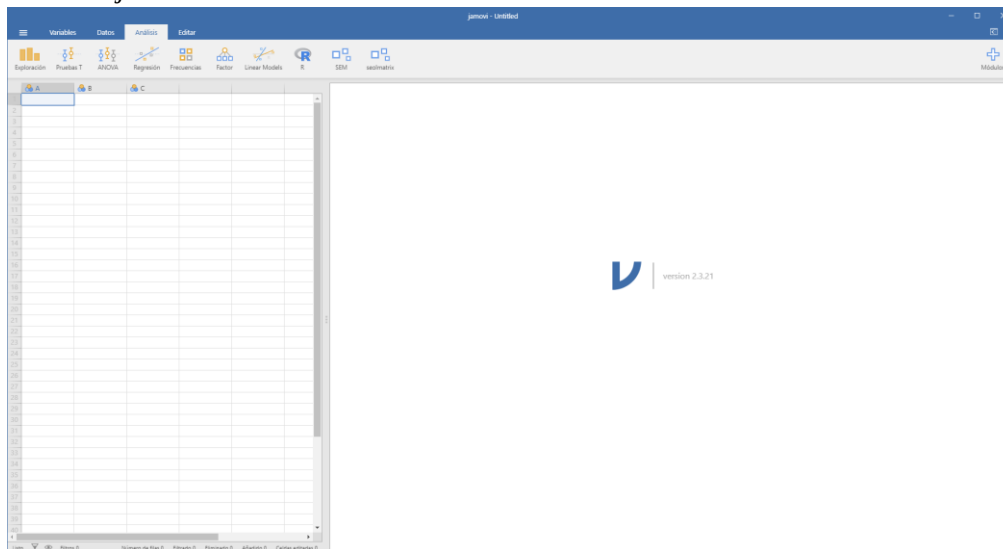
Las características de jamovi lo convierten en una herramienta accesible, versátil y potente para el tratamiento de datos y el análisis estadístico. De ellas, una de las más destacadas es el carácter gratuito y abierto del programa, de manera que cualquier persona puede descargarlo sin costo desde su página web y también puede contribuir a su mejora. Además, jamovi es un *software* multiplataforma que puede utilizarse en los principales sistemas operativos (Windows, Linux, MacOS y ChromeOS) e, incluso, puede ejecutarse en la nube sin necesidad de descargarlo, previo registro. Otra de las características de jamovi, la cual facilita sobremanera la migración de usuarios desde otros programas, es la eficacia en la gestión de ficheros externos, ya que permite importar archivos de datos de programas como SPSS, Stata, SAS, R, Excel o Libre Office, entre otros, de una manera muy simple. A la hora de analizar los datos, la instalación estándar de jamovi ofrece una serie de funcionalidades básicas que, no obstante, pueden ampliarse fácilmente gracias a la estructura modular que presenta el programa. Debido a esta configuración, pueden instalarse en jamovi hasta más de 40 módulos adicionales (similares a los paquetes o librerías de R) (The jamovi project, 2024c) que proporcionan pruebas estadísticas referidas a áreas específicas. Por último, es importante indicar que en jamovi los análisis se ejecutan en tiempo real y, en consecuencia, los resultados se actualizan automáticamente tras cualquier cambio (como, por ejemplo, la introducción de nuevos datos o la selección de nuevas opciones en un análisis). Con ello se consigue que cada instancia de análisis tenga asociada un conjunto de resultados único, evitando así duplicidades y confusiones (Åžahin y Aybek, 2020; Elosua y Egaña, 2020).

La apariencia de jamovi (Figura 1) es muy similar a la de otros programas dedicados al análisis de datos: en la parte izquierda de la pantalla principal hay una rejilla destinada a los datos; el espacio de la parte derecha está reservado para mostrar los resultados de los análisis que se realizan y en la parte superior se ubica la barra de menú, con las distintas opciones de análisis que ofrece el programa.

Con este trabajo se pretende mostrar las potencialidades y las limitaciones de jamovi para la representación gráfica de datos.

Figura 1

Apariencia de jamovi.



Nota: elaboración propia.

2. ANTECEDENTES

En su corta historia (la primera versión se lanzó en 2017), jamovi ha causado un gran impacto en la comunidad científica. El programa ha llamado la atención de muchos académicos (de la Torre-Rodríguez et al., 2023), quienes han comenzado a utilizarlo como herramienta para el tratamiento y el análisis de los datos en sus trabajos de investigación, en ámbitos tan diversos como el de las ciencias de la educación o las ciencias de la salud (véase, por ejemplo, Arellano-Pimentel y Canedo-Ibarra, 2022; Lambea-Gil et al., 2023; Ron-Cordero y Avello-Martínez, 2023). Como consecuencia, la popularidad de jamovi ha crecido exponencialmente en los últimos años, en detrimento de la de otros programas similares. No en vano, jamovi se erigió como uno de los mejores *softwares* estadísticos en el estudio que realizaron Shepherd y Richardson (2024) en el que compararon las características de siete programas. Según esta comparativa, jamovi obtuvo la máxima puntuación, 5 puntos sobre 5, en aspectos como la facilidad de descarga, la calidad de los recursos online para usuarios noveles y la disponibilidad de recursos para los docentes, igualando o superando la puntuación de otros programas como Excel, JASP, PSPP o R. Del mismo modo, consiguió una valoración de 4,5 puntos sobre 5 en aspectos como la facilidad de uso y la potencia analítica. Otro estudio comparativo entre programas para el análisis y la visualización de datos, el que llevaron a cabo Abbasnasab et al. en 2021, señaló a jamovi como el programa más adecuado para usuarios noveles u ocasionales frente a otras alternativas como R, R Commander o SPSS. Entre los motivos, los autores destacaron el diseño de la interfaz (que es muy similar al de otros programas con los que el usuario puede estar familiarizado) y la organización de los menús (que resulta más coherente que la de otras alternativas).

Jamovi también ha irrumpido con fuerza en el ámbito educativo y su presencia en las aulas comienza a ser significativa. Las características del programa, en general, y la sencillez de su interfaz, en particular, hacen que su uso sea muy recomendable para la docencia, sobre todo la de materias de estadística o de investigación (Abbasnasab et al., 2021; Roque-Hernández, 2022a, 2022b). A pesar de que aún son escasos (probablemente, debido a la novedad del

programa), la literatura está empezando a incorporar trabajos que muestran los resultados de experiencias docentes centradas en el uso de jamovi en las aulas, especialmente en el contexto de la educación superior. Así, por ejemplo, Lino-Calle et al. (2024a) recogieron las opiniones de estudiantes ecuatorianos tras el uso de jamovi en una asignatura de posgrado. Los estudiantes evaluaron la eficacia del programa como herramienta para el análisis de datos, concluyendo su capacidad para simplificar el análisis estadístico y mejorar la interpretación de los resultados. Lino-Calle et al. (2024b) también utilizaron en su estudio una muestra de estudiantes de posgrado de Ecuador. El estudio consistía en la resolución de pares de tareas estadísticas, de dificultad similar, utilizando dos programas: jamovi para una de las tareas y Excel para la otra. Al analizar los datos, los autores comprobaron que la media de la calificación obtenida por los estudiantes era significativamente superior en las tareas que habían sido resueltas con jamovi en comparación con la de las tareas que se habían resuelto con Excel. En España, Santabárbara-Serrano et al. (2020) llevaron a cabo un estudio que tenía por objeto analizar las percepciones de un grupo de estudiantes de un posgrado en ciencias de la salud acerca del uso de jamovi frente a SPSS. Los estudiantes habían utilizado el *software* comercial como herramienta para el análisis de datos durante el transcurso de una de las asignaturas del posgrado y tuvieron un breve contacto con jamovi en un taller. El análisis de las opiniones de los estudiantes reveló una percepción muy positiva de jamovi frente a SPSS y una inclinación hacia el reemplazo de SPSS por jamovi como programa para el análisis estadístico en un futuro. Tobar-Viera et al. (2024) obtuvieron resultados similares al analizar las percepciones de un grupo de estudiantes universitarios de psicología en Ecuador.

En definitiva, la limitada literatura científica que existe a día de hoy sobre jamovi (al margen de manuales y documentos de ayuda) ensalza las potencialidades del programa y lo señala como una buena alternativa a otros *softwares*, tanto libres como comerciales, y subraya, además, una experiencia de uso positiva por parte de quienes comienzan a utilizarlo. En los trabajos consultados, jamovi se emplea exclusivamente con estudiantes que cursan un grado o un posgrado universitario, lo que podría llevar a pensar que el uso del programa queda restringido a las aulas universitarias. No obstante, dada la sencillez y la facilidad de uso del programa, no resultaría difícil utilizar jamovi entre los estudiantes de educación secundaria. Por otra parte, es importante notar que los estudios que se han encontrado se centran en un uso generalista del programa, sin abordar tareas específicas, por lo que no ha sido posible encontrar trabajos que se ocupen exclusivamente de la evaluación del uso de jamovi para la visualización de datos.

3. METODOLOGÍA

Para la consecución del objetivo propuesto se utilizará el conjunto de datos procedente de la Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación en los Hogares para el año 2023 que lleva a cabo el Instituto Nacional de Estadística y que se ha descrito en el artículo principal de este monográfico.

Con el objeto de facilitar al lector la visualización de las representaciones gráficas que se van a presentar, se ha optado por restringir el análisis únicamente a los encuestados de ciertas comunidades autónomas españolas. En concreto, se ha utilizado la información de los individuos de Andalucía, de Castilla-La Mancha y de Extremadura, dejando al margen la de los encuestados del resto de regiones. De las variables que componen el conjunto de datos, se han seleccionado las siguientes como principales a la hora de realizar las representaciones gráficas:

Uso de ordenador desde el hogar en los últimos 3 meses, una variable cualitativa con tres posibles opciones de respuesta (sí, no y no sabe/no contesta), y *Número de personas en el hogar*, una variable cuantitativa discreta. Adicionalmente, se han considerado el género y la comunidad autónoma de residencia de los encuestados como variables secundarias o de agrupación.

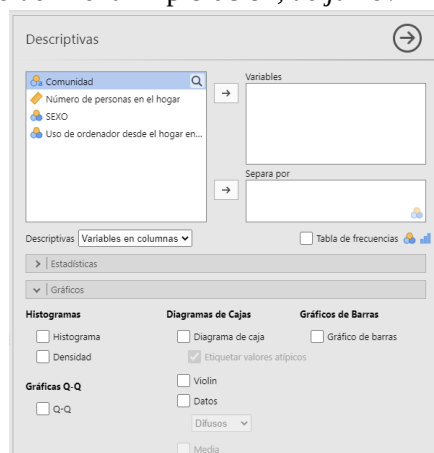
En este trabajo se van a mostrar las posibilidades que ofrece jamovi a la hora de realizar dos representaciones gráficas distintas: el gráfico de barras y el gráfico de violín. La primera de ellas, el gráfico de barras, constituye una herramienta fundamental para la representación de variables cualitativas y de variables cuantitativas discretas. Cuando en el gráfico de barras se representa una única variable, a la que se denomina principal, se dice que el gráfico es simple. No obstante, también es posible crear gráficos de barras agrupadas considerando en la representación, además de la variable principal, una o más variables secundarias que sirven para agrupar los datos. La relevancia de los gráficos de barras como herramienta para la representación de datos es tal que la construcción, lectura e interpretación de este tipo de gráficos figuran ya en el currículo de la Educación Primaria como habilidades a adquirir por parte de los estudiantes. El gráfico de violín, por su parte, es una representación gráfica más sofisticada que se emplea para mostrar la distribución de una variable cuantitativa. Con frecuencia, este tipo de gráfico se suele representar conjuntamente con el diagrama de caja y bigotes, proporcionando así más información sobre la variable. Al igual que ocurre con los gráficos de barras, la información que se muestra en los gráficos de violín puede aparecer representada en función de los valores de una o de varias variables de agrupación. El grado de abstracción que se necesita para comprender los elementos que conforman el gráfico de violín aconsejan presentar este tipo de representación en etapas educativas postobligatorias.

4. RESULTADOS

En esta sección se presentan, en primer lugar, los gráficos de barras que pueden realizarse utilizando jamovi. Como se ha indicado previamente, la facilidad para construir e interpretar este tipo de gráficos permite que se introduzcan en etapas educativas tempranas. Para obtener gráficos de barras con jamovi basta con acceder al menú *Exploración* y, después, al submenú *Descriptivas* en el programa (Figura 2).

Figura 2

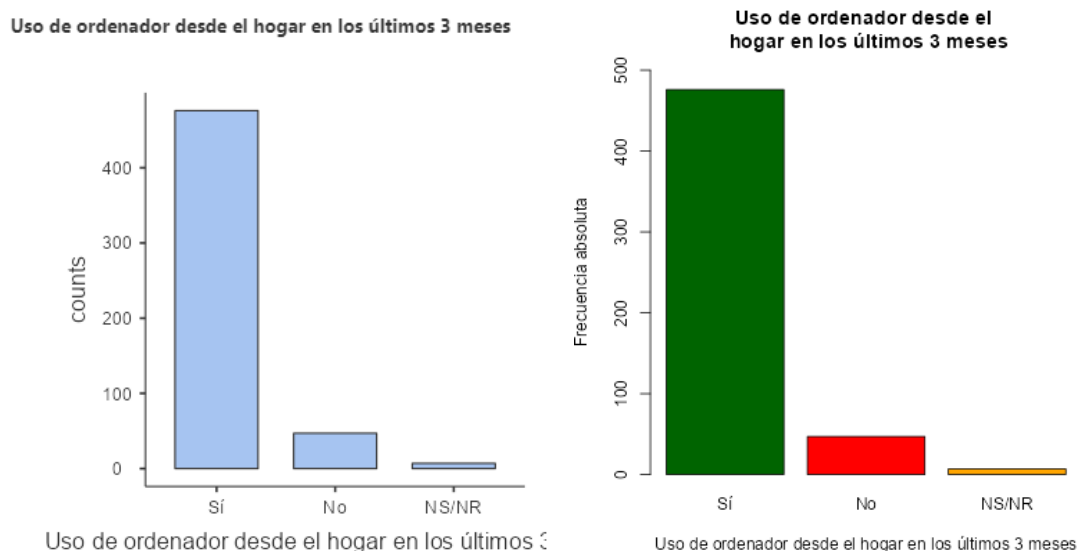
Submenú Descriptivas, dentro del menú Exploración, de jamovi



En la ventana que aparece, se coloca la variable principal en el espacio de *Variables* y, tras desplegar la sección de *Gráficos*, se marca la casilla *Gráfico de barras*.

Figura 3

Gráfico de barras simple para la variable Uso de ordenador desde el hogar en los últimos 3 meses utilizando las opciones del menú Descriptivas (izq.) y utilizando código de R (der.).



Nota: elaboración propia.

En la parte izquierda de la Figura 3 se muestra el gráfico de barras simple que resulta de considerar *Uso de ordenador desde el hogar en los últimos 3 meses* como variable principal. Según se aprecia en el gráfico, más de 400 individuos emplearon el ordenador desde el hogar en el trimestre anterior, mientras que el número de quienes no lo hicieron se sitúa en torno a 50.

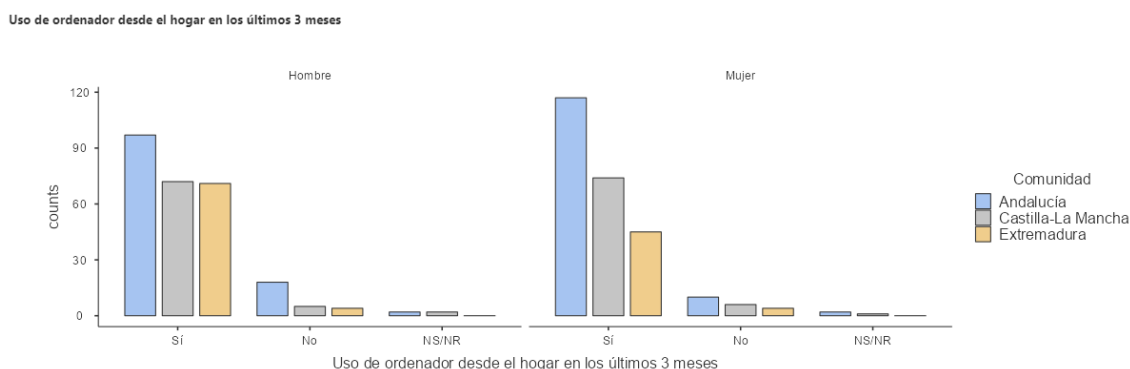
Un análisis detallado del gráfico a nivel técnico permite comprobar que el programa tiene en cuenta las características de los gráficos de barras a la hora de representarlos (se considera una barra por categoría, las barras aparecen equiespaciadas entre ellas y tienen todas la misma anchura, la altura de cada barra es proporcional a la frecuencia de aparición de la categoría correspondiente...) y que escoge automáticamente una escala apropiada para el eje vertical según la altura de las barras. Pero, por otro lado, se aprecian algunas carencias, la mayoría de las cuales se derivan de las limitaciones en la personalización de los análisis y de los resultados que permite jamovi. Una de las más evidentes hace referencia a las etiquetas de los ejes del gráfico, ya que una de ellas aparece cortada mientras que la otra está escrita en un idioma distinto al castellano. Pese a ello, el programa no ofrece la opción de editar ninguna de estas etiquetas. En este ejemplo se aprecia, además, que sería conveniente modificar los valores de referencia de la escala vertical para incluir un valor más que permitiese una mejor interpretación de la barra asociada al valor *Sí*, opción que tampoco ofrece jamovi. En resumen, en el momento de la creación del gráfico, el programa no permite editar ninguno de sus elementos lo que, en ocasiones, puede suponer un obstáculo a la hora de la interpretación. La única forma de personalizar los elementos de un gráfico en jamovi pasa por la instalación de un módulo adicional que permite la introducción y la ejecución de código de R desde jamovi. A modo de ejemplo, en la parte derecha de la Figura 3 se muestra un diagrama de barras equivalente al de la

parte izquierda (en el sentido de que ambos representan la misma información) pero que ha sido creado con código de R, lo que ha permitido la customización de sus elementos.

Otra de las representaciones visuales que jamovi permite realizar es el gráfico de barras agrupadas. En este tipo de gráficos se utilizan los valores de una o más variables cualitativas para agrupar los de la variable principal. Para realizar este tipo de gráficos en jamovi, se deben seguir las instrucciones que se han indicado para representar un gráfico de barras simple y, además, se deben colocar las variables de agrupación en el espacio denominado *Separa por*. A modo de ejemplo, en la Figura 4 se muestra un gráfico de barras agrupadas para la variable *Uso de ordenador desde el hogar en los 3 últimos meses* en el que se han considerado como variables de agrupación el género y la comunidad autónoma de residencia de los encuestados.

Figura 4

Gráfico de barras agrupadas para la variable Uso de ordenador desde el hogar en los últimos 3 meses considerando el género y la comunidad autónoma de residencia como variables de agrupación.



Nota: elaboración propia.

En este gráfico se aprecia que la mayoría de hombres y de mujeres de las tres regiones consideradas hicieron uso del ordenador desde el hogar en el trimestre previo a la realización de la encuesta. En concreto, en el caso de Andalucía, el número de mujeres que utilizó el ordenador desde sus casas supera al de hombres; en Extremadura ocurre justo lo contrario, y en Castilla-La Mancha la cantidad de hombres y de mujeres que emplearon este dispositivo electrónico fue bastante similar. Por otro lado, los hombres andaluces que no utilizaron el ordenador desde sus hogares superan en número a las mujeres de la misma región. En las otras dos comunidades autónomas, no se aprecian diferencias notables en función del género entre los individuos que no utilizaron el ordenador.

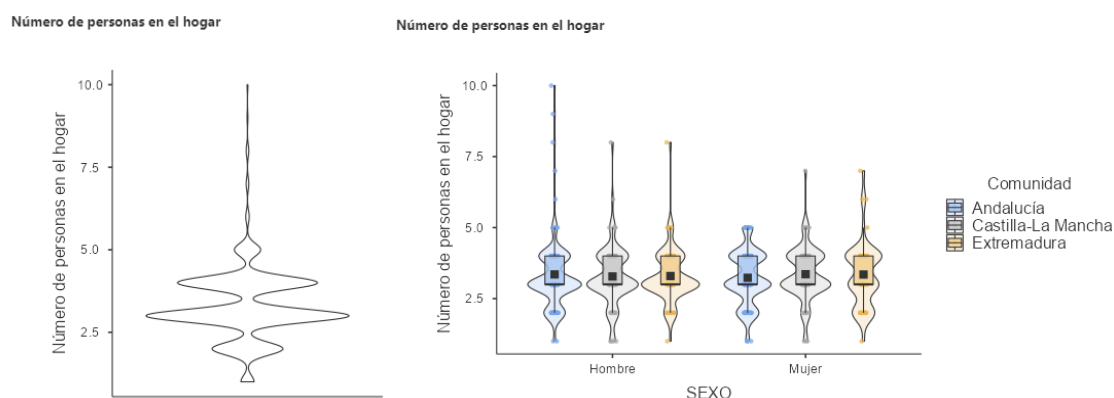
Como en el ejemplo anterior, el gráfico de barras agrupadas que realiza jamovi es correcto a nivel teórico. Sin embargo, al margen de los que ya se han mencionado, este gráfico pone de manifiesto ciertos aspectos que podrían mejorarse de cara a incrementar la usabilidad y la versatilidad del programa en lo relativo a visualización gráfica. Uno de los más trascendentes hace referencia al uso obligatorio que hace el programa de las frecuencias absolutas como valores de referencia en el eje vertical. En efecto, en los gráficos de barras que realiza jamovi la altura de cada una de las barras es proporcional a la frecuencia absoluta de la categoría representada. Esta opción, aunque correcta, puede no ser la más adecuada para la interpretación del gráfico en los casos en que se representan grupos de distinto tamaño. Esa es, precisamente, la situación que se plantea en la Figura 4, ya que el número de encuestados de las tres regiones

que se han considerado es diferente. En este tipo de escenarios es mucho más conveniente que la altura de las barras del gráfico sea proporcional a la frecuencia relativa o a la frecuencia porcentual de aparición de la categoría correspondiente.

Además de gráficos de barras, jamovi permite la creación de representaciones más complejas, como pueden ser los histogramas, los gráficos de puntos o los gráficos de violín, las cuales se suelen presentar al estudiante cuando cursa la educación secundaria. Para ilustrar las posibilidades que ofrece jamovi para la representación de este tipo de representaciones se han considerado, a modo de ejemplo, los gráficos de violín. Para realizar un gráfico de violín, es necesario acceder al menú *Exploración* y, después, al submenú *Descriptivas* de jamovi. Allí, se coloca la variable principal en el espacio *Variables* y las variables de agrupación, si las hay, en el espacio *Separa por*. Después, en la sección de *Gráficos* se debe marcar la casilla *Violín*. En la Figura 5 se muestran dos ejemplos de gráficos de violín realizados con jamovi.

Figura 5

Distribución del número de personas en el hogar para el total (izq.) y según el género (der.).



Nota: elaboración propia.

La representación que aparece en la parte izquierda de la Figura 5 se corresponde con un gráfico de violín simple que muestra la distribución del número de personas que residen en el hogar de los encuestados. El análisis de este gráfico permite identificar los valores más y menos frecuentes en la variable y, consecuentemente, acotar el valor de la moda y extraer información acerca de la simetría de la distribución. El análisis del gráfico permite comprobar que los valores de la variable son discretos, ya que la curva de la distribución de valores únicamente muestra picos en valores naturales. Se aprecia, además, que el número de personas en el hogar más frecuente es 3 y que existen pocos hogares compuestos por más de 5 miembros. Por otro lado, se intuye una cierta asimetría positiva en la distribución.

