**Estudiante:** Patricio Pantaleo  
**Asignatura:** Visualización de información 25/26-1C  
**Fecha:** 9/10/2025  
**Dataset:** Estadísticas de Bibliotecas de Castilla-La Mancha 2015  
**Fuente:** [datos.gob.es](https://datos.gob.es/es/catalogo/a08002880-estadisticas-de-bibliotecas-de-castilla-la-mancha)

## Introducción

El presente informe tiene por objetivo dar cuenta de los procesos, medidas y herramientas utilizadas para la realización de la actividad número 2 para el módulo de Visualización de la información del Máster Universitario en Bibliotecas, Archivos y Continuidad Digital 25/26, de la UC3M.

Inicialmente, cabe destacar que la actividad realizada tiene asignado un repositorio en mi cuenta personal de GitHub en donde se puede acceder a todo el material y códigos utilizados.

Se puede acceder al [**repositorio general de la actividad aquí**](https://github.com/Ppantaleo/Informes-Bibliometricos/tree/main/UC3M), y también se puede [**acceder a un archivo .md con las descripciones principales**](https://github.com/Ppantaleo/Informes-Bibliometricos/blob/main/UC3M/VisAct2.md) y los códigos utilizados.

## Manipulación de Dataset

### Cambios de nomenclatura

Se renombraron manualmente archivos y hojas aplicando camelCase para eliminar espacios y caracteres especiales, facilitando el trabajo programático en R:

*Archivo: Informe anual por biblioteca CLM 2015.xlsx → VisAct2Material.xlsx*

*Hoja: Informe Anual por biblioteca → InformeAnualPorBiblioteca*

**Procesamiento en R**

* Renombrado de columnas: Eliminación de espacios, tildes y puntos
* Filtrado: Exclusión de registros con valores NA o gastos = 0 en variables críticas
* Variable derivada: eficiencia = (visitas / gastos) × 1000 (visitas por cada 1.000€)

Los valores nulos o cero impiden calcular métricas y distorsionan la visualización de correlación.

## Visualización desarrollada

### Gráfico de Dispersión: Eficiencia de Bibliotecas

**Variables representadas: Eje X:** Gastos corrientes (€); **Eje Y:** Número de visitas anuales; **Tamaño del punto:** Volumen de la colección; **Color:** Provincia de ubicación

Con [este código](https://github.com/Ppantaleo/Informes-Bibliometricos/blob/main/UC3M/VisAct2.md#4-visualizaci%C3%B3n-principal), se exportó [este gráfico](https://github.com/Ppantaleo/Informes-Bibliometricos/blob/main/UC3M/VisAct2.md#gr%C3%A1fico).

### Correlación

Con [este código](https://github.com/Ppantaleo/Informes-Bibliometricos/blob/main/UC3M/VisAct2.md#5-estad%C3%ADsticas-clave), se creó [esta tabla](https://github.com/Ppantaleo/Informes-Bibliometricos/blob/main/UC3M/VisAct2.md#correlaci%C3%B3n).

## Herramienta utilizada

### R + ggplot2

**Descripción:** R es un lenguaje de programación de código abierto especializado en análisis estadístico y visualización de datos. El paquete ggplot2 implementa la "gramática de gráficos", permitiendo construir visualizaciones complejas de forma estructurada y reproducible.

**Paquetes utilizados:**

* readxl: Lectura de archivos Excel
* dplyr: Manipulación y transformación de datos
* ggplot2: Creación de visualizaciones
* scales: Formateo de etiquetas numéricas

## Razonamiento y análisis

### Selección del tipo de gráfico

Se eligió un **gráfico de dispersión** ya que permite visualizar la correlación entre dos variables cuantitativas (gastos e inversión vs. uso ciudadano). También, permite añadir dimensiones adicionales mediante el tamaño (colección) y color (provincia), enriqueciendo el análisis sin saturar la visualización. Por último, también facilita detectar bibliotecas eficientes (muchas visitas con bajo presupuesto) vs. ineficientes (pocas visitas con alto presupuesto).

### Razones para la elección de R

Se eligió R ya que es una herramienta de acceso abierto y altamente documentada. La documentación y la comunidad que residen detrás de su desarrollo, permite una fase de aprendizaje rápida y un mantenimiento adecuado en usos intensivos. Adicionalmente, la elección también se debió a la familiaridad que tengo con la herramienta en usos previos. Las librerías seleccionadas permiten la elaboración de gráficos elocuentes en base a inputs como xlsx.

La elección de R también se debió a que fue una herramienta abordada en las clases. El uso de R se realizó mediante la interfaz gráfica de R Studio.

### Análisis de los resultados y conclusiones

**Correlación identificada:** La correlación entre gastos y visitas resultó **positiva moderada** (r ≈ 0.65-0.75 según datos observados), lo que indica que existe relación entre inversión y uso, pero no es determinante.

**Diversidad en eficiencia operativa:** Existe una heterogeneidad significativa en el rendimiento de las bibliotecas de Castilla-La Mancha, con ratios de eficiencia que varían hasta 70 veces entre la más y la menos eficiente

## Uso de IA

En esta actividad **se usó** el asistente Claude de Anthropic para la revisión y apoyo del código R implementado.

En esta actividad, **no se usó** asistentes de IA para la instalación de R ni para la instalación o ejecución de R studio. Tampoco se utilizó para crear el repositorio de la asignatura en GitHub, ni para elaborar el directorio en local y vincularlo mediante comandos git al repositorio externo. De igual manera, la modificación de nombres de archivos, la localización de hojas del dataset, la contextualización de la actividad y la elaboración y carga del informe se realizó sin intervención de IA.

Finalmente, toda búsqueda realizada en motores de búsqueda puede haber estado medida, voluntaria o involuntariamente, por IA. Esto, atendiendo a la nuevas implementaciones de IA en búsqueda asistida según comento aquí: [Pantaleo, P. I. (2025, June 19)](https://doi.org/10.62059/9x4p2w).