En la representación de la parte derecha de la Figura 4 aparece otro gráfico de violín en el que las representaciones se muestran desagregadas por género y por comunidad autónoma de residencia. Como se puede apreciar, en este ejemplo se ha hecho uso de la opción que ofrece jamovi de incorporar elementos adicionales sobre el gráfico de violín para enriquecer la información que se proporciona. En concreto, cada diagrama de violín se ha complementado con el diagrama de caja y bigotes, con los datos brutos para la combinación de valores de la variable de agrupación (que aparecen marcados con puntos de colores) y con el valor medio de la variable numérica (que aparece marcado con un cuadrado negro) correspondientes. Incorporar toda esta información con jamovi es tan sencillo como marcar las casillas

correspondientes en la sección de *Gráficos*. Según este gráfico, la distribución del número de personas en los hogares en Andalucía y de Castilla-La Mancha es similar independientemente del género, aunque en los hogares de los hombres andaluces se observan algunos valores excepcionalmente altos en esta variable en comparación con los de las mujeres de la misma región. Las mayores diferencias en función del género en la distribución del número de miembros que conforman los hogares se aprecian en Extremadura. En esta región, el número más frecuente de personas en los hogares de los hombres encuestados es de 3, mientras que en los hogares de las mujeres, esta cifra es 4.

5. CONCLUSIONES

Jamovi es un programa para el análisis estadístico que ofrece opciones muy interesantes para la visualización de datos. A modo de ejemplo, en este trabajo se ha ilustrado el uso del programa para la realización de gráficos de barras y de gráficos de violín, aunque el catálogo de representaciones gráficas que permite crear jamovi es mucho más amplio.

A pesar de que la primera versión de jamovi fue lanzada hace apenas una década, su uso se ha propagado muy rápidamente entre científicos e investigadores y, también, entre docentes y estudiantes, principalmente, por la gratuidad del programa. En efecto, el hecho de ser un software libre de costo hace que cualquier usuario que disponga de un dispositivo electrónico con conexión a internet pueda descargar y usar el programa. Otro de los puntos fuertes de jamovi hace referencia a lo sencillo de su manejo. Jamovi ofrece al usuario una interfaz intuitiva y amigable, basada en menús, lo que hace que la curva de aprendizaje del programa sea relativamente suave, ya que el tiempo y el esfuerzo que se requieren para aprender a utilizarlo son reducidos. Por otra parte, es importante tener en cuenta que jamovi se basa en el software estadístico R, de quien hereda características como la versatilidad y la potencia. Esta relación entre jamovi y R permite que cualquier usuario pueda ejecutar código de R en jamovi y obtener los resultados correspondientes sin salir del programa, personalizando así la ejecución de los análisis a conveniencia. Esta posibilidad de realizar los análisis a través de los menús del programa o ejecutando código de R convierte a jamovi en una herramienta ideal para satisfacer las necesidades de usuarios tanto novatos como expertos.

Pero jamovi también presenta algunas limitaciones que pueden cuestionar su uso en algunos escenarios. La más importante de ellas se refiere a las restricciones en los análisis que el programa permite ejecutar. La instalación estándar de jamovi pone a disposición de los usuarios los módulos que incluyen las técnicas de análisis y de visualización de datos que se usan con más frecuencia. Esto implica que existen ciertos análisis que no pueden realizarse utilizando los menús de jamovi. Para solventar este problema, el programa cuenta con una biblioteca de módulos adicionales que se pueden instalar para extender las funcionalidades del programa y atender demandas específicas de los usuarios. No obstante, el número de módulos disponibles en jamovi es ciertamente reducido si lo comparamos, por ejemplo, con el número de paquetes de R, lo que supone una limitación en cuanto a las opciones de análisis de jamovi en comparación con este otro programa. Por otro lado, jamovi presenta los resultados de una manera clara, ordenada y siguiendo las recomendaciones de formato dictadas por la American Psychological Association. Sin embargo, la posibilidad de editar o personalizar estos resultados es prácticamente nula. Esto es especialmente limitante cuando se trata de representaciones gráficas, en las que la edición de sus elementos (nombres o escalas de ejes, título, leyenda, color de los elementos...) suele ser necesaria para adecuarlos a los requerimientos del usuario.

Tras la exposición de los beneficios y los inconvenientes del uso de jamovi, los autores de este trabajo consideran que los primeros superan ampliamente a los segundos. Se recomienda, por tanto, la utilización del programa como herramienta para la visualización de datos en el ámbito educativo. Teniendo en cuenta las características del programa, se suscribe la recomendación de autores como Ibáñez-López et al. (2023) y de la Torre et al. (2023), de utilizar jamovi para la visualización de datos en la educación superior, tanto a nivel de grado como de posgrado. Incluso, se propone extender esta recomendación a la enseñanza secundaria, tanto obligatoria como postobligatoria, en cuyas aulas se considera que jamovi podría convertirse en un *software* ideal para la representación gráfica de datos.

REFERENCIAS BIBLIOGRÁFICAS

- Abbasnasab Sardareh, S., Brown, G. T. L. y Denny, P. (2021). Comparing four contemporary statistical software tools for introductory data science and statistics in the social sciences. *Teaching Statistics* 43(S1), S157–S172. <https://doi.org/10.1111/test.12274>
- Arellano-Pimentel, J. J. y Canedo-Ibarra, S. P. (2022). EpAA: Entorno para el Aprendizaje de Algoritmos. Una experiencia educativa desde la perspectiva del aprendizaje flexible. *EDUTEC. Revista Electrónica de Tecnología Educativa*, 79, 63-79. <https://doi.org/10.21556/edutec.2022.79.2451>
- Āžahin, M. y Aybek, E. (2020). 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>
- Berrendero, J. R. (2021). *Una introducción a jamovi*. Universidad Autónoma de Madrid. <https://verso.mat.uam.es/~joser.berrendero/blog/tutorial-jamovi.pdf>
- de la Torre Rodríguez, M., Junco Bringa, D., Marrero Pérez, M. D. y Rodríguez Soto, I. (2023). Software Jamovi en la docencia de la asignatura Metodología de la Investigación. *Educación Médica Superior*, 37(4). http://scielo.sld.cu/scielo.php?pid=S0864-21412023000400009&script=sci_arttext
- Elosua, P. y Egaña, M. (2020). *Psicometría aplicada. Guía para el análisis de datos y escalas con jamovi*. Servicio Editorial de la Universidad del País Vasco.
- Ibáñez-López, F. J., Ponce Gea, A. I., Pedreño Plana, M. y Sánchez-Martín, M. (2023). Basic survival manual for descriptive statistical analysis. *Espiral. Cuadernos del Profesorado*, 16(32), 118-125. <https://doi.org/10.25115/ecp.v16i32.9134>
- Lambea-Gil Á., Saldaña-Inda I., Lamíquiz-Moneo I. y Cisneros-Gimeno A. I. (2023). Neurofobia entre los estudiantes de medicina de una universidad española: experiencias más allá de la anglosfera. *Revista de Neurología*, 76(11), 351-359. <https://doi.org/10.33588/rn.7611.2023102>
- Lino-Calle, V. A., Carvajal-Rivadeneira, D. D., Muñoz-Zambrano, J. L. e Intriago-Delgado, Y. M. (2024a). Jamovi como herramienta para el análisis de datos en la asignatura de Estadística y Diseño de Experimentos. *Alcance*, 7(1), 73-83. <https://doi.org/10.47230/ra.v7i1.62>
- Lino-Calle, V. A., Carvajal-Rivadeneira, D. D., Sornoza-Parrales, D., Vergara-Ibarra, J. L. e Intriago-Delgado, Y. M. (2024b). Herramienta tecnológica Jamovi en el análisis e interpretación de datos en proyectos de Ingeniería Civil. *Revista Innovaciones Educativas*, 26(41), 151-165. <https://doi.org/10.22458/ie.v26i41.5145>

- Ron-Cordero, A. A. y Avello-Martínez, R. (2023). Percepción de los estudiantes sobre la realidad aumentada como recurso didáctico para el aprendizaje de Ciencias Naturales. *Revista Latinoamericana de Ciencias Sociales y Humanidades*, 4(3), 1394-1404. <https://doi.org/10.56712/latam.v4i3.1167>
- Roque-Hernández, R. V. (2022a). El software estadístico en la docencia y la investigación. *Realidades Educativas Entrelazadas*, 1(1), 18-21. <https://doi.org/10.58400/ree.vi1.6>
- Roque-Hernández, R. V. (2022b). La enseñanza de la estadística para la investigación: algunas recomendaciones reflexionadas desde la praxis. *Revista Educación*, 46(2), 1-10. <https://doi.org/10.15517/revedu.v46i2.47569>
- Santabárbara Serrano, J., Lasheras, I. y Rubio, E. (2020). Taller de jamovi en residentes de medicina que cursan un postgrado de investigación: una experiencia docente. *Revista Española de Educación Médica*, 1(1), 90–95. <https://doi.org/10.6018/edumed.426491>
- Shepherd, M. A. y Richardson, E. J. (2024). Opting for open-source? A review of free statistical software programs. *Teaching Statistics*, 46, 53–63. <https://doi.org/10.1111/test.12360>
- The jamovi project. (2024a). *jamovi (Version 2.5)* [Computer Software]. <https://www.jamovi.org>
- The jamovi project. (2024b). *jamovi*. <https://www.jamovi.org/>
- The jamovi project. (2024c). *jamovi library*. <https://www.jamovi.org/library.html>
- Tobar-Viera, A., Núñez-Núñez, M., Mejía-Rubio, A. R. y Eugenio_Zumbana, L. (2024). Actitudes hacia la investigación y uso de Jamovi en estudiantes de Psicología Clínica. *Academo*, 11(3), 287-295. <https://doi.org/10.30545/academo.2024.set-dic.8>

Visualización de datos con PSPP

Inmaculada Serrano García

Cuerpo de Profesores de Enseñanza Secundaria, especialidad Matemáticas en I.E.S
Villanueva del Mar, La Herradura (Granada), isergar510@g.educaand.es

Ana Enríquez Molina

Cuerpo de Profesores de Enseñanza Secundaria, especialidad Matemáticas en I.E.S
Puerta del Mar, Almuñécar (Granada), aenrmol336@g.educaand.es

Resumen: En este artículo se explora la visualización de datos utilizando PSPP, un software de código abierto que sirve como alternativa gratuita a SPSS. Destaca la importancia de transformar datos complejos en representaciones visuales para facilitar su análisis y comprensión, especialmente en el ámbito educativo. PSPP permite realizar análisis estadísticos y generar gráficos básicos, lo que lo convierte en una herramienta accesible para estudiantes. A pesar de algunas limitaciones en comparaciones gráficas complejas, PSPP es efectivo para enseñar estadística y fomentar el interés por el análisis de datos en niveles educativos intermedios.

Palabras clave: PSPP, visualización, datos, estadística, educación.

Data's visualization with PSPP

Abstract: This article explores data visualization using PSPP, an open-source software that serves as a free alternative to SPSS. It highlights the importance of transforming complex data into visual representations to facilitate analysis and understanding, especially in the educational field. PSPP allows for statistical analysis and the generation of basic graphs, making it an accessible tool for students. Despite some limitations in complex graphical comparisons, PSPP is effective for teaching statistics and fostering interest in data analysis at intermediate educational levels.

Key words: PSPP, visualization, data, statistics, education.

1. INTRODUCCIÓN

La visualización de datos ha emergido como una disciplina esencial en la era de la información, permitiendo transformar datos complejos en representaciones visuales que facilitan su interpretación y análisis. Este enfoque no solo mejora la comprensión de grandes volúmenes de información, sino que también optimiza la comunicación efectiva de los hallazgos. Tradicionalmente, los gráficos estadísticos han sido la herramienta principal en este campo. Sin embargo, el avance tecnológico ha dado lugar a una diversidad de técnicas y herramientas innovadoras que amplían las posibilidades de representación gráfica, adaptándose a las necesidades cada vez más sofisticadas de la sociedad contemporánea.

Una de las herramientas disponibles para la visualización y análisis de datos es PSPP (GNU Project, 2015), un software de código abierto diseñado como una alternativa gratuita a SPSS (IBM Corp., 2023). Desarrollado por el proyecto GNU, PSPP es una opción accesible y potente

para estudiantes interesados en aprender sobre análisis estadísticos sin incurrir en gastos adicionales que permite realizar una amplia gama de análisis. Aunque PSPP está orientado principalmente al análisis estadístico, también incluye funciones para la creación de gráficos básicos, lo que facilita a los estudiantes la visualización y comprensión de datos de forma más eficaz. Su interfaz sencilla y la ausencia de limitaciones en su uso lo convierten en una herramienta ideal para que los jóvenes comiencen a explorar el mundo del análisis de datos de manera práctica y efectiva.

Algunos trabajos de investigación previos han comparado las características y funcionalidades del software PSPP frente al programa SPSS. Entre las principales ventajas de PSPP sobre SPSS, puede destacarse que es un programa abierto y gratuito, lo que lo hace accesible para estudiantes y educadores sin limitaciones de tipo económico. Además, se resalta la flexibilidad y el uso comercial de PSPP, dotando a los estudiantes de habilidades prácticas sobre la codificación, uso, análisis e interpretación de datos, lo que desarrollará en los mismos, competencias que puedan servirles frente a futuros empleos (Rybenská et al., 2014).

La comunidad de usuarios de la plataforma ha generado una variedad de recursos y tutoriales en línea que pueden ser muy útiles para aprender a utilizar la herramienta. Estos tutoriales incluyen guías en video, manuales escritos y foros de discusión que abordan desde conceptos básicos hasta funciones más avanzadas de PSPP. Cabe destacar que, para un uso más eficiente y fluido, se recomienda utilizar PSPP en un ordenador de escritorio o portátil.

Esta herramienta es accesible para el profesorado de matemáticas interesado en incorporar análisis y visualización de datos en sus clases. Su interfaz, similar a la de SPSS, facilita la adaptación para aquellos estudiantes que ya están familiarizados con el análisis estadístico, permitiendo una transición adecuada a este software libre. PSPP permite la realización desde análisis descriptivos básicos como tablas de frecuencias o medidas de tendencia central, hasta análisis inferencial como pruebas paramétricas o análisis de regresión, proporcionando una visualización clara y detallada de los datos. El manejo de archivos es intuitivo, ya que soporta formatos comunes como .sav y .csv, facilitando la importación y exportación de datos entre diferentes plataformas y aplicaciones.

Con PSPP, los profesores pueden gestionar grandes conjuntos de datos, personalizar tablas y gráficos para adaptarse a sus necesidades didácticas, y exportar resultados en formatos que se integren fácilmente en informes y presentaciones. Para estudiantes de edades tempranas es recomendable complementar el uso de PSPP con herramientas educativas diseñadas específicamente para facilitar el aprendizaje de conceptos estadísticos básicos como CODAP, proporcionando una experiencia educativa equilibrada.

Este trabajo tiene como meta proporcionar al profesorado de matemáticas de diferentes niveles educativos una guía básica sobre el uso del software PSPP para la exploración de datos a través de visualizaciones en línea, con el enfoque presentado en la primera sección del monográfico titulado “Explorando la Visualización de Datos” (Molina-Portillo et al., 2024). El propósito es atender a los diferentes niveles de demanda cognitiva y habilidades digitales del alumnado, presentando tanto las funcionalidades que PSPP ofrece, como el análisis de las visualizaciones generadas.

2. ANTECEDENTES

El constante desarrollo tecnológico de las últimas décadas ha puesto a disposición de los docentes una serie de recursos que se han convertido en herramientas fundamentales para el área educativa. La tecnología facilita la entrega efectiva de conocimientos, y permite el apoyo entre los educadores, el entrenamiento individual y la reflexión. Así, la integración de la tecnología en el aula se lleva a cabo a través de un proceso dinámico, destacándose como aspectos clave para tal fin, la necesidad de una administración eficiente de los recursos tecnológicos, superar las barreras que surgen durante su integración en los procesos de enseñanza-aprendizaje o el desarrollo profesional y competencial de los docentes, que deben recibir formación no solo del uso de la tecnología sino también de la integración de las herramientas digitales de forma pedagógica. También debe destacarse la necesidad de una experiencia práctica que permita aplicar lo aprendido en situaciones reales del aula, superar los obstáculos internos y externos que surgen en la integración tecnológica como es la falta de recursos o la resistencia al cambio y la importancia de desarrollar habilidades adicionales, como aquellas vinculadas a la competencia comunicativa con el alumnado para facilitar el desarrollo de las experiencias didácticas (Arteaga et al., 2023).

En línea con lo expuesto, han surgido muchos programas informáticos que facilitan esta integración tecnológica en el aula. De forma más específica, y en el área relacionada con la estadística, el análisis y la visualización de datos, PSPP es uno de los softwares gratuitos que pueden emplearse en la enseñanza en diferentes etapas educativas. El hecho de tener una interfaz intuitiva mediante vista de datos y variables, ofrecer una multitud de herramientas estadísticas, facilitar la manipulación de datos, su gratuidad y su compatibilidad con su versión de pago (SPSS), lo hacen una opción atractiva para las instituciones educativas (Yagnik, 2014).

En el contexto de la enseñanza, PSPP se ha utilizado en cursos de estadística y metodología de investigación para el desarrollo de competencias para el análisis de datos en los estudiantes. Concretamente, el trabajo de investigación realizado por Tindowen et al. (2019) en estudiantes de 16 y 17 años sobre la implementación de este software en el aula, demuestra que dicha herramienta puede mejorar la enseñanza de la estadística a nivel educativo. En esta línea, PSPP fue utilizado como recurso para facilitar el aprendizaje de conceptos estadísticos complejos, promoviendo un enfoque más práctico. Los resultados indican que no solo fortalece las habilidades procedimentales y conceptuales del alumnado, sino que también fomenta una actitud positiva hacia el aprendizaje de la estadística. Esto se traduce en un mayor compromiso y motivación y, en última instancia, en una mejora en su rendimiento académico.

Otro claro ejemplo es el estudio realizado por Cabrera (2022), quien compara la metodología Flipped Classroom y clase magistral en una propuesta didáctica para el bloque de Estadística de Matemáticas orientadas a las enseñanzas académicas de 4.º de ESO, haciendo uso en ambas metodologías del software estadístico PSPP como complemento a la enseñanza teórica de la asignatura. Entre los hallazgos de este trabajo de investigación, se subraya la conveniencia del uso del programa estadístico en ambas líneas metodológicas, pues involucra al alumnado en el aprendizaje, permitiendo que compruebe de manera práctica los resultados teóricos. Asimismo, este trabajo se completa con un manual de uso de PSPP para estudiantes de educación secundaria obligatoria que permite guiar a los estudiantes en las tareas prácticas de la propuesta didáctica llevada a cabo en el mismo (Cabrera, 2022).

3. METODOLOGÍA

3.1. Diseño y participantes

El objetivo de este estudio es ilustrar las características y propiedades del software PSPP en el análisis y visualización de datos, con especial énfasis en su aplicación en contextos educativos y formativos. En continuidad con lo descrito en el primer informe del monográfico, para tal fin se realizará un estudio exploratorio-descriptivo, de carácter transversal y ex post facto, a partir de una base de datos compuesta por 2565 participantes, todos ellos niños de entre 10 a 15 años (♂ = 51,9%; n = 1332; ♀ = 48,1% n = 1233). Este análisis se basa en la base de datos original proporcionada por el Instituto Nacional de Estadística sobre uso de dispositivos tecnológicos en niños en edad escolar "Menores de 10 a 15 años en la Encuesta sobre Equipamiento y Uso de las TICs de 2023", que comprende 2.565 individuos organizados en 21 variables (INE, 2024). El tamaño muestral incluye aproximadamente 2500 secciones censales y alrededor de 25000 viviendas. El INE empleó un diseño de muestreo trietápico, con estratificación de las unidades de primera etapa, que son las secciones censales, seguidas por las viviendas familiares principales como unidades de segunda etapa. En la tercera etapa, se selecciona un individuo de 16 años o más por vivienda, además de recopilar información de todos los niños de 10 a 15 años. A partir de esta fuente, se detalla el proceso metodológico seguido y los resultados obtenidos utilizando PSPP.

3.2. Instrumentos

El instrumento utilizado para la recogida de datos que constituye la base de datos empleada en este estudio es la “Encuesta sobre Equipamiento y Uso de las TICs” realizada por el INE en 2023 (INE, 2024). De forma específica, esta encuesta se llevó a cabo de forma continua con periodicidad anual, dirigida a la población que reside en viviendas familiares principales en todo el territorio nacional. El período de referencia para los resultados corresponde al primer trimestre del año, mientras que la mayor parte de las variables investigadas se relacionan con los doce y tres meses previos a la entrevista.

Dado que el objetivo del estudio es mostrar la visualización de datos con PSPP, resulta de interés señalar las variables que emplea la base de datos con la que se realiza este trabajo, así como sus características y forma de medida; las cuales son obtenidas mediante la encuesta señalada de forma previa. Éstas pueden ser consultadas en la tabla 1 y 2.

Tabla 1
Información de las variables de menores de 10 a 15 años en la Encuesta sobre Equipamiento y Uso de las TICs para el año 2023.

Variable	Posición	Etiqueta	Medida
ID	1	Identificador	Nominal
CODAUTO	2	Código Comunidad Autónoma	Nominal
Comunidad	3	Comunidad	Nominal
CPRO	4	Código Provincia	Nominal
Provincia	5	Provincia	Nominal
SERIAL	6	Número de cuestionario	Escala

Variable	Posición	Etiqueta	Medida
TIP_H	7	Tipo de Hogar	Nominal
TOT_MH	8	Miembros del Hogar	Escala
TOT10_15	9	Menores de 10 y 15	Nominal
ORD	10	Disponen de ordenador	Nominal
TABLET	11	Disponen de Tablet	Nominal
TELEF1	12	Disponen de teléfono fijo	Nominal
TELEF2	13	Disponen de teléfono móvil	Nominal
VIV_INTER	14	Acceso internet en la vivienda	Nominal
INTEFOR	15	Forma de conexión de banda ancha	Nominal
NPER	16	Número de persona seleccionada	Nominal
SEXO	17	Sexo	Nominal
EDAD	18	Edad	Escala
PC	19	Utilización de ordenador en los últimos 3 meses desde el hogar	Nominal
INT	20	Utilización de internet en los últimos 3 meses desde casa	Nominal
MOVIL	21	Teléfono Móvil	Nominal

Fuente: INE (2024). Elaboración propia.

Tabla 2
Etiquetas de los posibles valores de cada variable de la base de datos.

Nombre	Valor	Etiqueta
Tipo de Hogar	1	Hogar unipersonal
	2	Padre o madre solo/a que conviva con algún hijo
	3	Pareja sin hijos que convivan en el hogar
	4	Pareja con hijos que convivan en el hogar
	5	Otro tipo de hogar
Disponen de Ordenador	1	Si
	6	No
	9	NS/NR
Disponen de tablet	1	Si
	6	No
	9	NS/NR
Disponen de teléfono fijo	1	Si
	6	No
	9	NS/NR
Disponen de teléfono móvil	1	Si
	6	No
	9	NS/NR
Nombre	Valor	Etiqueta
Acceso a internet en la vivienda	1	Si
	6	No
	9	NS/NR
Forma de conexión de banda ancha	1	Fija
	2	Solo móvil
	9	NS/NR
Sexo	1	Hombre

	6	Mujer
	1	Si
Utilización de ordenador en los últimos 3 meses	6	No
	9	NS/NR
	1	Si
Utilización de internet en los últimos 3 meses	6	No
	9	NS/NR
	1	Si
Teléfono móvil	6	No
	9	NS/NR

Fuente: INE (2024). Elaboración propia.

3.3. Procedimiento

En primer lugar, se ha realizado una revisión de la literatura con el fin de introducir el software PSPP, describiendo algunas de sus características básicas, su interfaz y las consideraciones iniciales para comenzar a trabajar con él.

Posteriormente, se ha explicitado el procedimiento a seguir para trabajar con el conjunto de datos ofrecido por el INE (Encuesta sobre Equipamiento y Uso de las TICs de 2023), trabajando inicialmente con un fichero .xlsx. A partir de este, se realiza una importación en PSPP, permitiendo la concreción de las variables objeto de estudio, su denominación por tipología y la inserción de etiquetas de valor. Finalmente, se genera un fichero .sav que permite el visionado del conjunto de datos y la vista de variables de forma operativa.

A continuación, se procede con el análisis de los datos, con el fin de mostrar las opciones que ofrece este software en entornos educativos. En primer lugar, se presenta una tabla con los estadísticos descriptivos de las variables ‘miembros del hogar’ y ‘edad’. Seguidamente, se muestra en un gráfico de barras la frecuencia de encuestados en función de la Comunidad Autónoma. Tras ello, se ilustra la relación entre el tipo de hogar y el número de miembros que lo componen mediante un gráfico de puntos.

Posteriormente, se muestra un histograma que representa la distribución del número de miembros del hogar y se superpone la curva normal sobre la distribución. Además, se exploran las herramientas que PSPP ofrece para analizar la relación entre más de una variable a partir de un gráfico de barras. Asimismo dado que se emplean variables cualitativas nominales, se ha realizado una tabla de contingencia para analizar si existe una relación estadísticamente significativa entre las variables de la base de datos. Como resultado se ha identificado una correlación entre el número de miembros del hogar y la disponibilidad de ordenador. A partir de este análisis se visualiza dicha relación mediante diagramas de sectores.

Por tanto, y a través de todo lo expuesto, este enfoque busca demostrar la eficacia de PSPP para realizar análisis estadísticos exhaustivos y crear visualizaciones que faciliten la comprensión y comunicación de los resultados en estudios descriptivos y exploratorios.

3.4. Análisis de los datos

Para el análisis de los datos se usa el software PSPP – GNU Project (Free Software Foundation). Concretamente, se emplean análisis descriptivos básicos (frecuencias y medidas de tendencia central) de las variables de objeto de estudio, así como diferentes tipos de gráficos para la visualización de la distribución de datos (gráficos de barras, gráfico de puntos, diagrama de sectores e histograma). También se emplean tablas de contingencia para el contraste de

variables de tipo nominal, empleando el estadístico Chi-cuadrado de Pearson para determinar la existencia de asociación estadísticamente significativa entre las variables. En concreto, se considerarán significativos todos los resultados con un p-valor asociado inferior a 0,05. El Índice de Confiabilidad (IC) en los análisis realizados se concretó en el 95%.

4. RESULTADOS

Se ha llevado a cabo un análisis exhaustivo de la herramienta PSPP orientado principalmente hacia su aplicación en la visualización de datos, con un enfoque particular en el ámbito educativo de la materia de matemáticas. La ventana principal del programa se divide en dos vistas: la vista de datos, que permite visualizar y editar los registros de la base de datos, y la vista de variables, que posibilita la definición y modificación de las características de las variables empleadas en el análisis.

Adicionalmente, PSPP ofrece una serie de menús desplegables que proporcionan una amplia gama de opciones para realizar análisis estadísticos, generar gráficos y exportar resultados. Esto hace que el proceso analítico sea accesible incluso para usuarios con un conocimiento básico de estadística. PSPP admite múltiples formatos de archivo, como .sav (propio de SPSS) y .csv (archivos de texto separados por comas), y permite la importación de datos desde bases de datos externas u hojas de cálculo.

Una vez cargados los datos, es fundamental revisar y definir las variables en la "Vista de Variables". En esta sección, se pueden asignar nombres a las variables, definir su tipo (por ejemplo, numéricas o de texto), establecer etiquetas que faciliten su interpretación y ajustar propiedades adicionales, como el ancho o el formato de los datos. En la "Vista de Datos", es posible añadir nuevas entradas o editar las existentes, lo que permite la introducción manual de datos o la corrección de errores en los datos importados. Siguiendo la secuencia "Archivo" → "Muestra información del archivo de datos" → "Archivos de trabajo", se obtiene una primera comprensión de la base de datos utilizada, como se observa en los resultados anteriores concretamente en la Tablas 1 y 2, donde se resumen las características de cada una de las variables.

A continuación, se presenta la Tabla 3, que resume los estadísticos descriptivos de las variables ‘Edad’ y ‘Miembros del Hogar’ a modo de ejemplo, indicando el valor mínimo, el valor máximo, la media y la desviación estándar. Estos datos proporcionan una descripción clara y concisa de las características de la muestra. Además, PSPP permite calcular otros indicadores estadísticos, como la varianza, la curtosis y la asimetría.

Tabla 3
Estadísticos descriptivos de las variables objeto de estudio.

	N	Mínimo	Máximo	Media	Desviación estándar
Edad	2565	10	15	12,57	1,698
Miembros del Hogar	2565	2	13	4,03	1,070

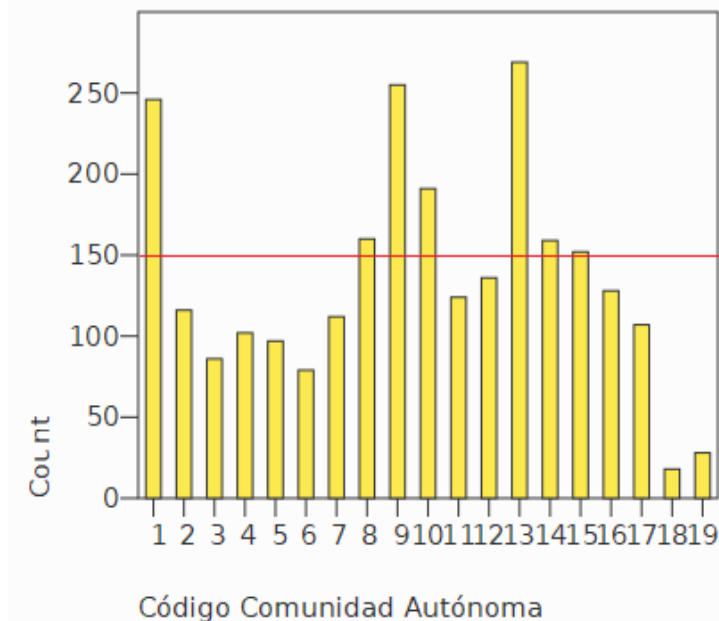
Fuente: INE (2024). Elaboración propia.

En el siguiente análisis se examina la distribución geográfica de los encuestados. La Figura 1 presenta un diagrama de barras que detalla el número de individuos según su Comunidad Autónoma de origen, representada mediante códigos numéricos. Se destaca que la mayor parte de los encuestados procede de las Comunidades Autónomas de Andalucía (código 1, n=246), Castilla-La Mancha (código 8, n=160), Cataluña (código 9, n=255), Comunidad Valenciana

(código 10, n=191), Madrid (código 13, n=269), Murcia (código 14, n=159) y Navarra (código 15, n=152), todas ellas con una frecuencia superior a 150 individuos. Este tipo de visualización gráfica puede resultar especialmente útil en la enseñanza en niveles educativos inferiores, facilitando la comprensión de la distribución espacial de la muestra.

Figura 1

Número de individuos encuestados según Comunidad Autónoma de origen.

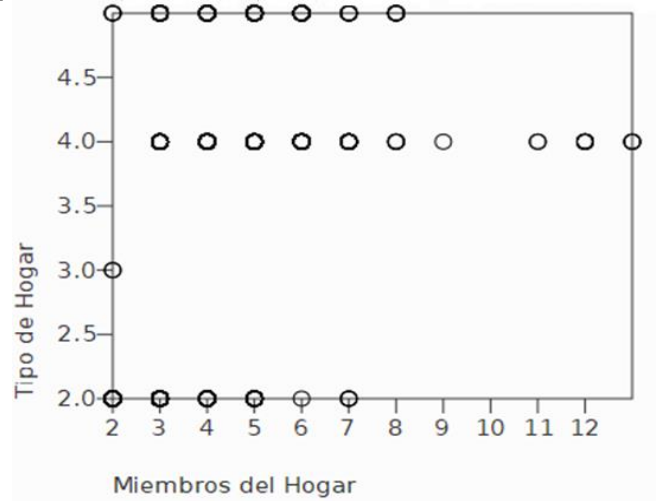


Nota: 1, Andalucía; 2, Aragón; 3, Asturias; 4, Baleares; 5, Canarias; 6, Cantabria; 7, Castilla y León; 8, Castilla-La Mancha; 9, Cataluña; 10, Comunidad Valenciana; 11, Extremadura; 12, Galicia; 13, Comunidad de Madrid; 14, Murcia; 15, Navarra; 16, País Vasco; 17, La Rioja; 18, Ceuta; 19, Melilla. Fuente: INE (2024). Elaboración propia.

La Figura 2, representada mediante un gráfico de puntos, ilustra la relación entre el tipo de hogar y el número de miembros que lo componen. Los resultados indican que, en hogares con más de ocho miembros, el único tipo de hogar observado corresponde a ‘pareja con hijos que conviven en el hogar’, identificado con el código 4. Por el contrario, en los hogares con ocho o menos de ocho miembros se aprecia una mayor heterogeneidad en la estructura familiar, con la presencia del tipo 2, ‘padre o madre solo/a que convive con algún hijo’, así como del tipo 5, clasificado como ‘otro tipo de hogar’.

Figura 2

Relación entre el tipo de hogar y el número de miembros que lo componen.

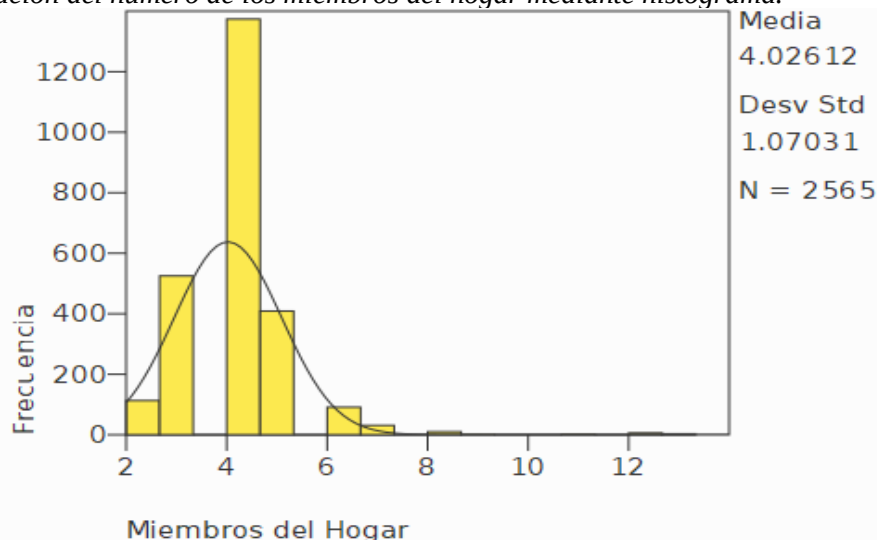


Nota: En el eje horizontal se muestra el número de miembros que componen el hogar y en el eje vertical el tipo de hogar, siendo 2, Padre o madre solo/a que conviva con algún hijo; 3, Pareja sin hijos que convivan en el hogar; 4, Pareja con hijos que convivan en el hogar; 5, Otro tipo de hogar. Fuente: INE (2024). Elaboración propia.

Como otra de las opciones de visualización que ofrece PSPP, en la Figura 3 se muestra un histograma que ilustra la distribución de la variable 'miembros del hogar', junto con la superposición de una curva normal para facilitar la interpretación de los datos. Como se puede observar, los valores más altos en la distribución corresponden a hogares con 3, 4 y 5 miembros, lo que indica que estos tamaños de familia son los más comunes entre los encuestados. Adicionalmente, PSPP junto con el histograma indica el valor de la media, la desviación estándar y la frecuencia absoluta, proporcionando una visión más detallada de los datos para el análisis.

Figura 3

Representación del número de los miembros del hogar mediante histograma.

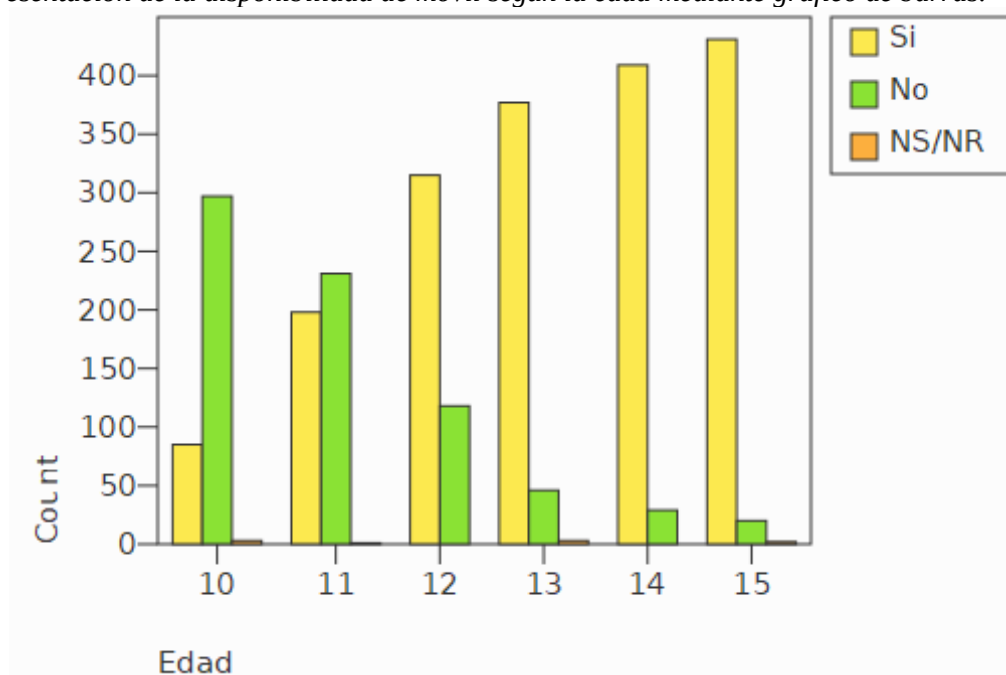


Fuente: INE (2024). Elaboración propia.

Como otra de las opciones de visualización que ofrece PSPP, se pueden realizar comparaciones entre dos variables mediante gráficos de barras. En la Figura 4 se ilustra la relación entre la variable 'edad' y la 'posesión de un teléfono móvil'. Este gráfico permite observar una tendencia clara: con el aumento de la edad, se incrementa el porcentaje de personas que poseen un teléfono móvil.

Figura 4

Representación de la disponibilidad de móvil según la edad mediante gráfico de barras.

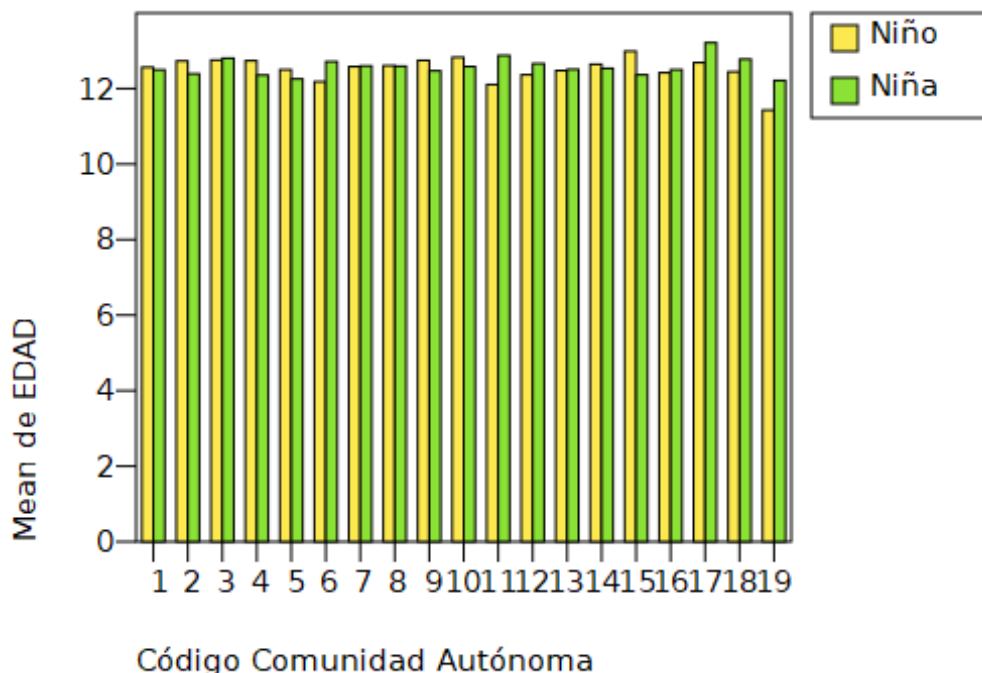


Fuente: INE (2024). Elaboración propia.

PSPP también permite la construcción de diagramas de barras que relacionan tres variables, proporcionando una visualización más detallada de los datos. En la Figura 5 se representa un diagrama de barras que muestra la variable 'edad' en el eje vertical, mientras que en el eje horizontal se presentan los códigos de las Comunidades Autónomas, clasificados por la variable 'sexo'. Este tipo de gráfico permite explorar patrones demográficos esenciales. El patrón observado sugiere que los datos recogidos se concentran en un grupo de edad específico, de manera consistente en todas las regiones. Asimismo, se aprecia una distribución homogénea tanto por edad como por sexo en las diferentes Comunidades Autónomas, lo que refleja que la muestra de datos es equilibrada en estos aspectos.

Figura 5

Representación del sexo y edad media de la muestra por comunidad autónoma mediante gráfico de barras.



Nota: 1, Andalucía; 2, Aragón; 3, Asturias; 4, Baleares; 5, Canarias; 6, Cantabria; 7, Castilla y León; 8, Castilla-La Mancha; 9, Cataluña; 10, Comunidad Valenciana; 11, Extremadura; 12, Galicia; 13, Comunidad de Madrid; 14, Murcia; 15, Navarra; 16, País Vasco; 17, La Rioja; 18, Ceuta; 19, Melilla. Fuente: INE (2024). Elaboración propia.

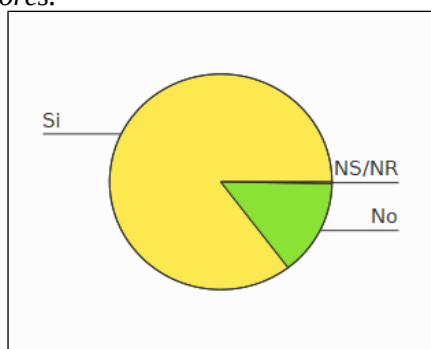
Por otro lado, PSPP permite realizar comparaciones entre variables de tipo nominal, análisis que puede ser ejecutado a través de la secuencia “Análisis estadísticos” → “Tablas cruzadas”. En esta línea, y a modo de ejemplo, en la Tabla 4 se muestra la relación entre el tipo de hogar y la disponibilidad de ordenador de los encuestados. Para este análisis, se ha concretado la variable ‘Dispone de ordenador’ en las columnas. Así mismo, las categorías de la variable ‘Tipo de hogar’ se han dispuesto por filas. Además, se ha marcado la opción “Chi-cuadrado” de la casilla “Estadísticos”, con el fin de poder determinar la existencia de asociación estadísticamente significativa entre las variables; cabe señalar que PSPP ofrece otro tipo de pruebas de contraste para el nivel de significatividad. Para las pruebas del Chi-cuadrado reportadas, se obtuvo una significación asintótica bilateral de p inferior a 0,05. Esto quiere decir que las variables ‘Tipo de hogar’ y ‘Dispone de ordenador’ no son independientes y que existe cierto grado de asociación entre ellas.

Tabla 4*Tabla cruzada del tipo de hogar según disponibilidad de ordenador.*

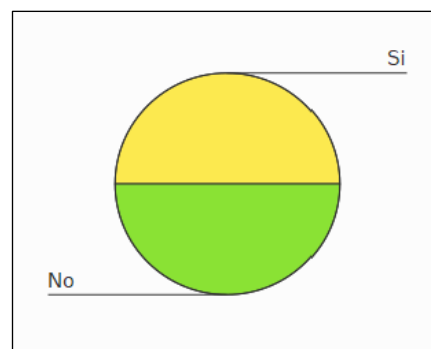
		Disponen de ordenador			Total	
		Si	No	NS/NR		
Tipo de Hogar	Padre o madre solo/a que conviva con algún hijo	Recuento	332	56	1	389
		% dentro de Tipo de Hogar	85,3%	14,4%	0,3%	100,0%
		% dentro de Disponen de ordenador	14,0%	29,6%	25,0%	15,2%
		% del total	12,9%	2,2%	0,0%	15,2%
	Pareja sin hijos que convivan en el hogar	Recuento	1	1	0	2
		% dentro de Tipo de Hogar	50,0%	50,0%	0,0%	100,0%
		% dentro de Disponen de ordenador	0,0%	0,5%	0,0%	0,1%
		% del total	0,0%	0,0%	0,0%	0,1%
	Pareja con hijos que convivan en el hogar	Recuento	1884	105	3	1992
		% dentro de Tipo de Hogar	94,6%	5,3%	0,2%	100,0%
		% dentro de Disponen de ordenador	79,4%	55,6%	75,0%	77,7%
		% del total	73,5%	4,1%	0,1%	77,7%
Otro tipo de hogar	Recuento	155	27	0	182	
	% dentro de Tipo de Hogar	85,2%	14,8%	0,0%	100,0%	
	% dentro de Disponen de ordenador	6,5%	14,3%	0,0%	7,1%	
Total	% del total	6,0%	1,1%	0,0%	7,1%	
	Recuento	2372	189	4	2565	
	% dentro de Tipo de Hogar	92,5%	7,4%	0,2%	100,0%	
	% dentro de Disponen de ordenador	100,0%	100,0%	100,0%	100,0%	
	% del total	92,5%	7,4%	0,2%	100,0%	

Fuente: INE (2024). Elaboración propia.

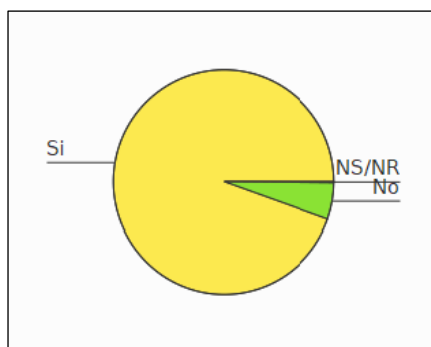
A partir de este análisis, es posible visualizar gráficamente dicha asociación mediante un diagrama. En este caso, utilizaremos un diagrama de sectores para representar los valores de la variable ‘disponibilidad de ordenador’ según ‘tipo de hogar’. Dado que PSPP no permite cruzar los valores porcentuales en un diagrama de barras, ya que este utiliza la frecuencia absoluta, ha sido necesario organizar los datos por grupos, seleccionando ‘tipo de hogar’ como variable. Para ello, se siguió la ruta “Dividir datos” dentro de la opción de “Datos”. A continuación, se procedió a generar el diagrama de sectores para la variable ‘disponibilidad de ordenador en el hogar’, obteniendo así la Figura 6.

Figura 6*Representación de la disponibilidad de ordenador según el tipo de hogar mediante diagrama de sectores.*

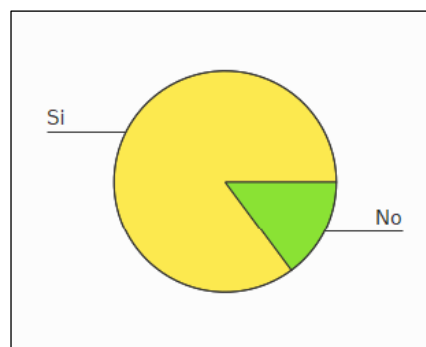
Padre o Madre solo/a que conviva con algún hijo



Pareja sin hijos que convivan en el hogar



Pareja con hijos que convivan en el hogar



Otro tipo de hogar

Fuente: INE (2024). Elaboración propia.

5. CONCLUSIONES

El análisis realizado con el software PSPP ha demostrado su utilidad para la visualización y análisis de datos en el ámbito educativo, específicamente en la enseñanza de la estadística. A lo largo de este estudio, se ha evidenciado cómo PSPP facilita la manipulación de conjuntos de datos relevantes, como la Encuesta sobre Equipamiento y Uso de TICs en menores de 10 a 15 años, proporcionando resultados claros y accesibles mediante la creación de gráficos y tablas.

Esta herramienta resulta especialmente útil en el ámbito educativo, tanto en primaria como en secundaria, ya que permite a los estudiantes trabajar con datos reales y desarrollar competencias clave del currículo de matemáticas, como la interpretación de gráficos, el análisis de proporciones y la detección de patrones. Además, este tipo de ejercicios fomenta el pensamiento crítico y el razonamiento lógico, habilidades fundamentales recogidas en los objetivos de aprendizaje de la asignatura.

Entre los resultados más destacados, la representación gráfica de variables como la edad, el tipo de hogar y la disponibilidad de dispositivos tecnológicos permitió observar patrones consistentes en la muestra analizada. Por ejemplo, se observó que los hogares más grandes tienden a estar compuestos principalmente por parejas con hijos, mientras que los hogares más pequeños presentan una mayor heterogeneidad en su composición. Asimismo, las comparaciones entre variables nominales, como la relación entre el tipo de hogar y la disponibilidad de un ordenador, evidenciaron diferencias estadísticamente significativas que reflejan la prevalencia tecnológica en ciertos tipos de hogares.

Otro aspecto importante es que el uso de PSPP en el aula facilita la enseñanza de la estadística mediante la generación de gráficos como histogramas, diagramas de barras y gráficos de sectores, que no solo permiten a los estudiantes visualizar patrones, sino también interpretar los resultados estadísticos con mayor claridad. Esto es particularmente valioso en la educación secundaria y bachillerato, donde se fomenta el análisis crítico de los datos. PSPP no solo mejora la comprensión de conceptos estadísticos, sino que también promueve habilidades críticas y analíticas en los estudiantes, preparándolos para enfrentar desafíos en el análisis de datos en el futuro.

Aunque PSPP permite generar gráficos, las opciones de personalización y presentación gráfica son limitadas lo que puede resultar un inconveniente si se necesita un alto nivel de detalle. Además, los gráficos generados en PSPP no son interactivos, y su exportación a otros

formatos presenta limitaciones, por ejemplo, la exportación de gráficos a otros formatos como PNG o PDF puede no estar disponible o ser básica. Además, PSPP permite exportar datos principalmente a formatos como CSV, TXT y HTML. Sin embargo, la exportación a formatos más complejos como Excel (XLSX) o formatos específicos de software estadístico puede no estar disponible o ser limitada. Además, al exportar a formatos como CSV o TXT, es posible que se pierdan algunos metadatos, como etiquetas de variables, formatos de datos o información sobre los valores perdidos.

A pesar de ello, PSPP cumple con los requisitos necesarios para la enseñanza básica de la estadística y la exploración de datos en niveles educativos intermedios, presentándose como una herramienta accesible y funcional para introducir a los estudiantes en el análisis de datos reales. En conclusión, los resultados del estudio sugieren que PSPP es una herramienta efectiva para facilitar la enseñanza de la estadística a través de la visualización de datos. Su interfaz intuitiva y la capacidad de generar análisis claros proporciona a los estudiantes una base sólida para el desarrollo de habilidades estadísticas, a la vez que fomenta su interés por el análisis de datos. Esto refuerza la idea de que PSPP puede ser una herramienta útil en el aula, especialmente en entornos donde se promueve el análisis práctico de datos sin la necesidad de aprender lenguajes de programación complejos.

REFERENCIAS BIBLIOGRÁFICAS

- Arteaga-Alcívar, Y., Guaña-Moya, J., Begnini-Domínguez, L., Cabrera-Córdova, M. F., Sánchez-Cali, F., y Moya-Carrera, Y. (2022). Integración de la tecnología con la educación. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E54), 182-193.
- Cabrera, R. (2022). *Análisis comparativo de dos metodologías: flipped classroom y clase magistral en una propuesta didáctica para el bloque de Estadística en 4º ESO de Matemáticas Académicas* [Zenodo]. <https://doi.org/10.5281/zenodo.13786744>
- Cabrera, R. (2022). *Manual PSPP*. [Zenodo]. <https://doi.org/10.5281/zenodo.13788409>
- Erickson, T. y Chen, E. (2021). Introducing data science with data moves and CODAP. *Teaching Statistics*, 43, S124-S132.
- Fisk G. PSPP for Beginners. 2022. <https://garyfisk.com/pspp/>
- GNU Project (2015). GNU PSPP (Versión 2.0.1) [Computer Software]. Free Software Foundation. <https://www.gnu.org/software/pspp/>
- Haldar, L. C., Wong, N., Heller, J. I. y Konold, C. (2018). Students making sense of multi-level data. *Technology Innovations in Statistics Education*, 11(1). <http://dx.doi.org/10.5070/T5111031358>
- IBM Corp. (2023). IBM SPSS Statistics for Windows (Version 29.0) [Computer software]. IBM Corp. <https://www.ibm.com/products/spss-statistics>
- Instituto Nacional de Estadística [INE] (2024). Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares, año 2023. <https://tinyurl.com/2bq2py8r>
- Mojica, G. F., Azmy, C. N. y Lee, H. S. (2019). Exploring data with CODAP. *The Mathematics Teacher*, 112(6), 473-476. <https://doi.org/10.5951/mathteacher.112.6.0473>
- Molina-Portillo, E., Contreras, J.M., Molina-Muñoz, D. y Cabrera, R. (2024). Explorando la visualización de datos. *Epsilon* 118, 9-20.
- Nelson, E. y Korey, J. *PSPP: A basic tutorial*. Social Science Research and Instructional Center; 2020. <http://ssric.org/node/699>

- Rybenská, K., Kudová, L., y Sedivy, J. (2014). Comparative analysis of the use of open source software for statistical data processing in teaching social sciences. *International Journal of Education and Information Technologies*, 8, 132-137.
- Shreiner, T. L. y Guzdial, M. (2022). The information won't just sink in: Helping teachers provide technology-assisted data literacy instruction in social studies. *British Journal of Educational Technology*, 53(5), 1134-1158. <https://doi.org/10.1111/bjet.13255>
- The Concord Consortium (2024). <https://codap.concord.org/>
- Tuva Labs Inc. (2024). *Tuva: Herramienta de análisis de datos. Tuva*. <https://tuvalabs.com>
- Yagnik, J. (2014). PSPP: A Free and Open Source Tool for Data Analysis. *Voice of Research*, 2(4), 73-76.
- Tindowen, D. J., Mendezabal, M. J., Quilang, P., y Agustin, E. T. (2019). The Use of PSPP Software in Learning Statistics. *European Journal of Educational Research*, 8(4), 1127-1136.

Visualización de datos con Power BI

José Miguel Contreras García
Universidad de Granada, jmcontreras@ugr.es

Rubén Cabrera Lozano
IES José Luis López Aranguren, rcabreralozano@educa.madrid.org

Resumen: Este estudio explora el uso de Power BI, herramienta de Business Intelligence (BI), para fomentar la alfabetización en datos en educación a través de la visualización gráfica. Power BI permite representar información mediante gráficos interactivos y dinámicos, facilitando la comprensión de conceptos complejos. Entre sus características destacadas se encuentran los filtros interactivos y la facilidad para cambiar entre distintos tipos de gráficos, lo que fomenta un aprendizaje significativo y contextualizado. Los resultados, mediante ejemplos aplicados a datos sobre acceso y uso de Tecnologías de la Información y Comunicación en menores de 12 a 15 años, muestran que Power BI es una herramienta accesible y poderosa, adecuada para estudiantes de cualquier ciclo educativo, ya que no requiere conocimientos avanzados de programación. A pesar de sus limitaciones, como la dependencia de datos estructurados y las restricciones de licencias, su implementación en el aula contribuye al desarrollo de competencias analíticas y críticas esenciales en la sociedad actual.

Palabras clave: Power BI, alfabetización en datos, visualización interactiva, educación obligatoria y bachillerato.

Data Visualization with Power BI

Abstract: This study explores the use of Power BI, a Business Intelligence (BI) tool, to promote data literacy through graphical visualization. Power BI allows information to be presented through interactive and dynamic graphics, facilitating the understanding of complex concepts. Key features include interactive filters and the ability to easily switch between different chart types, which foster meaningful and contextualized learning. The results, through examples applied to data on access to and use of Information and Communication Technologies by children aged 12 to 15, show that Power BI is an accessible and powerful tool suitable for students at any educational level as it does not require advanced programming skills. Despite its limitations, such as reliance on structured data and licensing restrictions, its implementation in the classroom contributes to the development of analytical and critical skills that are essential in today's data-driven society.

Key words: Power BI, data literacy, interactive visualization, compulsory and secondary education.

1. INTRODUCCIÓN

En la era digital actual, la alfabetización en datos se ha convertido en una competencia esencial para desenvolverse en un mundo impulsado por la información. La capacidad para interpretar, analizar y comunicar datos es cada vez más valorada en diversas disciplinas, desde la ciencia y la economía hasta las artes y las ciencias sociales. En este sentido, la educación obligatoria y Bachillerato representan etapas críticas para dotar a los estudiantes de habilidades que les

permitan comprender y utilizar datos de manera efectiva. Introducir la alfabetización en datos en el currículo no solo facilita el desarrollo del pensamiento crítico, sino que también prepara a los estudiantes para tomar decisiones fundamentadas en su vida personal y profesional (Williamson et al., 2020).

Además, la alfabetización en datos se alinea con los objetivos educativos actuales, que buscan desarrollar competencias transversales en los estudiantes, como la resolución de problemas y la capacidad de trabajo en equipo (OECD, 2019). Las actividades de análisis y visualización de datos fomentan el pensamiento lógico y el uso de herramientas digitales, habilidades que son cada vez más demandadas en el mercado laboral (Schwab y Malleret, 2020). De esta manera, introducir a los estudiantes en el análisis de datos en educación contribuye no solo a su desarrollo académico, sino también a su preparación para los retos del futuro.

Tableau y Power BI son dos herramientas de visualización de datos, del tipo Business Intelligence (BI), ampliamente utilizadas en la industria y en el ámbito académico. Su propósito principal es ayudar a los usuarios a analizar y comprender grandes volúmenes de datos mediante gráficos interactivos y dashboards que facilitan la toma de decisiones (Rafif, 2019). En el contexto educativo, estas herramientas ofrecen una plataforma visual intuitiva y potente que permite a los estudiantes explorar y comprender conjuntos de datos de manera interactiva y dinámica, lo que facilita la comprensión de conceptos abstractos en disciplinas como Matemáticas, Ciencias y Estudios Sociales (Carlisle, 2018). Esta interactividad promueve la curiosidad y el análisis exploratorio, ya que los estudiantes pueden modificar y filtrar los datos en tiempo real para descubrir patrones y tendencias. Esta experiencia de aprendizaje activo ayudará a los estudiantes a entender la importancia de la visualización de datos y a desarrollar habilidades analíticas, convirtiendo el aprendizaje en una actividad más significativa y motivadora.

Una de las razones por las que Tableau y Power BI son especialmente relevantes en el entorno educativo es su capacidad para simplificar la creación de visualizaciones de datos sin la necesidad de conocimientos avanzados de programación. La interfaz de arrastrar y soltar y los gráficos prediseñados que ambas plataformas ofrecen, permite que los estudiantes puedan centrarse en el análisis y la interpretación de los datos, en lugar de dedicar tiempo a aprender conceptos complejos de codificación (Hammad, et al., 2021). Este enfoque hace que el proceso de creación de visualizaciones sea accesible para estudiantes de diferentes niveles, fomentando un aprendizaje más inclusivo.

Asimismo, la popularidad de Tableau y Power BI en el ámbito profesional respalda su inclusión en el currículo educativo. Estas herramientas son ampliamente utilizadas en áreas como la Ciencia de Datos o la Economía, y tener conocimientos básicos sobre ellas ofrece a los estudiantes una ventaja competitiva en el futuro mercado laboral (Davenport y Patil, 2012). Al aprender a utilizar estas plataformas en un entorno educativo, los estudiantes desarrollan una familiaridad con herramientas que pueden serles útiles en su futuro profesional, ya sea en roles técnicos o no técnicos.

El objetivo de este trabajo es mostrar el potencial de las herramientas de Business Intelligence (BI) en la visualización y análisis de datos dentro del ámbito educativo, centrándose en la utilización del software Power BI. A través de su capacidad para explorar datos mediante visualizaciones interactivas y dinámicas, se busca ofrecer al profesorado de matemáticas una herramienta versátil y accesible que pueda ser aplicada en distintos niveles educativos. Power BI no solo permite representar información de manera clara y efectiva, sino que también facilita

la interacción con los datos, promoviendo un aprendizaje activo y contextualizado, que conecta conceptos matemáticos abstractos con situaciones prácticas y reales.

2. ANTECEDENTES

La incorporación de herramientas de visualización de datos como Tableau y Power BI en el ámbito educativo ha sido objeto de diversas investigaciones que destacan sus beneficios en la enseñanza y el aprendizaje. Por ejemplo, los hallazgos del estudio realizado por López-Robledo (2023) en la Universidad de Puerto Rico respecto al uso de Power BI para la visualización de datos en instituciones educativas sugiere que el desarrollo de dashboards con Power BI permite una presentación efectiva de datos, lo que facilita la toma de decisiones informadas y orienta los esfuerzos y recursos en la dirección correcta.

Por otro lado, Tableau ha sido reconocido por su capacidad para mejorar la alfabetización en datos entre los estudiantes. Según un artículo publicado por Tableau (2021), la empresa lanzó un plan de estudios sobre alfabetización de datos dirigido a la educación superior, con el objetivo de cerrar la brecha relacionada con las habilidades en el manejo de datos. Este plan de estudios proporciona a los instructores materiales que facilitan la enseñanza de competencias en análisis y visualización de datos, preparando a los estudiantes para un entorno laboral cada vez más orientado a los datos.

Además, un análisis comparativo entre Tableau y Power BI realizado por DatData (2024) destaca que ambas herramientas ofrecen interfaces intuitivas y una amplia gama de visualizaciones que pueden ser aprovechadas en contextos educativos. La elección entre una u otra dependerá de las necesidades específicas del entorno educativo y de las competencias que se deseen desarrollar en los estudiantes.

El uso de herramientas como Tableau y Power BI ha mostrado ventajas claras para el fortalecimiento de competencias matemáticas específicas. Estas plataformas permiten a los estudiantes visualizar conceptos abstractos, como funciones y modelos estadísticos, de forma concreta y dinámica, facilitando así su comprensión. Según Integratecnología (2024), la incorporación de cuadros de mando y visualizaciones en el entorno educativo optimiza no solo el aprendizaje, sino también la organización de los recursos y el monitoreo individual de los alumnos, mejorando significativamente los procesos de enseñanza y gestión académica.

Otra característica es la accesibilidad de las interfaces de Tableau y Power BI, que permite a estudiantes sin experiencia en programación involucrarse en el análisis de datos. Estas herramientas ofrecen una experiencia de usuario intuitiva, facilitando la creación de visualizaciones interactivas que promueven la exploración y comprensión de datos complejos. Según un artículo de la UNIR (2020), tanto Tableau como Power BI destacan por su facilidad de uso, incluso para personas con conocimientos limitados en analítica de datos, gracias a su interfaz sencilla e intuitiva.

En resumen, la integración de Tableau y Power BI en la educación secundaria y bachillerato no solo facilita la comprensión de conceptos matemáticos complejos, sino que también preparan a los estudiantes para enfrentar los desafíos de un mundo cada vez más orientado a la información y al análisis de datos.

3. METODOLOGÍA

Para mostrar el uso de herramientas de Business Intelligence (BI) en la visualización y análisis de datos en el ámbito educativo, este estudio adopta un enfoque exploratorio-descriptivo

centrado en el análisis de datos sobre el acceso y uso de Tecnologías de la Información y la Comunicación (TIC) entre menores de 10 a 15 años (INE, 2024). A partir de los datos presentados en el primer estudio del monográfico, que proporciona una descripción detallada de fuentes, selección de casos, variables y resultados, este apartado explicará los métodos de análisis y visualización empleados.

Aunque Tableau ofrece una licencia gratuita de un año para estudiantes y profesores, Tableau (2024), <https://www.tableau.com/es-es/academic/teaching>, permite un acceso completo a su plataforma. En este trabajo nos centraremos en Power BI debido a la limitación de espacio y a que dispone de una plataforma online que permite trabajar sin tener que instalar el software, <https://app.powerbi.com>.

En el desarrollo de este estudio en Power BI, inicialmente se describen los pasos para cargar el conjunto de datos y realizar una limpieza básica (si es necesario), además de proporcionar una breve explicación de cada variable relevante y mostrar la potencialidad gráfica de este software. Los datos analizados proceden de la Encuesta sobre Equipamiento y Uso de TICs (INE, 2024), que abarca información sobre el acceso a tecnologías en el hogar y su uso entre menores, según el número de niños en la familia y la distribución geográfica en España, desglosada por comunidades y ciudades autónomas.

Para representar visualmente la información en Power BI, se han seleccionado gráficos que permiten explorar y describir los datos desde múltiples perspectivas:

- Gráfico de barras: adecuado para comparar el acceso a TIC entre comunidades autónomas y el número de menores en el hogar.
- Gráfico de líneas: permite observar tendencias o variaciones en el uso de TIC a lo largo del tiempo o por regiones.
- Gráfico de sectores: muestra la proporción de hogares con acceso a diferentes tecnologías, facilitando una visión general de la distribución.
- Histograma: ayuda a comprender la frecuencia de uso de TIC en diversas franjas de edad.
- Gráfico de dispersión: útil para analizar la relación entre el uso de TIC y variables como el número de menores en el hogar o la comunidad autónoma.
- Diagrama de cajas y bigotes: proporciona una visión de la variabilidad y distribución en el acceso a TIC por región.

Cada visualización en Power BI se acompaña de una pequeña guía explicativa que detalla los pasos para ajustar los parámetros de cada gráfico y configurar opciones de visualización. Además, se explican las opciones de interacción de Power BI, que permiten a los estudiantes filtrar y explorar datos en tiempo real, mejorando así la comprensión de los patrones y tendencias.

4. RESULTADOS

En este trabajo, nos enfocaremos en algunas de las características más destacadas de Power BI, en su versión gratuita, como su capacidad para realizar filtros interactivos de datos y su facilidad para cambiar entre diferentes tipos de gráficos. Estas funcionalidades no solo permiten a los estudiantes explorar de manera dinámica los conjuntos de datos, sino que también fomentan un aprendizaje más significativo y contextualizado.

El uso de filtros interactivos posibilita a los alumnos analizar información específica en tiempo real, identificando patrones y tendencias que de otro modo podrían pasar desapercibidos. Por su parte, la flexibilidad para cambiar el tipo de gráfico les ayuda a comprender qué

representaciones visuales son más adecuadas según el tipo de datos que estén manejando. Esto no solo fortalece su capacidad para interpretar y comunicar datos, sino que también los guía hacia una selección más precisa y fundamentada de las herramientas de visualización.

Además, estas características facilitan la extracción de información relevante de manera rigurosa y eficiente, promoviendo el pensamiento crítico y el análisis detallado. Los estudiantes no solo aprenden a visualizar datos, sino también a cuestionar y reflexionar sobre lo que los datos significan en diferentes contextos.

Desde el primer momento, Power BI nos permite interactuar con los datos de manera intuitiva. Por ejemplo, al seleccionar el archivo de trabajo, la herramienta proporciona una vista previa detallada de su contenido. Esta funcionalidad es particularmente útil para que los usuarios puedan verificar rápidamente la estructura de los datos, identificar posibles inconsistencias o errores, y decidir cómo proceder en el análisis.

La vista previa, Figura 1, incluye información como los nombres de las columnas, los tipos de datos detectados automáticamente y una muestra de las filas. Esto permite realizar ajustes iniciales, como cambiar delimitadores, corregir errores en la codificación de caracteres o modificar tipos de datos incorrectamente identificados. Además, Power BI ofrece la posibilidad de realizar transformaciones básicas en esta etapa, como la eliminación de columnas innecesarias o la división de celdas.

Estas primeras interacciones no solo facilitan la limpieza de los datos, sino que también establecen una base sólida para el análisis posterior. Al brindar un entorno visual amigable, Power BI elimina barreras técnicas y permite a los usuarios centrarse en el propósito de los datos, promoviendo un análisis más efectivo y organizado desde el inicio del proyecto. Este enfoque es particularmente valioso en entornos educativos, donde los estudiantes pueden comenzar a familiarizarse con los conjuntos de datos antes de profundizar en el análisis y la visualización.

Figura 1

Vista previa de los datos del ejemplo

ID	CODAUTO	Comunidad	CPRO	Provincia	SERIAL	TIP_H	TOT_MH	TOT10_15	ORD	TABLET	TELEF1	TELEF2	VIV_INTER	INTERFOR	NPER	SEXO	EDAD	PC
01000030...	16	País Vasco	1	Araba/Alava	3	4	4	1	6	1	6	1	1	1	1	3	6	12
01000300...	16	País Vasco	1	Araba/Alava	30	4	3	1	1	1	6	1	1	1	1	3	6	15
01001000...	16	País Vasco	1	Araba/Alava	100	4	4	2	1	1	1	1	1	1	1	3	6	15
01001000...	16	País Vasco	1	Araba/Alava	100	4	4	2	1	1	1	1	1	1	1	4	6	12
01001020...	16	País Vasco	1	Araba/Alava	102	4	4	1	1	1	1	1	1	1	1	3	1	11
01001100...	16	País Vasco	1	Araba/Alava	110	4	3	1	1	1	6	1	1	1	1	3	1	13
01001140...	16	País Vasco	1	Araba/Alava	114	4	5	1	1	6	1	1	1	2	3	1	12	9
01001180...	16	País Vasco	1	Araba/Alava	118	2	3	2	1	1	1	1	1	1	2	1	10	1
01001180...	16	País Vasco	1	Araba/Alava	118	2	3	2	1	1	1	1	1	1	3	1	10	1
01001210...	16	País Vasco	1	Araba/Alava	121	4	4	1	1	6	1	1	1	1	1	4	6	12
01001260...	16	País Vasco	1	Araba/Alava	126	4	3	1	1	1	1	1	1	1	1	3	6	12
01001320...	16	País Vasco	1	Araba/Alava	132	4	6	2	1	6	6	1	1	1	1	3	1	14
01001320...	16	País Vasco	1	Araba/Alava	132	4	6	2	1	6	6	1	1	1	1	4	1	14
01001360...	16	País Vasco	1	Araba/Alava	136	3	2	1	1	1	6	1	1	1	1	2	6	12
01001370...	16	País Vasco	1	Araba/Alava	137	4	4	1	1	1	1	1	1	1	1	2	1	11
01001380...	16	País Vasco	1	Araba/Alava	138	2	3	1	1	1	1	1	1	1	1	2	1	13
02001430...	8	Castilla-La Mancha	2	Albacete	143	4	4	1	1	6	1	1	1	1	1	3	6	11
02001470...	8	Castilla-La Mancha	2	Albacete	147	4	4	1	1	1	6	1	1	1	1	4	6	14
02001530...	8	Castilla-La Mancha	2	Albacete	153	4	4	1	1	1	6	1	1	2	3	6	12	1
02001550...	8	Castilla-La Mancha	2	Albacete	155	4	4	1	1	1	1	1	1	1	1	3	1	11
02001560...	8	Castilla-La Mancha	2	Albacete	156	5	5	2	1	1	1	6	1	1	1	3	6	13
02001560...	8	Castilla-La Mancha	2	Albacete	156	5	5	2	1	1	1	6	1	1	1	4	6	13
02001570...	8	Castilla-La Mancha	2	Albacete	157	2	3	2	1	1	1	1	1	1	1	2	6	15
02001570...	8	Castilla-La Mancha	2	Albacete	157	2	3	2	1	1	1	1	1	1	1	3	1	13

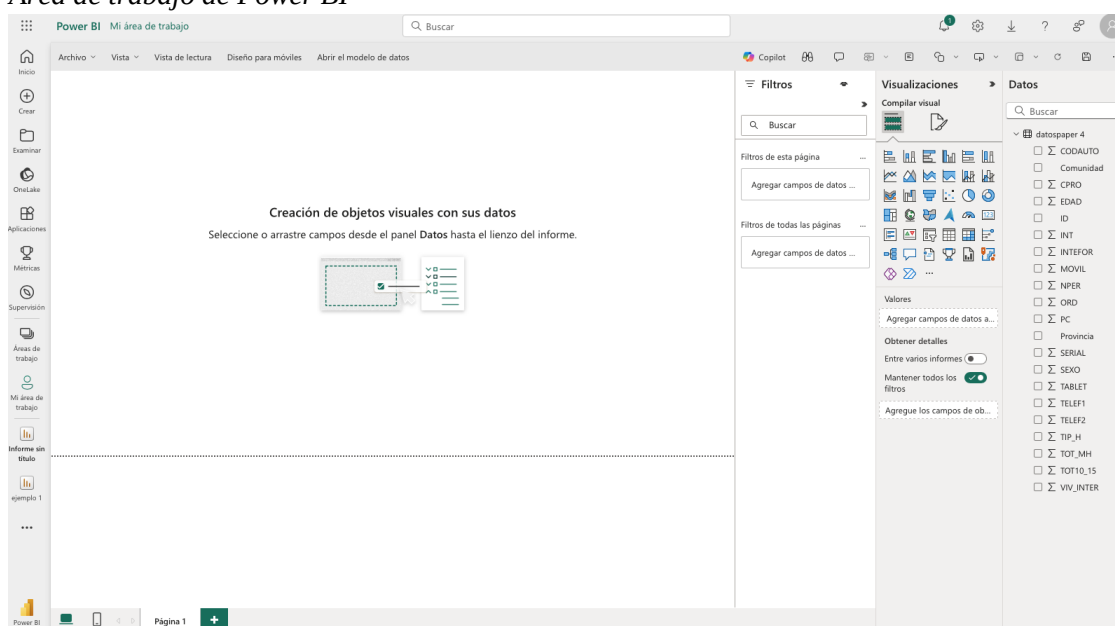
Fuente: *Encuesta sobre Equipamiento y Uso de TIC (año 2023)*. INE (2024a). Elaboración propia.

Para comenzar, trabajamos desde el área de trabajo de Power BI, representada en la Figura 2. Este entorno visual ha sido diseñado específicamente para simplificar la creación y gestión de informes y visualizaciones interactivas. En esta área, los usuarios tienen la posibilidad de construir gráficos y tablas dinámicas basados en los datos previamente cargados, utilizando herramientas intuitivas como el panel de visualizaciones y el selector de campos.

La interfaz permite arrastrar y soltar elementos de manera sencilla, configurar filtros dinámicos, ajustar parámetros de diseño y personalizar cada visualización según las necesidades del análisis. Además, la organización clara y lógica del área de trabajo facilita la navegación y el flujo de tareas, optimizando el tiempo y reduciendo la curva de aprendizaje.

Figura 2

Área de trabajo de Power BI

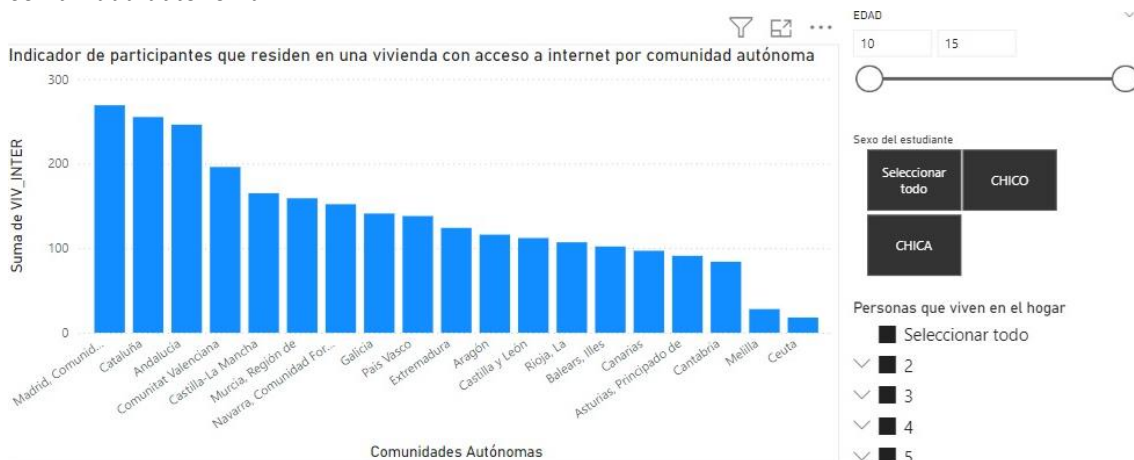


Como ejemplo práctico, hemos generado un diagrama de barras de la variable "Indicador de viviendas con acceso a Internet" en función de la "comunidad autónoma" de los estudiantes, Figura 3. Este gráfico, pone de manifiesto una de las principales virtudes de Power BI: su capacidad de interactividad.

En esta visualización, se han incorporado diversos filtros para personalizar el análisis y explorar los datos de manera más detallada. En primer lugar, se utiliza un control deslizante (slider) que permite filtrar los datos por edad de los estudiantes, ajustándose dinámicamente al rango seleccionado. Además, se incluye un filtro para segmentar los datos por sexo del estudiante y otro para analizar la información en función del número de integrantes del hogar. Estas herramientas no solo permiten una exploración más profunda de los datos, sino que también facilitan la identificación de patrones y relaciones significativas entre las variables.

Figura 3

Representación Indicador de participantes que residen en viviendas con acceso a internet según comunidad autónoma

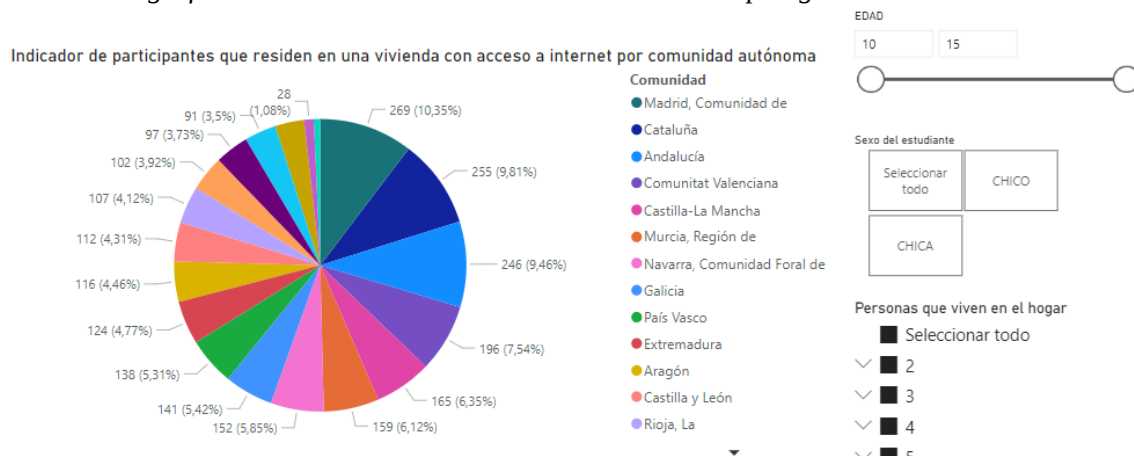


Fuente: *Encuesta sobre Equipamiento y Uso de TIC (año 2023)*. INE (2024a). Elaboración propia.

La potencialidad de Power BI para la enseñanza radica, entre otras cosas, en la facilidad para cambiar entre diferentes tipos de gráficos. Por ejemplo, con tan solo copiar la página (utilizando la pestaña inferior del área de trabajo) y arrastrar un nuevo tipo de visualización desde el panel de visualizaciones sobre el gráfico existente, se puede generar una representación completamente diferente de los mismos datos (Figura 4).

Figura 4

Alternativa gráfica a la visualización anterior cambiando de tipología



Fuente: *Encuesta sobre Equipamiento y Uso de TIC (año 2023)*. INE (2024a). Elaboración propia.

Esta funcionalidad no solo fomenta la flexibilidad en el análisis de datos, sino que también permite a los estudiantes explorar cuál es el tipo de gráfico más adecuado para representar la información de manera clara y efectiva. Además, esta capacidad para alternar entre diferentes visualizaciones ofrece una oportunidad para que los estudiantes desarrollen habilidades

analíticas al comparar cómo distintos enfoques gráficos influyen en la interpretación de los datos. Este proceso no solo enriquece su aprendizaje, sino que también promueve un análisis más profundo y crítico de la información presentada.

Otro aspecto destacado de Power BI es su herramienta de filtrado, que permite a los usuarios "interactuar" con los datos de manera dinámica. Esta funcionalidad facilita explorar cómo las variaciones en el contexto afectan la información presentada, ofreciendo una visión más completa y significativa del análisis.

Figura 5

Alternativa gráfica a la visualización anterior aplicando filtros



Fuente: *Encuesta sobre Equipamiento y Uso de TIC (año 2023)*. INE (2024a). Elaboración propia.

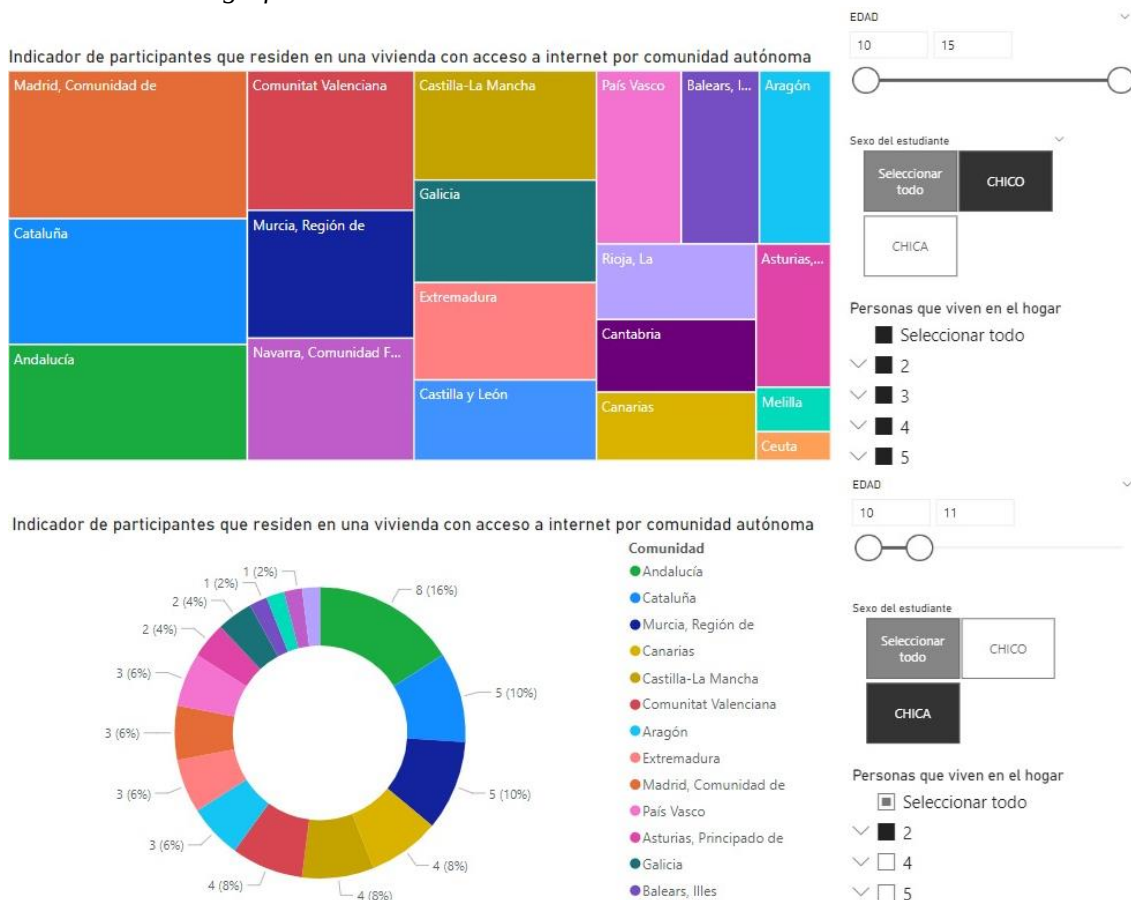
Por ejemplo, al aplicar filtros específicos (Figura 5), como rango de edades, género o cualquier otra variable categórica o cuantitativa, los usuarios pueden observar en tiempo real cómo cambian los resultados y las visualizaciones en función de estas elecciones. Esta capacidad interactiva no solo enriquece la experiencia del usuario, sino que también promueve un aprendizaje activo al incentivar a los estudiantes a plantear preguntas y analizar cómo diferentes contextos impactan en la interpretación de los datos.

En definitiva, permitir que los estudiantes interactúen activamente con los datos, explorando diferentes tipologías de gráficos —ya sea buscando representaciones más actuales, vistosas o llamativas para ellos—, combinado con el uso de filtros para modificar el contexto de los datos, puede ser de gran valor en su alfabetización en datos (Figura 6). Este enfoque no solo fomenta la creatividad y el pensamiento crítico, sino que también les ayuda a comprender cómo la forma en que se presentan los datos puede influir en su interpretación y en la toma de decisiones.

Además, estas actividades promueven una conexión más significativa entre los estudiantes y los datos que analizan, incentivándolos a plantear hipótesis, explorar patrones y descubrir relaciones ocultas en los conjuntos de datos. Este proceso interactivo no solo fortalece sus habilidades analíticas, sino que también contribuye al desarrollo de competencias transversales como la resolución de problemas, el razonamiento lógico y la capacidad de comunicar de manera efectiva conclusiones basadas en datos. En un mundo cada vez más impulsado por la información, estas son habilidades esenciales que prepararán a los estudiantes para enfrentar los desafíos del futuro con una base sólida en el análisis y la visualización de datos.

Figura 6

Alternativas a los gráficos tradicionales



Fuente: *Encuesta sobre Equipamiento y Uso de TIC (año 2023)*. INE (2024a). Elaboración propia.

5. CONCLUSIONES

Este estudio demuestra cómo la utilización de Power BI puede contribuir significativamente a la promoción de la alfabetización en datos en el ámbito educativo. El desarrollo de competencias relacionadas con la lectura de gráficos, la interpretación de estadísticas y el análisis de tendencias no solo fortalece el conocimiento técnico de los estudiantes, sino que también fomenta el análisis crítico de la información. Estas habilidades son esenciales en un mundo donde los datos ocupan un lugar central en la toma de decisiones informadas (Donoho, 2017).

En el contexto educativo, fomentar la alfabetización en datos permite a los estudiantes abordar temas complejos de una manera más accesible y significativa, conectando los aspectos teóricos con situaciones prácticas de la vida real. Este enfoque promueve un aprendizaje profundo, contextualizado y motivador (Gould, 2021). En este sentido, Power BI ofrece una plataforma intuitiva y poderosa que no requiere habilidades avanzadas de programación, pero que permite la creación de representaciones gráficas avanzadas y dinámicas. Su facilidad de uso, que se asemeja al manejo de software cotidiano, lo hace especialmente adecuado para estudiantes de Educación Secundaria, quienes ya cuentan con las habilidades digitales básicas necesarias para su manipulación.

Además, Power BI destaca por su capacidad de interactividad, que no solo enriquece el análisis de datos, sino que también incentiva a los estudiantes a explorar, plantear hipótesis y evaluar diferentes representaciones gráficas. Estas actividades promueven la conexión activa entre los datos y su contexto, desarrollando habilidades analíticas, críticas y creativas que son fundamentales en la sociedad actual.

Aunque el manejo e interpretación de datos se han convertido en herramientas indispensables para la resolución de problemas y la participación en una sociedad basada en la información, es importante señalar algunas limitaciones. Una de ellas es la dependencia de conjuntos de datos previamente estructurados y procesados, lo que puede limitar la capacidad de los estudiantes para trabajar con datos en bruto y enfrentarse a desafíos reales de limpieza y organización. Asimismo, aunque Power BI es accesible en términos de facilidad de uso, y dispone de una versión gratuita, su versión más completa requiere licencias, lo que podría representar una barrera en algunos contextos educativos con recursos limitados.

A pesar de estas limitaciones, el enfoque promovido en este trabajo no solo satisface las expectativas del mercado laboral, que demanda individuos con habilidades analíticas, sino que también contribuye a formar ciudadanos reflexivos, responsables y críticos. Al interactuar de manera activa con los datos y evaluar su validez y relevancia, los estudiantes desarrollan una visión más informada y comprometida sobre los desafíos que enfrentan en su vida diaria y futura.

REFERENCIAS BIBLIOGRÁFICAS

- Carlisle, S. (2018). Software: Tableau and Microsoft Power BI. *Technology, Architecture + Design*, 2(2), 256-259.
- DatData (2024). Tableau vs Power BI: Para Analistas de Datos. Recuperado de <https://www.datdata.com/blog/tableau-vs-power-bi-para-analistas-de-datos>.
- Donoho, D. (2017). 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, 26(4), 745-766.
- Davenport, T. H. y Patil, D. J. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, 90(10), 70-76.
- Gould, R. (2021). Toward data-scientific thinking. *Teaching Statistics*, 43, S11-S22.
- Hamad, M. T. J., Yassin, M. M., Shaban, O. S. y Amoush, A. H. (2023, May). Using Business Intelligence Tools in Accounting Education. In *Conference on Sustainability and Cutting-Edge Business Technologies* (pp. 163-177). Cham: Springer Nature Switzerland.
- INE (2024). *Encuesta sobre equipamiento y uso de tecnologías de la información y comunicación en los hogares*. Instituto Nacional de Estadística. Recuperado de <https://www.ine.es/>
- Integratecnología (2024). *Power BI y el uso de la visualización de datos en el sector educativo*. Integratecnología. Recuperado de <https://www.integratecnologia.es/la-innovacion-necesaria/power-bi-y-el-uso-de-la-visualizacion-de-datos-en-el-sector-educativo/>
- López-Robledo, D. M. (2023). Power BI para la visualización de datos en instituciones educativas. *HETS Online Journal*, 13(2), 6-22.
- OECD (2019). *OECD Future of Education and Skills 2030*. OECD Publishing.
- Rafif, M. F. (2019). Comparison study of the development of Self-Service Business Intelligence between Power BI and Tableau desktop (Case study: Microsoft Adventure Works).

- Schwab, K. y Malleret, T. (2020). *COVID-19: The Great Reset*. Forum Publishing.
- Tableau (2021). Plan de estudios sobre alfabetización de datos de Tableau para educación superior. Recuperado de <https://www.tableau.com/es-es/blog/announcing-new-tableau-data-literacy-curriculum-higher-education>.
- UNIR (2020). *Power BI vs Tableau: ¿Cuál es la mejor herramienta de visualización de datos?* UNIR Revista. Recuperado de <https://www.unir.net/revista/ingenieria/poweb-bi-vs-tableau/>
- Williamson, B., Eynon, R. y Potter, J. (2020). Pandemic politics, pedagogies and practices: Digital technologies and distance education during the coronavirus emergency. *Learning, Media and Technology*, 45(2), 107-114.

Exploración y visualización de datos en Educación Matemática con Jupyter y Python

José Miguel Contreras García
Universidad de Granada, jmcontreras@ugr.es

Cristina Hidalgo Castillo
Universidad de Granada, chidcas@correo.ugr.es

Resumen: Este artículo explora el uso de Python a través de Jupyter Notebook como herramienta educativa para la visualización y análisis de datos en el ámbito de la enseñanza de matemáticas. A través de ejemplos prácticos, se muestra cómo Jupyter permite crear gráficos variados (barras, líneas, sectores, histogramas, entre otros) usando bibliotecas de Python como Matplotlib y Seaborn, facilitando el análisis exploratorio de datos. Dirigido a profesores, el artículo destaca la interactividad de Jupyter para que los estudiantes puedan experimentar y desarrollar habilidades en programación y visualización. Así, Jupyter se presenta como una plataforma accesible y eficaz para promover la alfabetización de datos en educación obligatoria.

Palabras clave: Visualización de datos, Jupyter Notebook, Python, educación matemática, alfabetización en datos, aprendizaje interactivo.

Data Exploration and Visualization in Mathematics Education with Jupyter and Python

Abstract: This study examines the use of Python through Jupyter Notebook as an educational tool for data visualization and analysis in mathematics education. Through practical examples, it illustrates how Jupyter enables the creation of diverse graphical representations (bar, line, pie charts, histograms, among others) using Python's libraries such as Matplotlib and Seaborn, thereby facilitating exploratory data analysis. Aimed at educators, the article focuses on Jupyter's interactive features, which allow students to experiment and develop skills in programming and visualization. Consequently, Jupyter is positioned as an accessible and effective platform to foster data literacy in compulsory education.

Key words: Data visualization, Jupyter Notebook, Python, mathematics education, data literacy, interactive learning.

1. INTRODUCCIÓN

La visualización de datos es una habilidad fundamental en la alfabetización de datos. Para comprender un conjunto de datos, es crucial explorar y conocer en profundidad las variables que lo componen. Aunque no se espera que un estudiante sea un experto en los datos que maneja, siempre será necesario investigarlos a fondo antes de realizar cualquier otro análisis. La visualización de datos es la herramienta más importante para lograr esto, por lo que es una habilidad esencial que debe dominarse en la educación obligatoria.

Jupyter con Python es una herramienta poderosa para la visualización y análisis de datos que combina un entorno interactivo de desarrollo con el lenguaje de programación Python.

Cómo citar: Contreras García, J. M. y Hidalgo Castillo, C. (2024). Exploración y Visualización de Datos en Educación Matemática con Jupyter y Python. *Epsilon*, 118, 81-89.

Jupyter Notebook es un software de código abierto que permite a los usuarios crear y compartir documentos que contienen código en vivo, ecuaciones, visualizaciones y texto narrativo. Esta herramienta es ampliamente utilizada en educación, ciencia de datos e investigación, y es gratuita para todos los usuarios. Desarrollada como parte del Proyecto Jupyter, esta plataforma es compatible con numerosos lenguajes de programación (aunque Python es el más común) y proporciona un entorno intuitivo para la exploración de datos, el desarrollo de modelos y la documentación de procesos de análisis.

Jupyter Notebook es una herramienta poderosa para el análisis de datos interactivo, que permite a los usuarios crear gráficos y visualizaciones dinámicas que facilitan la comprensión de patrones y tendencias en los datos. Su interfaz, accesible desde un entorno web, es fácil de usar y altamente compatible con bibliotecas populares de Python como Matplotlib, Pandas, y Seaborn, lo que la hace ideal tanto para principiantes como para profesionales en ciencia de datos. Además, Jupyter se integra fácilmente con otras herramientas y plataformas de desarrollo, ofreciendo un ecosistema robusto para el análisis de datos y la creación de contenido interactivo.

En el ámbito educativo, Jupyter resulta particularmente valioso para los profesores (Kluyver et al., 2016), ya que permite la preparación de materiales didácticos interactivos y la inclusión de ejemplos prácticos que los estudiantes pueden explorar y modificar de forma autónoma, todo dentro del mismo entorno. Su diseño, basado en celdas de código y texto, permite a los educadores combinar explicaciones teóricas con ejercicios prácticos, promoviendo un enfoque de aprendizaje activo. Esto no solo enriquece el proceso de enseñanza, sino que también brinda a los estudiantes una experiencia práctica en programación, análisis de datos y visualización, fomentando su capacidad para experimentar y aprender de manera independiente y personalizada.

En el contexto de la visualización de datos, Jupyter se integra de manera eficiente con bibliotecas como Matplotlib, Plotly y Seaborn, permitiendo a los usuarios generar gráficos y visualizaciones directamente en la aplicación. Estos gráficos pueden variar desde representaciones simples, como histogramas o diagramas de dispersión, hasta visualizaciones más complejas e interactivas en 3D, facilitando una exploración visual más profunda. Estas capacidades de visualización son esenciales para el análisis exploratorio de datos (EDA, por sus siglas en inglés), ya que ayudan a identificar patrones, relaciones y anomalías dentro de los conjuntos de datos.

Siguiendo el enfoque del documento de referencia de la primera sección, titulado “Explorando la visualización de datos” (Molina-Portillo et al., 2024), este trabajo tiene como objetivo ofrecer a los profesores de matemáticas de las etapas de Secundaria y Bachillerato una guía práctica sobre el uso de Python mediante Jupyter Notebook para explorar datos mediante su visualización. Se presenta una selección de ejemplos que ilustran cómo adaptar la complejidad de las visualizaciones y el análisis a diferentes niveles de demanda cognitiva y habilidades digitales, destacando las posibilidades de esta herramienta para enriquecer la enseñanza de matemáticas y el análisis visual de datos.

2. ANTECEDENTES

El uso de Jupyter Notebook en educación ha sido ampliamente documentado, destacando su capacidad para facilitar un aprendizaje interactivo, a través de sus notebooks, y reproducible en diversas disciplinas, especialmente en matemáticas y ciencias de la computación.

Por ejemplo, Barba et al. (2019) describen el uso de Jupyter Notebook como una herramienta educativa que puede mejorar el aprendizaje en diversas disciplinas, incluyendo matemáticas, debido a su naturaleza interactiva. Los autores destacan cómo los notebooks permiten a los estudiantes y profesores combinar explicaciones teóricas con código ejecutable y visualizaciones, creando un entorno de aprendizaje dinámico y exploratorio, particularmente útil en la educación secundaria y superior.

Granger y Pérez (2021) también exploran el papel de Jupyter en la educación, subrayando cómo su capacidad de integrar código y texto ha sido beneficiosa para la enseñanza de materias como Matemáticas y Ciencia de Datos. Argumentan que el enfoque interactivo de Jupyter permite a los estudiantes entender mejor los conceptos teóricos a través de la experimentación práctica, facilitando un entorno de aprendizaje activo y personalizado.

En un estudio más específico, Gajdoš et al. (2022) analizan el uso de Jupyter Notebook en combinación con SageMath para la enseñanza de contenidos matemáticos. Este trabajo demuestra la efectividad de los notebooks como herramienta educativa al permitir a los estudiantes explorar y visualizar conceptos matemáticos complejos de manera interactiva, promoviendo así una comprensión más profunda.

Desde una perspectiva de ciencia de datos, Perkel (2018) en *Nature* resalta que Jupyter Notebook ha emergido como la herramienta preferida por los científicos de datos debido a su capacidad para soportar la computación interactiva. Este artículo destaca que los notebooks no solo facilitan la exploración de datos, sino que también permiten la comunicación clara de resultados a través de una combinación de texto explicativo y código ejecutable. Esto también se aplica al ámbito educativo, en el que los notebooks se utilizan para enseñar conceptos matemáticos y de programación de manera visual e interactiva.

La arquitectura de Jupyter, que se centra en la interactividad y la facilidad de uso, también ha sido objeto de atención en la literatura. Ragan-Kelley et al. (2014) discuten cómo el diseño de Jupyter permite un enfoque unificado en la investigación computacional, desde la exploración interactiva hasta la comunicación y publicación de resultados. Esta capacidad ha sido fundamental para su adopción en entornos educativos, en los que se valora la capacidad de los estudiantes para explorar y manipular conceptos matemáticos y científicos de manera práctica.

Rule et al. (2018) exploran cómo los notebooks computacionales, incluidos los Jupyter Notebook, facilitan tanto la exploración como la explicación en el aprendizaje. Destacan su utilidad en la enseñanza de programación y matemáticas al integrar código, visualizaciones y narrativas que permiten una comprensión más profunda y dinámica de conceptos complejos.

Finalmente, en Thompson y Mills (2020) se utiliza Python a través de Jupyter Notebooks en un aula de secundaria para la exploración de datos. En el estudio se muestra que el uso de Python puede ser útil a la hora de visualizar y analizar datos reales. Sin embargo, concluye que en este nivel educativo se debe utilizar de manera guiada ya que los estudiantes pueden llegar a sentirse confusos dado que se trata de un lenguaje de programación al que no suelen estar acostumbrados.

Estos estudios proporcionan una muestra sobre los beneficios de usar Jupyter Notebook en la educación, especialmente en matemáticas y programación. Su capacidad para integrar contenido diverso y su enfoque en la interactividad y reproducibilidad hacen de Jupyter una herramienta poderosa para educadores y estudiantes, promoviendo un aprendizaje más profundo y significativo.

Para este estudio tomamos en consideración la fundamentación teórica descrita en Molina-Portillo et al. (2024)

3. METODOLOGÍA

Para ilustrar el uso del lenguaje de programación Python en la visualización y análisis de datos, este estudio adopta un enfoque exploratorio-descriptivo centrado en el análisis del uso de Tecnologías de la Información y la Comunicación (TIC) entre menores de 10 a 15 años (INE, 2024). Basándonos en los datos presentados en el primer estudio del monográfico, que ofrece una descripción detallada de las fuentes, selección de casos, variables y resultados documentados, este apartado se enfoca en explicar los métodos de análisis y visualización empleados.

El trabajo se lleva a cabo en Jupyter Notebook, una plataforma que facilita la combinación de código, gráficos y documentación en un único entorno. En el notebook, se incluyen inicialmente instrucciones sobre cómo cargar el conjunto de datos, junto con una breve descripción de cada variable relevante. Los datos analizados proceden de la Encuesta sobre Equipamiento y Uso de TIC (2023), y contienen información sobre el acceso y uso de TIC en el hogar por menores, según el número de niños en la familia y su distribución geográfica en España, desglosada por comunidades y ciudades autónomas.

Para representar visualmente la información, se han seleccionado diversos tipos de gráficos que permiten explorar y describir los datos desde diferentes ángulos y niveles de complejidad.

Se muestran ejemplos de los siguientes tipos de gráficos:

- Gráfico de barras: ideal para mostrar comparaciones entre comunidades autónomas y el acceso a TIC según la cantidad de menores en el hogar.
- Gráfico de líneas: útil para observar tendencias en el tiempo o variaciones por regiones en el uso de diferentes tipos de TIC.
- Gráfico de sectores (o de pastel): se emplea para representar la proporción de hogares con acceso a diferentes tecnologías, facilitando una vista de la distribución general.
- Histograma: ayuda a comprender la frecuencia de uso de TIC en diferentes franjas de edad.
- Gráfico de dispersión: utilizado para analizar relaciones entre variables, como el uso de TIC frente al número de menores por hogar o por comunidad.
- Diagrama de cajas y bigotes: proporciona una visión detallada de la variabilidad y distribución de los datos de acceso a TIC por región.

Cada visualización se acompaña de una explicación detallada sobre el uso de bibliotecas de Python específicas, como Matplotlib, Pandas y Seaborn, adaptadas al tipo de gráfico en cuestión. Estas explicaciones guían al lector en el proceso de creación de cada gráfico, ajustando los parámetros y configuraciones para obtener representaciones visuales claras y precisas.

Para fomentar el aprendizaje activo, cada sección incluye una tarea práctica que invita al lector a aplicar los conceptos aprendidos y realizar ajustes adicionales a las visualizaciones. A continuación de cada tarea, se presenta una solución detallada que facilita la comprensión del proceso y refuerza las habilidades prácticas en el uso de Jupyter Notebook y sus bibliotecas para la visualización de datos.

Dada la limitación del texto, toda la descripción de los ejemplos se realiza en el ejemplo virtual, describiendo solo una parte de uno de ellos en este trabajo.

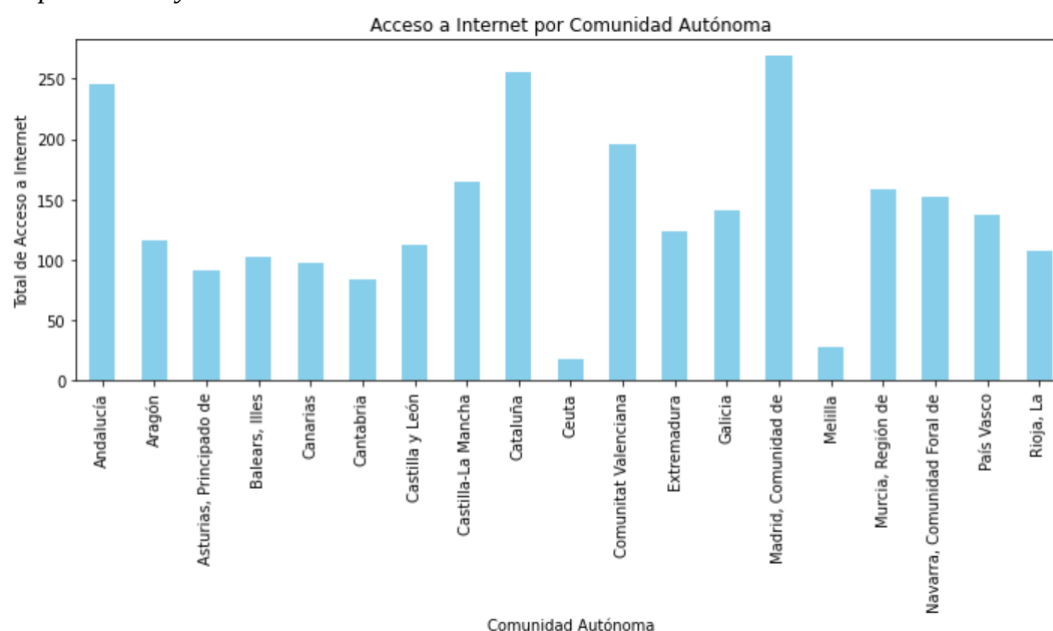
4. RESULTADOS

En los resultados se utiliza Jupyter Notebook para la visualización de los distintos ejemplos diseñados para mostrar la potencialidad de éste, que se pueden encontrar en el enlace <https://github.com/jmcontrerasgarcia/EDUCACION-EN-VISUALIZACION-DE-DATOS>. Al utilizar un lenguaje de programación algo complejo, como es Python, este notebook se enfoca a partir de la etapa de Secundaria. Los gráficos que se utilizan son el gráfico de barras, gráfico de líneas, gráfico de sectores, histograma, gráfico de dispersión y diagrama de cajas y bigotes. Estos recogen los datos de la Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares para el año 2023. A continuación, mostramos algunos ejemplos que se pueden realizar usando este software. Dada la limitación de espacio, mostramos un solo ejemplo de gráfico (los demás ejemplos los puede encontrar en el anterior enlace).

En la Figura 1, se muestra un gráfico de barras creado con la biblioteca Matplotlib en Python, el cual representa el acceso a Internet por comunidad autónoma.

Figura 1

Gráfico de barras que representa el acceso a Internet por comunidad autónoma en la librería Matplotlib de Python

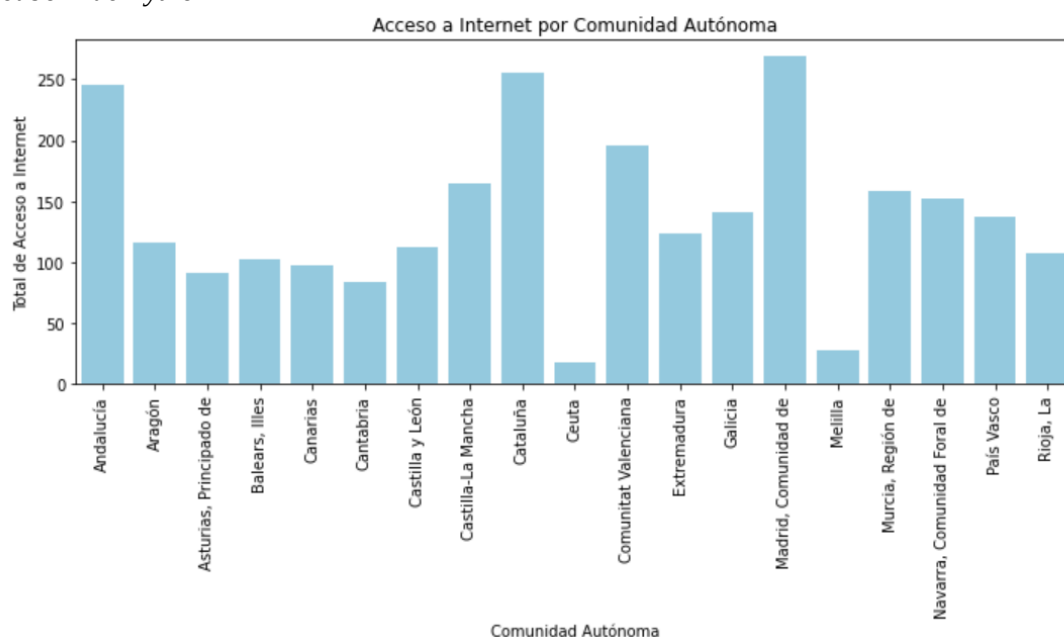


Fuente: elaboración propia

La Figura 2 presenta el mismo gráfico, pero utilizando la biblioteca Seaborn. Aunque ambas herramientas generan gráficos similares, Seaborn ofrece ciertas ventajas, como estilos predeterminados y una mayor simplicidad al gestionar datos. Mientras que Matplotlib requiere configuraciones manuales para ciertos efectos visuales, Seaborn facilita la creación de gráficos estéticamente agradables con menos líneas de código, gracias a sus paletas de colores predefinidas y su mejor integración con estructuras de datos de Pandas. Ambas bibliotecas son útiles, pero Seaborn puede resultar más conveniente para quienes buscan una presentación visual más atractiva con un código más conciso.

Figura 2

Gráfico de barras que representa el acceso a Internet por comunidad autónoma en la librería Seaborn de Python



Fuente: elaboración propia

En la Figura 3 se muestra un gráfico de barras creado con Seaborn, el cual destaca por la mejora en la estética visual y la legibilidad de los datos. A diferencia de las figuras anteriores, este gráfico utiliza una paleta de colores gradientes que facilita la distinción entre las diferentes comunidades autónomas. Además, se han añadido etiquetas numéricas en cada barra, lo que permite una lectura rápida de los valores sin necesidad de referirse al eje horizontal. Estas mejoras hacen que la información sea más accesible y fácil de interpretar, contribuyendo a una experiencia visual más agradable y clara.

En la Figura 4 se presentan diversas visualizaciones de datos que reflejan diferentes aspectos del uso de Internet y tecnologías de la información en España en menores de 10 a 15 años. El gráfico de sectores (arriba a la izquierda) muestra la distribución del acceso a Internet por niveles educativos, proporcionando una visión clara de cómo varía el acceso según la formación. A su lado, un diagrama de cajas y bigotes resalta las variaciones en el acceso a Internet por diferentes comunidades autónomas, mostrando la mediana y los rangos intercuartílicos de los datos. El gráfico de puntos (arriba a la derecha) permite visualizar el uso de Internet por provincias, ofreciendo una perspectiva clara de la relación entre las provincias y el acceso a la red. En la parte inferior, un histograma compara la distribución de edades en relación con el acceso a Internet, y un gráfico de barras agrupadas (abajo a la derecha) compara el uso de Internet por provincias, facilitando una rápida comparación entre ellas. Estos gráficos, demuestran la versatilidad de las herramientas de visualización de datos en Jupyter Notebook y su capacidad para presentar diferentes dimensiones de un conjunto de datos de manera clara y comprensible.

Figura 3

Gráfico de barras que representa el acceso a Internet promedio por comunidad autónoma en la librería Seaborn de Python

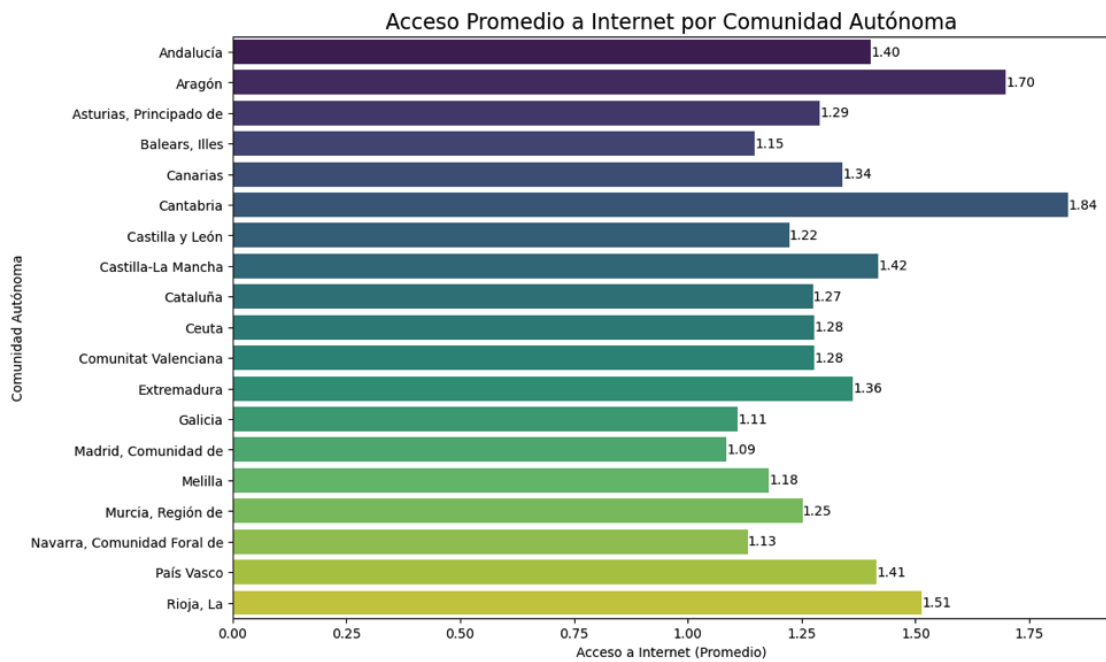
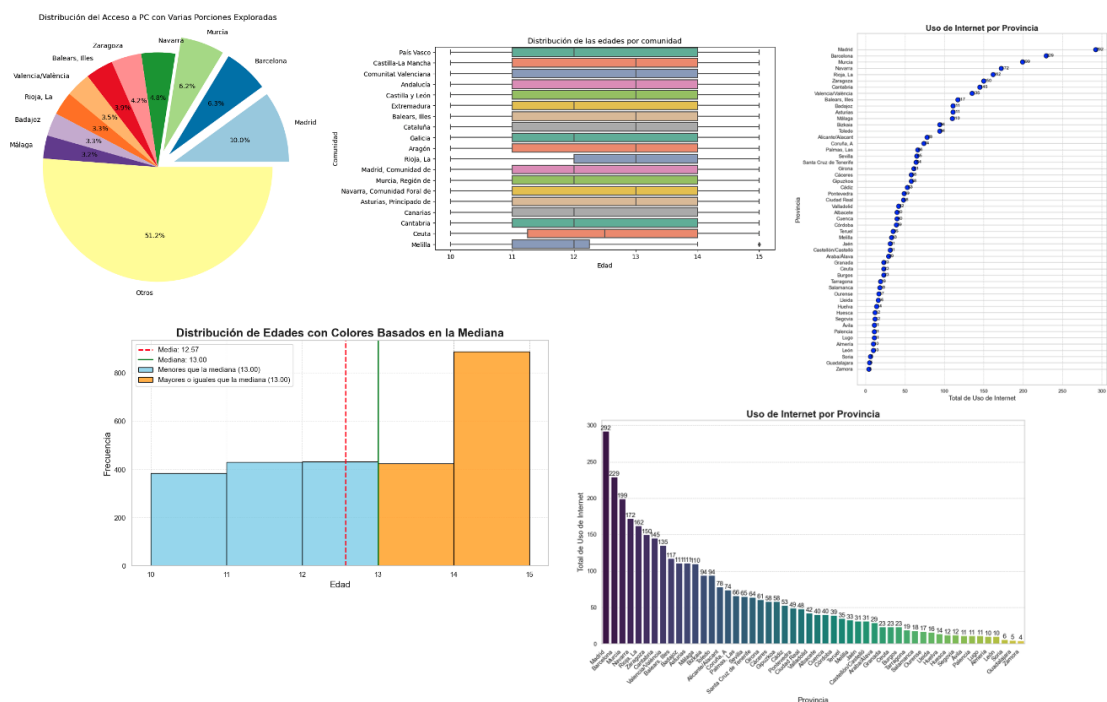


Figura 4

Muestra de gráficos realizados en Python



5. CONCLUSIONES

Jupyter Notebook se ha consolidado como una herramienta poderosa y versátil para la visualización y análisis de datos, particularmente relevante en los ámbitos de ciencia de datos y educación. Su entorno de desarrollo interactivo facilita la integración de código, texto explicativo, gráficos y resultados en un solo documento, lo cual mejora significativamente la experiencia de aprendizaje y la comunicación de hallazgos. Este trabajo muestra cómo Jupyter puede utilizarse para desarrollar visualizaciones que permiten una comprensión profunda de los datos, proporcionando al profesorado de matemáticas un recurso de gran valor para abordar la alfabetización de datos en distintos niveles educativos.

La compatibilidad de Jupyter con bibliotecas de visualización populares de Python, como Matplotlib y Seaborn, permite a los usuarios crear gráficos de alta calidad y personalizarlos para satisfacer distintas necesidades analíticas y pedagógicas. Matplotlib ofrece una flexibilidad completa en la configuración de gráficos, mientras que Seaborn facilita el diseño de visualizaciones más estéticas con menor cantidad de código, integrándose naturalmente con estructuras de datos de Pandas. Esta dualidad de opciones brinda a los educadores la libertad de elegir la biblioteca que mejor se adapte a su nivel de experiencia y a los objetivos específicos de su enseñanza.

El enfoque visual que ofrece Jupyter no solo facilita el análisis exploratorio de datos, sino que también permite presentar la información de manera clara y accesible. A lo largo del trabajo (especialmente en el enlace anexo), hemos mostrado cómo diferentes tipos de gráficos, como barras, líneas, sectores, histogramas, gráficos de dispersión y diagramas de cajas y bigotes, se pueden emplear para visualizar y comprender distintos aspectos de los datos sobre el uso de Tecnologías de la Información y la Comunicación (TIC) en menores. Estas visualizaciones, además de ser intuitivas, permiten al usuario identificar patrones, tendencias y posibles anomalías en los datos de forma rápida y efectiva.

Para los profesores de matemáticas, Jupyter Notebook representa una herramienta didáctica valiosa, ya que ofrece un entorno en el que los estudiantes pueden interactuar con datos reales, crear sus propias visualizaciones y explorar diferentes perspectivas de análisis de forma autónoma. Al poder manipular celdas de código en tiempo real, los estudiantes pueden experimentar y ver los efectos de cada cambio de manera inmediata, lo cual fomenta el pensamiento crítico y el aprendizaje activo. Asimismo, el diseño en celdas de Jupyter facilita la organización del contenido en secciones claras, combinando explicaciones teóricas con ejercicios prácticos y fomentando así una comprensión integral del análisis de datos.

Este trabajo también pone de manifiesto la importancia de la alfabetización de datos en la educación obligatoria. En un mundo cada vez más impulsado por los datos, la capacidad de explorar y representar visualmente la información es una competencia esencial que debe desarrollarse desde edades tempranas. Herramientas como Jupyter no solo acercan a los estudiantes al lenguaje de programación Python, sino que también les proporcionan un entorno amigable para la práctica de habilidades fundamentales en el análisis de datos, alineado con los diferentes niveles de demanda cognitiva y habilidades digitales. Como limitación, Python puede ser confuso si se utiliza sin tener conocimientos previos o si no se utiliza de manera guiada.

En conclusión, Jupyter Notebook es más que una simple herramienta de programación; es una plataforma integral para la educación en ciencia de datos, que permite a los docentes integrar visualización y análisis de datos en sus lecciones de manera accesible y efectiva. Esta capacidad de combinar teoría y práctica en un solo entorno contribuye a una enseñanza más

dinámica y adaptada a las necesidades de los estudiantes actuales, preparándolos para un futuro en el que la capacidad de analizar y comunicar datos será cada vez más relevante.

REFERENCIAS BIBLIOGRÁFICAS

- Barba, L. A., Barker, L. J., Blank, D. S., Brown, J., Downey, A. B., George, T., Heagy, L. J., Mandli, K. T., Moore, J. K., Lippert, D., Niemeyer, K. E., Watkins, R. R., West, R. H., Wickes, E., Willing, C., y Zingale, M. (2019). *Teaching and learning with Jupyter*. Recuperado: <https://jupyter4edu.github.io/jupyter-edu-book>, 1-77.
- Gajdoš, A., Hanč, J., y Hančová, M. (2022). Interactive Jupyter Notebooks with SageMath in Number Theory, Algebra, Calculus, and Numerical Methods. En *2022 20th International Conference on Emerging eLearning Technologies and Applications (ICETA)* (pp. 178-183). IEEE.
- Granger, B., y Pérez, F. (2021). Jupyter: Thinking and Storytelling with Code and Data. *Computing in Science & Engineering*, 23(2), 7-14.
- Instituto Nacional de Estadística [INE] (2024). Encuesta sobre equipamiento y uso de tecnologías de información y comunicación en los hogares, año 2023. <https://tinyurl.com/2bq2py8r>
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., y Willing, C. (2016). Jupyter Notebooks—a publishing format for reproducible computational workflows. En *Positioning and power in academic publishing: Players, agents and agendas* (pp. 87-90). IOS press.
- Molina-Portillo, E., Contreras, J.M., Molina-Muñoz, D. y Cabrera, R. (2024). Explorando la visualización de datos. *Epsilon*, 118, 9-20.
- Perkel, J. M. (2018). Why Jupyter is data scientists' computational notebook of choice. *Nature*, 563(7732), 145-146.
- Ragan-Kelley, M., Perez, F., Granger, B., Kluyver, T., Ivanov, P., Frederic, J., y Bussonnier, M. (2014). The Jupyter/IPython architecture: a unified view of computational research, from interactive exploration to communication and publication. En *Proceedings of the 13th Python in Science Conference (SciPy)*, 51-56.
- Rule, A., Tabard, A., y Hollan, J. D. (2018). Exploration and explanation in computational notebooks. En *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).
- Thompson, J., Wu, S. P., y Mills, J. (2020). The use of computer programming in a secondary mathematics class. En *2020 ASEE Virtual Annual Conference Content Access*.

Visualización de datos con R

Carlos Santiago Sánchez Muñoz

Universidad de Granada, carlossamu7@correo.ugr.es

Manuel Alejandro Jiménez Morales

Universidad de Granada, majm86@correo.ugr.es

Resumen: Este artículo explora el potencial del lenguaje R en la enseñanza de visualización y análisis de datos a nivel de Educación Secundaria y Bachillerato. Se introduce la herramienta y se proveen diversas fuentes susceptibles de guiar al profesorado en la creación de experiencias de clase ligadas a la enseñanza de habilidades de programación, análisis y visualización de datos, y pensamiento computacional. Por último, se proponen un conjunto de actividades que introducen, promoviendo un aprendizaje activo y experimental, el empleo de una base de datos real, ilustrando la gran versatilidad de R para la manipulación, análisis y representación gráfica de la información relevante.

Palabras clave: visualización de datos, ciencia de datos, estadística, pensamiento computacional, programación en R.

Data Visualization with R

Abstract: This article explores the potential of the R programming language in teaching data visualization and analysis at the secondary education level. The tool is introduced and various resources are provided to guide teachers in creating class experiences focused on programming, data analysis, visualization skills, and computational thinking. Lastly, it proposes a set of activities that promotes active, hands-on learning, by using a real database, illustrating R's versatility for data manipulation, analysis, and graphical representation of relevant information.

Key words: data visualization, data science, statistics, computational thinking, R programming.

1. INTRODUCCIÓN

El uso de R en el entorno educativo para el tratamiento y la visualización de datos se ha convertido en una herramienta poderosa que fomenta el aprendizaje de la estadística, la ciencia de datos y la programación de manera práctica y accesible. R es un lenguaje de programación de código abierto, ampliamente utilizado tanto en el ámbito académico como profesional, que permite a los estudiantes y educadores analizar grandes volúmenes de datos y representarlos gráficamente de manera clara y comprensible.

Una de las principales ventajas de R es su versatilidad y la amplia gama de paquetes y librerías disponibles para el manejo y visualización de datos, como *dplyr*, *ggplot2* y *shiny*. Estos paquetes permiten a los usuarios crear gráficos interactivos y personalizados que ayudan a ilustrar patrones y tendencias en los datos, facilitando su interpretación y análisis.

En el contexto educativo, R proporciona una oportunidad única para que los estudiantes desarrollen habilidades de pensamiento crítico y resolución de problemas a través del manejo de

datos reales. Al igual que el resto de herramientas, R se puede utilizar para representar visualmente la información, pero con el valor añadido de ofrecer una mayor flexibilidad y control sobre el proceso de análisis y visualización.

Además, R puede adaptarse a diferentes niveles educativos, desde la enseñanza secundaria hasta la educación superior. Su implementación en el aula puede ser progresiva, comenzando con la introducción a la programación básica y la creación de gráficos sencillos, hasta llegar a la manipulación avanzada de datos y la construcción de aplicaciones interactivas. Este lenguaje cuenta con una comunidad activa que proporciona tutoriales, foros y recursos educativos accesibles para estudiantes y docentes. Sin embargo, a diferencia de otras herramientas basadas en la web, el uso de R requiere un mayor grado de familiaridad con la programación, lo cual puede ser un reto, pero también una ventaja educativa, ya que fomenta el desarrollo de competencias digitales avanzadas.

Esta combinación de análisis de datos y visualización permite a los estudiantes no solo entender conceptos matemáticos y estadísticos, sino también explorar y comunicar sus hallazgos de manera efectiva a través de gráficos dinámicos y reportes reproducibles, alineándose con las demandas actuales de trabajo.

Para el desarrollo de esta práctica resulta imprescindible contar con una instalación del ecosistema R & RStudio, la cual se puede obtener de forma gratuita en su web oficial <https://posit.co/download/rstudio-desktop/>.

2. ANTECEDENTES

La gran capacidad y versatilidad de R, en combinación con el hecho de tratarse de un lenguaje de código abierto, ha propiciado su explotación desde una miríada de aplicaciones, tanto fuera como dentro del aula, si bien estos mismos rasgos lo ha hecho tradicionalmente más adecuado para entornos de educación superior. Buena muestra de ello lo dan Cuevas et al. (2019) o Chan et al. (2020), donde se exploran sus ventajas en comparación con otras similares y se contrasta el grado de implantación de la herramienta a nivel universitario. No obstante, trabajos como el de Ruiz et al. (2009) o Martínez et al. (2022) ofrecen un primer acercamiento a las características y funcionalidades de R que servirían para introducir al profesorado eventualmente más novel en habilidades de programación o visualización de datos.

De entre una gran variedad de recursos en línea, el libro de Estrellado et al. (2020) constituye un compendio excelente que abarca no sólo al lenguaje, sino también la forma de abordar el estudio de Ciencia de Datos en educación. Un paso más allá lo proporciona el recurso web de Bean (2021) que, si bien extiende la complejidad matemática más allá del objetivo que aquí se persigue, expone gran variedad de ejemplos y casos de estudio que podrían inspirar la creación de actividades y otras formas más heterodoxas de tratar los contenidos curriculares en clase.

Las aplicaciones de R en Educación Secundaria han sido menos reportadas, si bien hay ejemplos que dan buena muestra de las posibilidades de este recurso a pie de aula. Al tratarse de un lenguaje no ideado especialmente para la enseñanza, su empleo requiere de cierta elaboración por parte del docente para hacerlo más amigable. Aun así, estudios como los de Calahorra et al. (2019) y de Briz et al. (2018), que exponen la experiencia de uso con adolescentes a nivel de E.S.O. y Bachillerato, confirman el potencial de esta herramienta en la enseñanza de las matemáticas a nivel preuniversitario, así como hacia la integración del pensamiento computacional en el currículum de matemáticas.

3. METODOLOGÍA

A modo de ejemplo ilustrativo del uso de R, seguidamente se exponen una serie de actividades de visualización y análisis de datos susceptibles de ser empleadas en clase. El nivel de dificultad es creciente, y con ello se persigue un acercamiento tanto a las particularidades del lenguaje R como a la exploración de datos en general.

La base de datos que se va a emplear es la misma que en el resto de capítulos. Se trata de un documento en formato CSV (*Comma Separated Values*), que se leerá y formateará correspondientemente utilizando este software libre. Este capítulo propone tres grandes retos:

1. El primero se corresponde a la lectura y preprocesamiento del conjunto de datos. Para ello, es necesario alterar la gran cantidad de variables categóricas que incluye este dataset. Se realizarán los siguientes cambios:
 - 1.1. Para las variables dicotómicas 1= “Sí”, 6= “No” y 9= “NS/NR”.
 - 1.2. Para la variable *INTEFOR*, 1= “Fija”, 2= “Solo móvil” y 9= “NS/NR”.
 - 1.3. Para la variable *SEXO*, 1= “Hombre” y 6= “Mujer”.
2. Con el conjunto de datos adecuadamente procesado y cargado en memoria, nos encontramos en condiciones de construir cualquier gráfico deseado. Como se ha indicado anteriormente, se utilizará la librería *ggplot2*. Las tareas correspondientes a este apartado consistirán en:
 - 2.1. Construir un gráfico de barras que refleje el número de participantes de cada comunidad autónoma. Guarda esta imagen en una carpeta “img” ubicada en la misma ruta de los datos, con el nombre “participantes_CA.png” y con un ancho de 8cm y un alto de 6cm.
 - 2.2. Dibujar otro gráfico de barras con las edades de los participantes. Guarda esta imagen en una carpeta “img” ubicada en la misma ruta de los datos, con el nombre “participantes_EDAD.png” y con un ancho de 8cm y un alto de 6cm.
 - 2.3. Pintar un gráfico de sectores con la participación de cada sexo en la encuesta. Guarda esta imagen en una carpeta “img” ubicada en la misma ruta de los datos, con el nombre “sectores_SEXO.png” y con un ancho de 8cm y alto de 6cm.
 - 2.4. Construir un gráfico de barras que indique para cada sexo si se ha usado internet en los últimos tres meses. De forma análoga se realizará para la edad. Pista: utilizar el argumento en la función *fill*. Guarda estas imágenes en una carpeta “img” ubicada en la misma ruta de los datos, con el nombre “INT_según_SEXO.png” y “INT_según_EDAD.png”, y con un ancho de 8cm y un alto de 6cm.
 - 2.5. Prueba a personalizar los gráficos anteriores tanto como desees. Por ejemplo, se puede cambiar:
 - el color de las barras/sectores.
 - el título del gráfico y de los ejes.
 - el alineamiento del título (puedes centrarlo).
 - el tamaño de letra del título y de los ejes.
 - el color de fondo de la imagen. Pista: en *ggsave* utiliza el parámetro “bg”.
3. Finalmente, se va a aumentar levemente la complejidad solicitando cálculos estadísticos que implican algunos filtros o selecciones. Más concretamente se plantean dos actividades:
 - 3.1. Calcular una tabla de contingencia entre las variables *TELEF1* y *TELEF2*. Interpreta los resultados.

- 3.2. Calcula el porcentaje de personas con 15 años que no utilizan internet. Para ello, identifica las variables relevantes, realiza el filtro correspondiente y, por último, calcula el porcentaje.

4. RESULTADOS

Esta sección expone detalladamente los pasos a seguir para completar los retos planteados. Se indicarán los pormenores del proceso a la vez que los resultados que se van obteniendo. Comenzamos abriendo un nuevo script dentro de RStudio: *File > New File > R Script* y lo guardamos con el nombre que deseemos. La presentación de esta herramienta consiste en 4 paneles correspondientes al código fuente, la consola, el entorno y los otros recursos (como archivos, paquetes o menú de ayuda entre otros).

Las subsecciones siguientes abordan cada una de las tres grandes tareas propuestas en la Metodología.

4.1. Lectura y preprocesamiento de datos

Todos los lenguajes de programación tienen una guía de buenas prácticas con sugerencias y consejos que evitan errores y permiten realizar mantenimientos más eficientes. El primer consejo en el uso de R es siempre incluir una línea que borre cualquier información cargada en el sistema previamente mediante la ejecución de `rm(list=ls())`. Más adelante, se suelen incluir las librerías necesarias para poder ejecutar el software (si no dispone de alguna puede instalarlas ejecutando `install.packages("nombre_libreria")`). En estos ejemplos se van a emplear las librerías *dplyr* para el procesamiento de datos y *ggplot2* para la creación de gráficas. A continuación, se insertan la(s) ruta(s) de trabajo (cuidado con los ordenadores con sistema operativo Windows ya que copian las barras de forma invertida). Por último, se carga el conjunto de datos mediante `read.csv`, función contenida en el software base de R. Obsérvese que se indica que el elemento separador es el carácter “;” y la codificación “latin1”.

```
rm(list=ls())

library(dplyr)
library(ggplot2)

# Rellenar ruta adecuadamente
ruta <- ""

# Lectura del conjunto de datos
data <- read.csv(paste0(ruta, "datospaper.csv"), sep=";", encoding = "latin1")
```

Tras esta lectura del fichero, hemos de formatear aquellas variables que vamos utilizar para el análisis:

```
T_SINO_lev <- c(1, 6, 9)
T_SINO_lab <- c("Sí", "No", "NS/NR")

# Cambiamos la variable categóricas por su factor correspondiente
data <- data %>%
  mutate(SEXO = factor(SEXO, levels = c(1, 6), labels = c("Hombre", "Mujer")),
         ORD = factor(ORD, levels = T_SINO_lev, labels = T_SINO_lab),
         TABLET = factor(TABLET, levels = T_SINO_lev, labels = T_SINO_lab),
         TELEF1 = factor(TELEF1, levels = T_SINO_lev, labels = T_SINO_lab),
         TELEF2 = factor(TELEF2, levels = T_SINO_lev, labels = T_SINO_lab),
         VIV_INTER = factor(VIV_INTER, levels = T_SINO_lev,
```

```

        labels = T_SINO_lab),
    INTEFOR = factor(INTEFOR, levels = c(1, 2, 9),
        labels = c("Fija", "Solo móvil", "NS/NR")),
    PC = factor(PC, levels = T_SINO_lev, labels = T_SINO_lab),
    INT = factor(INT, levels = T_SINO_lev, labels = T_SINO_lab),
    MOVIL = factor(MOVIL, levels = T_SINO_lev, labels = T_SINO_lab)
)

```

4.2. Visualización de gráficas

El lenguaje de programación R tiene una gran potencia para la visualización de datos gracias a la librería ggplot. Ésta, permite crear gráficos altamente personalizables de manera intuitiva. Para empezar, se debe usar la función ggplot() donde se especifica el dataframe (por ejemplo, data) y, dentro de aes(), se definen las variables estéticas, como los ejes.

Para la consecución de la gráfica pedida en 2.1, indicaremos que x es la Comunidad y utilizaremos la función geom_bar. La función theme_minimal() le da un aspecto más minimalista eliminando los ejes. La función labs permite establecer el valor del título y de los ejes. Con theme es posible personalizar aspectos estéticos del gráfico. Por último, se almacena con ggsave, indicando la ruta (se añade “/img” y previamente se crea la carpeta manualmente) y el ancho y alto (con “width” y “height” respectivamente).

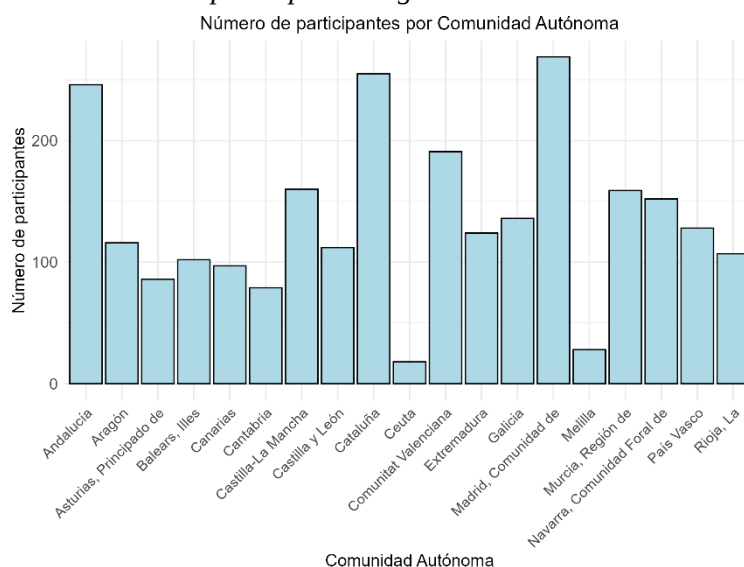
```

# Graficar el número de participantes por Comunidad Autónoma
ggplot(data, aes(x = Comunidad)) +
  geom_bar(fill = "lightblue", color = "black") +
  theme_minimal() +
  labs(title = "Número de participantes por Comunidad Autónoma",
       x = "Comunidad Autónoma", y = "Número de participantes") +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 11),
    axis.text.y = element_text(size = 12),
    axis.title.x = element_text(size = 13),
    axis.title.y = element_text(size = 13),
    plot.title = element_text(hjust = 0.5, size = 14)
  )
ggsave(paste0(ruta, "participantes_CA.png"), width = 8, height = 6)

```

Figura 1

Gráfico de barras del número de participantes según la Comunidad Autónoma.



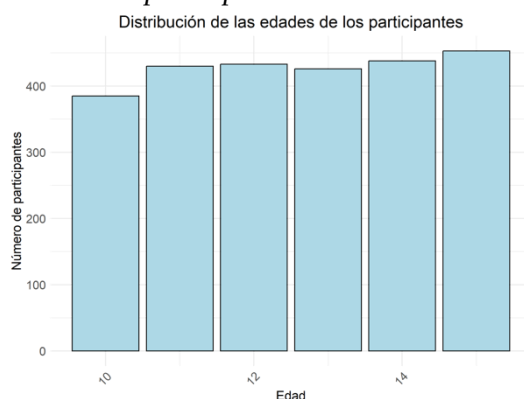
En la Figura 1 se puede observar que las comunidades con menor participación son Ceuta y Melilla y la de mayor participación es la Comunidad de Madrid.

El procedimiento es idéntico para la edad, con la salvedad de cambiar el atributo correspondiente.

```
ggplot(data, aes(x = EDAD)) +
  geom_bar(fill = "lightblue", color = "black") +
  theme_minimal() +
  labs(title = "Distribución de las edades de los participantes",
        x = "Edad", y = "Número de participantes")
ggsave(paste0(ruta, "participantes_EDAD.png"), width = 8, height = 6)
```

Figura 2

Gráfico de barras con la edad de los participantes de la encuesta.

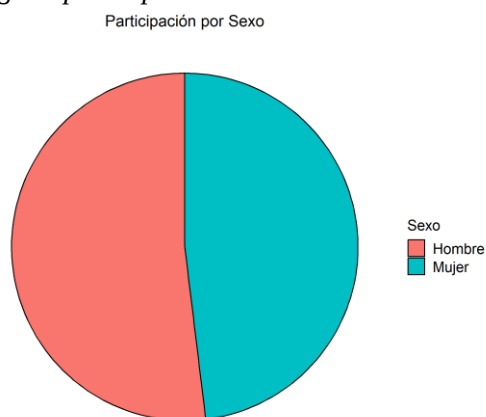


Se observa que la distribución por edades es uniforme. Para el gráfico de sectores es necesario pasarlo a coordenadas polares, utilizando la función `coord_polar(theta = "y")`. Además, eliminamos todos los ejes con `theme_void()`.

```
# Graficar el gráfico de sectores
ggplot(data, aes(x = "", fill = factor(SEXO))) +
  geom_bar(width = 1, color = "black") +
  coord_polar(theta = "y") +
  labs(title = "Participación por Sexo", fill = "Sexo") +
  theme_void() # Elimina los ejes
ggsave(paste0(ruta, "sectores_SEXO.png"), width = 8, height = 6)
```

Figura 3

Gráfico de sectores que recoge la participación de cada sexo en la encuesta.



El gráfico de sectores de la Figura 3 muestra que existe una ligera mayor participación de hombres que de mujeres en la encuesta.

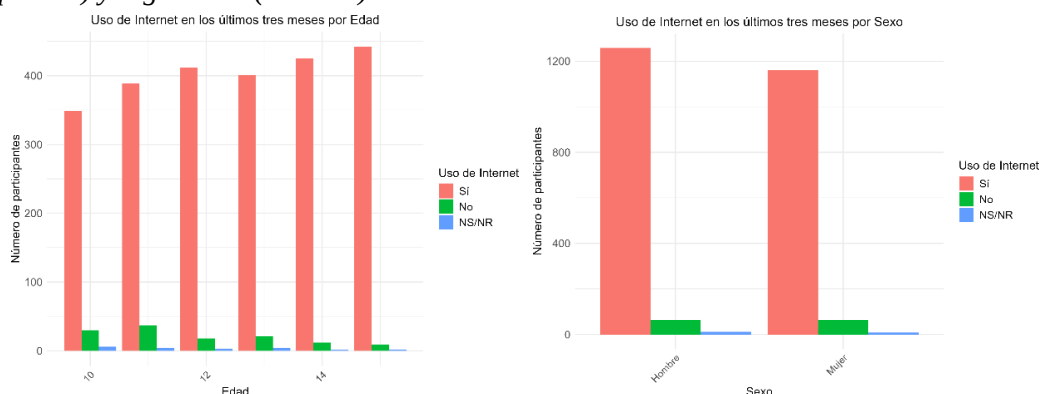
Para graficar una variable en función de otra se emplean parámetros de estética, como “fill”, “color” o “size” entre otros. Además, se ha empleado “dodge” para la posición de las barras de modo que una se ubique al lado de la otra.

```
# Graficar el uso de Internet por sexo
ggplot(data, aes(x = factor(SEXO), fill = factor(INT))) +
  geom_bar(position = "dodge") +
  labs(title = "Uso de Internet en los últimos tres meses por Sexo",
       x = "Sexo", y = "Número de participantes", fill = "Uso de Internet") +
  theme_minimal()
ggsave(paste0(ruta, "INT_según_SEXO.png"), width = 8, height = 6)

# Graficar el uso de Internet por edad
ggplot(data, aes(x = EDAD, fill = factor(INT))) +
  geom_bar(position = "dodge") +
  labs(title = "Uso de Internet en los últimos tres meses por Edad",
       x = "Edad", y = "Número de participantes", fill = "Uso de Internet") +
  theme_minimal()
ggsave(paste0(ruta, "INT_según_EDAD.png"), width = 8, height = 6)
```

Figura 4

Gráfico de barras que expresa el uso de internet en las familias españolas según edad (izquierda) y según sexo (derecha).



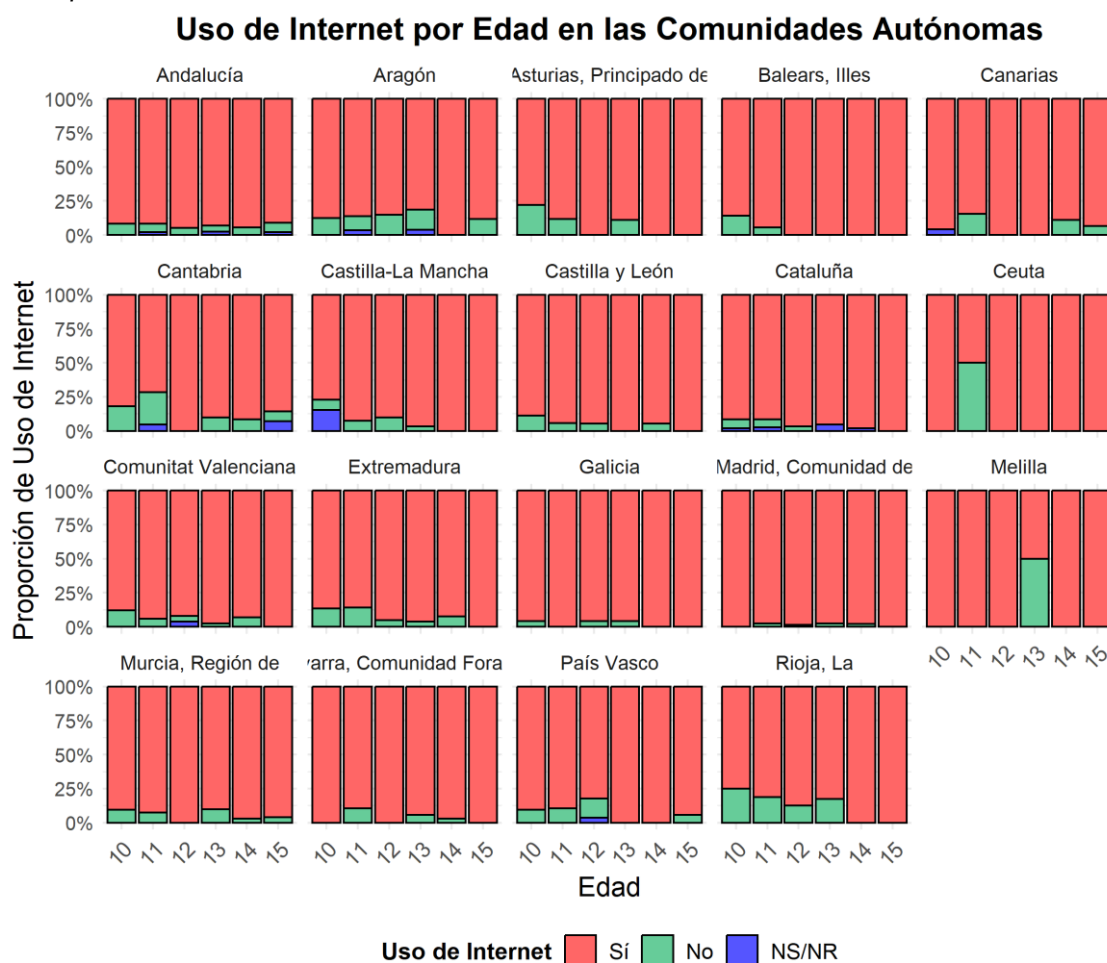
Por último, se puede personalizar al gusto la gráfica. Encuentre a continuación el código de un ejemplo:

```
ggplot(data, aes(x = EDAD)) +
  geom_bar(fill = "orange", color = "darkred") +
  theme_minimal() +
  labs(title = "Distribución de las Edades de los Participantes",
       subtitle = "Encuesta de Uso de Internet",
       x = "Edad de los Participantes", y = "Número de participantes") +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 11),
    axis.text.y = element_text(size = 12),
    axis.title.x = element_text(size = 13),
    axis.title.y = element_text(size = 13),
    plot.title = element_text(size = 14)
  )
```

A modo adicional, se va a elaborar un gráfico avanzado que combina múltiples variables para ilustrar el potencial y la versatilidad de R en el ámbito de las ciencias de datos. Este gráfico demuestra cómo R facilita el análisis y la visualización de datos complejos de forma efectiva.

Figura 5

Gráfico de barras que refleja el porcentaje de uso de internet en las familias españolas según la edad, facetando la Comunidad Autónoma.



A simple golpe de vista la gráfica expone que la inmensa mayoría de jóvenes de España sí usan internet.

4.3. Cálculos estadísticos

Este lenguaje nació en el contexto de la Estadística por lo que otro de sus puntos fuertes es contar con funcionalidades y herramientas que permitan hallar casi cualquier cálculo deseado.

El ejercicio 3.1 requiere del empleo de “table”, que tabula una variable en función de la otra. El uso de “prop.table” realiza el mismo cálculo pero en porcentaje.

```
# Calcular la tabla de contingencia
tabla_contingencia <- table(data$TELEF1, data$TELEF2)
print(tabla_contingencia)

# Mostrar la tabla en porcentaje sobre el total
tabla_contingencia_prop <- prop.table(tabla_contingencia) * 100
print(tabla_contingencia_prop)
```

La tabla resultante de la ejecución anterior es la siguiente:

Tabla 1

Tabla de contingencia de las variables TELEF1 y TELEF2.

		TELEF2		
		Sí	No	NS/NR
TELEF1	Sí	1542	2	0
	No	1011	0	0
	NS/NR	10	0	0

La interpretación de esta tabla permite identificar varias ideas:

- Casi todos los hogares cuentan con móviles.
- El uso de teléfono fijo se está perdiendo (un 40% de hogares no tienen).
- Cuando el hogar no cuenta con teléfono fijo es porque hay al menos un móvil.

Para resolver la tarea 3.2, se construye un nuevo conjunto de datos filtrando (función “filter”) a que la edad sean 15 años. Se cuenta el número de participantes con esa edad que son 453. Se aplica un nuevo filtro que indica que no se usa internet, resultando en 9 participantes. Finalmente, se halla el porcentaje con el cálculo $9/453 \times 100$, obteniendo un 1,99% de participantes con 15 años que no han utilizado internet en los últimos tres meses.

```
# Filtrar las personas que tienen 15 años
personas_15 <- data %>%
  filter(EDAD == 15)

total_personas_15 <- nrow(personas_15)

# Calcular el número de personas de 15 años que no utilizan internet
personas_15_sin_internet <- personas_15 %>%
  filter(INT == "No")

total_personas_15_sin_int <- nrow(personas_15_sin_internet)

# Calcular el porcentaje de personas de 15 años que no utilizan internet
porcentaje_sin_internet <- (total_personas_15_sin_int / total_personas_15) * 100

# Mostrar el porcentaje
print(paste("El porcentaje de personas de 15 años que no utilizan internet es:",
  round(porcentaje_sin_internet, 2), "%"))
```

5. CONCLUSIONES

La gran difusión que el lenguaje R ha alcanzado en la exploración y visualización de datos da buena muestra de sus posibilidades en cuanto a su uso a pie de aula. Especialmente en la enseñanza de la Estadística, esta herramienta gratuita y de código abierto permite a estudiantes y profesores acceder a potentes funcionalidades sin coste alguno, fomentando además un enfoque colaborativo y de código reproducible.

Las actividades aquí presentadas persiguen la adquisición de nociones fundamentales sobre manipulación y visualización de datos en el contexto educativo, haciendo uso de R. A través de estas tareas, los estudiantes no solo se familiarizan con herramientas útiles para su futura inserción laboral, sino que también desarrollan habilidades de pensamiento crítico y computacional. El análisis de un conjunto de datos sobre el uso de la tecnología e internet en diferentes regiones de España enriquece el aprendizaje, ofreciendo una aplicación práctica y

relevante para el alumnado, hacia una experiencia educativa más rica, práctica y adaptable a la diversidad de disciplinas existente.

REFERENCIAS BIBLIOGRÁFICAS

- Bean, J. (2021). Using R for Educational Research.
- Briz, A. y Serrano, Á. (2018). Aprendizaje de las matemáticas a través del lenguaje de programación R en Educación Secundaria. *Educación Matemática*, 30(1), 133-162.
- Calahorra, J., Aguilar, T., Diciembre, S. y Sanchiz, D. (2019). Aproximación didáctica a las matemáticas a través de la programación en R. *Números*, 102, 161-184.
- Chan, D. y Galli M.G. (2020). Aplicación de técnicas estadísticas multivariadas con el lenguaje de programación R en investigaciones educativas del nivel superior. *Revista Argentina de Educación Superior*, 20, 123-136.
- Cuevas, H., Solís, C. y Silva, I. (2019). Programación computacional y análisis de datos en educación estadística. *Areté: Revista Digital del Doctorado en Educación de la Universidad Central de Venezuela*, 5(9), 11-27.
- Estrellado, R.A., Freer, E.A., Motsipak, J., Rosenberg, J.M. y Velásquez, I.C. (2020). *Data science in education using R*. Londres, ed. Routledge.
- Martínez J., Joel F., Herrera, J.B. y Paredes, F.A. El programa R: Una estrategia inicial para su entendimiento y aprendizaje. *Revista Digital Universitaria*, 23(4) (<http://doi.org/10.22201/cuaieed.16076079e.2022.23.4.4>)
- Puig, L. (1997). Análisis fenomenológico. En L. Rico (Coord.) *La educación matemática en la enseñanza secundaria*, 61-94. Horsori.
- Ramírez, R., Brizuela, B. y Blanton, M. (2020). Kindergarten and first-grade students' understandings and representations of arithmetic properties. *Early Childhood Education Journal*, 50, 1-12.
- Ruiz, M. y López, E. (2009). El entorno estadístico R: ventajas de su uso en la docencia y la investigación. *Revista Española de Pedagogía*, 243, 255-274.

Tercer Anuncio del 19 CEAM Huelva, del 12 al 14 de abril de 2025

Comités Local de Organización y Científico
ceam@saemthales.es

Resumen: Presentamos el tercer anuncio de nuestro 19 CEAM en la revista EPSILON, a la vez que también en web, que usaremos para informar de todo lo relacionado con él. En este tercer anuncio aparece: a) fechas y sede; b) tópicos; c) conferencias plenarias; d) programa provisional; e) programa de acompañantes; f) alojamiento; g) cena de gala; i) segunda edición premios 19 CEAM, TFG y TFM; j) cuotas e inscripción del congreso; k) comité organizador.

¡Ven a compartir ideas y construir una educación matemática inclusiva y enriquecedora!

- Insíbete y aporta tu visión mediante una experiencia de aula o investigación, en diversos formatos: TALLER, COMUNICACIÓN, PÓSTER, ZOCO ...
- Analiza y conoce la visión y las experiencias de otros docentes.
- Forma parte de este encuentro organizado por la Delegación de SAEM Thales, en Huelva!

NOTA IMPORTANTE: Para la presentación de trabajos es necesario estar registrado primero en la web y posteriormente seguir los pasos en las instrucciones para su presentación como aparece en el apartado de **Bienvenida a la WEB del XIX CEAM:** <https://ceam.saemthales.es/node/20>.

1. FECHA Y SEDE

El 19 CONGRESO DE ENSEÑANZA Y APRENDIZAJE DE LAS MATEMÁTICAS (19 CEAM) se celebrará del sábado, 12 de abril, al lunes, 14 de abril de 2024 en Huelva. La sede será la Facultad de Derecho de la Universidad de Huelva, Campus El Carmen - Avda. de las Fuerzas Armadas, s/n. - 21007 Huelva.

2. TÓPICOS

El congreso se estructurará en los siguientes tópicos:

2.1. Tópico 1. ¿Qué matemática es necesaria en una sociedad matematizada?

Una sociedad matematizada es aquella en la que la matemática desempeña un papel fundamental en el funcionamiento de multitud de aspectos de la misma.

Como lenguaje que es, la matemática representa un medio de comunicación potente, ¿qué uso le darías a la matemática para conseguir que el alumnado se desenvuelva en la vida cotidiana de una manera digna?

2.2. Tópico 2. Las matemáticas como proceso dinámico.

En la formación de profesores de matemáticas es imprescindible la aplicación del pensamiento matemático para la solución de situaciones problemáticas. Es necesario aplicar pensamientos

flexibles, analíticos y sobre todo dinámicos que provocarán unas correctas decisiones en los docentes. Tus experiencias enriquecerán los debates.

2.3. Tópico 3. Multiculturalismo, interculturalidad y diversidad en educación.

¿Es necesario trabajar la diversidad cultural en el aula? Es fundamental porque culmina con su integración real. En el XIX congreso de Matemática tendrás la oportunidad de afrontar la diversidad cultural en el aula como reto para formar personas respetuosas, tolerantes y libres.

2.4. Tópico 4. Visión intercultural de las Matemáticas escolares y Universitarias. Estudios de casos.

Implementar la interculturalidad en el área de las matemáticas a nivel conceptual y a nivel práctico puede ser un importante factor a tener en cuenta en nuestras aulas. Buscamos la motivación en el aprendizaje de nuestros estudiantes, pero ¿tenemos en cuenta los elementos de su entorno sociocultural? ¿Has diseñado y desarrollado SDA cercanos a la realidad de todo estudiante teniendo en cuenta su cultura? ¿Quieres conocer prácticas pedagógicas y conocimientos de matemáticas de otras partes del mundo? ¿Conoces el término etnomatemáticas? Qué nos podrá aportar?

2.5. Tópico 5. La historia de las Matemáticas escolares y universitarias. Estudio de casos.

El alumnado percibe la matemática como algo frío, aburrido y difícil. Es necesario humanizarla haciéndole ver que esta disciplina es el resultado de la actividad humana, que se ha ido construyendo a lo largo del tiempo por diversas civilizaciones. Desempolva tus antiguos libros de texto y únete a nosotros en un viaje a través del tiempo matemático.

2.6. Tópico 6. Gestión del talento matemático en un entorno multicultural.

En abril de 2025, en nuestra Huelva multicultural y descubridora, compartiremos ideas para reconocer y desarrollar el talento matemático en un contexto de diversidad cultural. Te invitamos a explorar enfoques que hagan nuestras aulas más inclusivas y adaptadas a una realidad multicultural.

3. CONFERENCIAS PLENARIAS

Las conferencias plenarias serán impartidas por:

3.1. Uwe Gellert

Profesor de Educación Matemática en la Freie Universität Berlin, Alemania. Sus investigaciones se centran en las relaciones entre el currículo y las (des-)igualdades sociales. Fue Presidente de la Commission Internationale pour l'Étude et Amélioration de l'Enseignement des Mathématiques (CIEAEM); Director de la School of Education en Berlín; y Editor de la revista Journal of Mathematics Teacher Education. Ha publicado numerosos

artículos, en Journal of Curriculum Studies, Educational Studies in Mathematics y Enseñanza de las Ciencias.

Conferencia: *Dinámicas de la matematización de la sociedad: estructura y ejemplos.*

Puede sonar paradójico: La matematización de la sociedad comienza mucho antes de que exista un cuerpo de conocimientos que llamamos matemáticas. Hace 5.000 años, el comercio a larga distancia comenzó a registrar la cantidad de mercancías, primero con símbolos y más tarde con números. En este contexto se desarrolló la primera protoaritmética, que se teorizó hasta convertirse en una matemática de los números en los milenios siguientes. Por otra parte, surgió una forma de comercio de mercancías en la que los números a menudo se hacían invisibles (por ejemplo, ocultos en códigos de barras). Este es un ejemplo de cómo el proceso de matematización de un campo de acción social (compra, venta, entrega, facturación) va acompañado de un proceso de desmatematización. El elemento de desmatematización va acompañado de una devaluación del valor de utilidad inmediata de las matemáticas escolares. En el estado matematizado, las matemáticas que se esconden detrás de los dispositivos técnicos, las normas de distribución, los reglamentos, las aplicaciones informáticas, etc., suelen permanecer ocultas o sólo aparecen en la superficie. Sin embargo, pueden simplificar la vida cotidiana. Los modelos matemáticos que estructuran la realidad son cada vez más complejos. En consecuencia, las matemáticas estructuran la interacción social de una manera que hace cada vez más difícil comprender cómo funcionan los dispositivos y cómo se toman las decisiones. No cabe duda de que el problema al que aquí se alude ha cobrado un impulso considerable gracias al desarrollo de la inteligencia artificial.

Es una cuestión pertinente qué conclusiones se puede extraer de la matematización/desmatematización en curso para la definición de los objetivos educativos en matemáticas.

3.2. Lola Cabrillana

Maestra de Educación Infantil y Pedagogía Terapéutica desde hace más de veinte años en un barrio de difícil desempeño. Mediadora intercultural especialista en la cultura e historia del pueblo gitano. Colaboradora de la cadena SER Málaga con un programa educativo y de literatura. Escritora de tres novelas, dos de ellas con la editorial Grijalbo. Soy activista por la defensa de la visualización del pueblo gitano.

Conferencia: *Matemáticas y diversidad: una oportunidad para sumar.*

Todas las culturas utilizan las matemáticas, pero no cabe duda de que el desarrollo, uso y transformación de estas tienen una identidad propia. En el aula, esas matemáticas se universalizan, creando una sola realidad que intenta dotar al alumnado de habilidades y estrategias y rara vez tiene en cuenta su contexto, su etnia o procedencia.

Cuando sumamos las oportunidades que nos ofrecen las distintas culturas que integran el aula, utilizando elementos que creen un contexto rico en estímulos, encontramos un escenario perfecto para comenzar a construir: exploramos en la diversidad.

Disfrutando de esa diversidad, en el aula de educación infantil, comenzamos a crear los cimientos de un pensamiento matemático; indicamos el camino para iniciarse en las habilidades lógico-matemáticas.

A través de esta experiencia, dentro del aula, observaremos como el niño y la niña adquiere el conocimiento cuando el juego está lleno de elementos conocidos, familiares, que lo acogen y les ofrecen un medio cercano. Mostraremos cómo es fácil aprender de otras culturas si los maestros y las maestras nos convertimos en fomentadores del talento de nuestro alumnado, cuidando que la generalización de lo aprendido también se pueda llevar a cabo dentro de su propia cultura.

3.1. Javier Díez Palomar

Profesor de Didáctica de las Matemáticas del Departamento de Educación Lingüística, Científica y Matemática de la Universidad de Barcelona. Vicepresidente de la Comisión Internacional para la Mejora de la Enseñanza y Aprendizaje de las Matemáticas (CIEAEM). Investigador principal del Grupo de Investigación consolidado (reconocido por la Generalitat de Catalunya): Grupo de Investigación en Educación Superadora de Desigualdades (GRESUD), miembro de CREA (Community of Research on Excellence for All). Miembro del Grupo de Expertos en Alfabetización Numérica del PIAAC (OECD). Miembro del Comité Ejecutivo (Trustees Committee) de la asociación internacional de investigación “Adults Learning Mathematics.” Investigador principal de varios proyectos de investigación en educación matemática (I+D y proyectos europeos).

Conferencia: *Matemática dialógica para la inclusión.*

Esta ponencia se centra en la matemática dialógica. A través de la presentación de los siete principios del aprendizaje dialógico, se discutirá como este enfoque permite una aproximación inclusiva de la diversidad, y en especial, de los grupos más vulnerables de nuestra sociedad, como es el caso de los niños y de las niñas gitanos y gitanas. Se presentarán actuaciones educativas que han mostrado tener éxito en la mejora tanto de los resultados de aprendizaje de matemáticas y la eliminación del absentismo escolar, como de aspectos de convivencia. Se pondrá énfasis especial en el trabajo con la comunidad, y la importancia de trabajar conjuntamente con las familias y otras personas relevantes del entorno para construir comunidades de aprendizaje basadas en evidencias científicas que tienen como resultado los mejores resultados de aprendizaje de matemáticas, y de más excelencia, para todos y todas.

4. PROGRAMA PROVISIONAL

El congreso está estructurado en comunicaciones, talleres, sesión de pósteres, zoco matemático, exposiciones según el programa adjunto que aparece en la Figura adjunta:

Figura 1*Programa provisional*

19 Congreso sobre Enseñanza y Aprendizaje de las Matemáticas			
SAEM THALES			
HORA	Sábado, 12 de abril	Domingo, 13 de abril	Lunes, 14 de abril
09:00-10:00	Documentación	Talleres	Junta de la SAEM Thales
10:00-10:30	Acto Inaugural	Talleres	Comunicaciones
10:30-11:00	Pausa-Café	Pausa-Café	Pausa-Café
11:00-12:30	Sesión Plenaria-Apertura	Talleres	Comunicaciones
12:30-13:45	Comunicaciones	Sesión Plenaria	Sesión Plenaria-Clausura
14:00-16:00	Almuerzo	Almuerzo	
16:00-18:30	Talleres	Comunicaciones	
18:30-20:00	Comunicaciones	Talleres	
20:00-20:30	Visita Expos-Póster	Visita Expos-Póster	
21:30-01:00	Cena-Cóctel		
NOTA: La asignación de espacios y ubicación definitivas de aulas para comunicaciones, talleres y póster se anunciará oportunamente.			

5. PROGRAMA DE ACOMPAÑANTES

La organización del CEAM ha puesto especial cuidado en el diseño del programa de acompañantes para que su visita a Huelva durante el desarrollo del Congreso sea una experiencia inolvidable. Puede consultarse en <https://forms.gle/UFhZjcatxGoUEbKv7>.

El Comité Local de Organización del 19 CEAM está cerrando con las instituciones un programa de actividades a los acompañantes para la visita de los Lugares Colombinos, como una ruta, declarada Conjunto Histórico-Artístico, que discurre en torno a Moguer y Palos de la Frontera, lugares andaluces que tuvieron una especial relevancia en la preparación y la realización del primer viaje de Cristóbal Colón y el Descubrimiento de América:

Día 13: Visita guiada a lugares colombinos (Muelle de las Carabelas, Monasterio de La Rábida, Iglesia de San Jorge en Palos de la Frontera).

Día 14: Visita guiada a la ciudad de Huelva.

6. ALOJAMIENTO

A continuación, puedes encontrar una lista de hoteles ubicados muy cerca del lugar del congreso y en el centro de la ciudad, que cuentan con precios especiales para los congresistas. Al hacer su reserva siga las indicaciones indicadas para beneficiarse de estas tarifas especiales. Le recomendamos reservar su hotel lo antes posible, ya que habrá otros eventos en Huelva.

Residencia Universitaria Loops Home El Carmen https://www.loop-homes.com/loop-homes-el-carmen-huelva/	Habitación Individual-44€ Habitación en Twodio-38,50€ por persona <i>No se sirve desayuno.</i>
Hotel NH- LUZ https://www.nh-hotels.com/es/hotel/nh-luz-huelva	Habitación simple con desayuno: 105 € Habitación doble con desayuno: 115 €

Hotel AC-Huelva https://www.marriott.com/es/reservation/rateListMenu.mi	<i>Habitación simple con desayuno: **</i> <i>Habitación doble con desayuno: **</i> AVISO: Mirar los precios actuales en la web del hotel, están entre 85€ y 94€. Se está negociando una rebaja que se comunicará a la mayor brevedad.
Hotel Senator https://www.senatorhr.com/senator-huelva-hotel/	<i>Habitación simple con desayuno: **</i> <i>Habitación doble con desayuno: **</i> AVISO: Mirar los precios actuales en la web del hotel, están entre 84€ y 95€. Se está negociando una rebaja que se comunicará a la mayor brevedad.
Hotel Exe-Tartessos https://www.eurostarshotels.com/exe-tartessos.html	<i>Habitación simple con desayuno: **</i> <i>Habitación doble con desayuno: **</i> AVISO: Mirar los precios actuales están entre 75€ y 96€. Se está negociando una rebaja que se comunicará a la mayor brevedad.

NOTA: Hay que hacer notar que estamos en vacaciones de semana santa y el Comité Local de Organización está haciendo las gestiones oportunas para ajustar los precios lo más posible. Para un listado más completo de hoteles de Huelva y provincia:

<https://es.search.yahoo.com/search?fr=mcafee&type=E210ES91082G91820&p=hoteles+en+huelva>

7. CENA DE GALA DEL CONGRESO

Nos complace anunciar que la cena de gala se llevará a cabo el Sábado 12 de abril a partir de las 21 h 30 min en el complejo **La Casa del Conquero**, lugar mágico, que, sin duda, nos dejará a todos con un recuerdo inolvidable, que podéis visitar en el enlace: <https://lacasadelconquero.com/spaces>.

La Casa del Conquero es un lugar incomparable que pone en valor un marco incomparable desde su atalaya, de las marismas de Huelva gran potencial de la ciudad de Huelva. Está ubicado junto al Mirador del Conquero, en un enclave paisajístico inigualable, siendo la edificación una de las casas más reconocibles de la ciudad.

El precio de la cena será de 25€ para los no socios y acompañantes y 20€ para los socios. La cena será ofrecida por la empresa La COMPANY que ofrecerá un menú típicamente onubense elegido del catálogo que adjuntamos, entre otros platos.

Figura 2

Platos típicos de Huelva





8. PREMIOS CEAM –TFG-TFM-SEGUNDA EDICIÓN

Relativo a esta segunda edición, en estos momentos, se está en conversaciones con Facultades y Departamentos, para dar a conocer los CEAM y la Sociedad Andaluza de Educación Matemática THALES y a fin de conectar a los actuales maestros/as y profesores/as de matemáticas, se con-vocan los PREMIOS CEAM TFG (TRABAJO FIN DE GRADO) Y TFM (TRABAJO FIN DE MÁSTER) y hacerlo extensivo a otras disciplinas con alta contenido en Matemáticas. Se realiza un concurso para recién graduados o estudiantes del Grado en Educación Primaria, Grado en Educación Infantil, Grado en Matemáticas, Grado en Estadística, Máster en Estadística Aplicada y Máster en Profesorado de Enseñanza Secundaria Obligatoria y Bachillerato, Formación Profesional y Enseñanzas de Idiomas que hayan defendido el TFG o el TFM durante los cursos 2023/2024 2024/2025 (en este curso hasta el mes de febrero de 2025).

Todos/as los/as interesados/as podrán enviar un resumen, como máximo de 1500 palabras en formato WORD siguiendo el **formato de entrega de trabajos** que aparece en la página web del congreso: <https://ceam.saemthales.es/instrucciones> .

9. CUOTAS DE INSCRIPCIÓN

Los plazos para inscribirse en el 19 CEAM aparecen en la Tabla 1.

Tabla 1

Plazos de inscripción

	Hasta el 15 de marzo de 2025	Desde el 16 de marzo de 2025 hasta el 5 de abril de 2025
Socios (de THALES o de la FESPM)	120 €	160 €
No socios	170 €	210 €
Estudiantes/Jubilados	65 €	70 €

10. COMITÉ LOCAL DE ORGANIZACIÓN

Figura 3

Comité Local de organización



11. MÁS INFORMACIÓN-CONTACTO

Para una información más completa puede consultarse la web: <https://ceam.saemthales.es/>.

También se puede seguirnos en **@Ceam_SaemThales** en **Facebook**, **Instagram** y **X**, donde os animamos a compartir los contenidos con vuestros contactos.

Más y puntual información irán apareciendo en los próximos anuncios que debidamente se publicarán en la web del 19 CEAM. Para consultas y dudas sobre el 19 CEAM, puede contactar con nosotros vía correo electrónico: ceam@saemthales.es.

¡Nos vemos en Huelva!

Comité Local de Organización y Comité Científico

