

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

Tecnologías de la información



Extracción de conocimiento en base de datos

V.1. Reporte de investigación de técnicas de visualización (50%)

IDGS91N

Presenta:

Brahn Raudales

Docente:

Enrique Mascote

sábado, 29 de noviembre de 2025

1. Introducción

Vivimos en una era en la cual la cantidad de datos generados y disponibles crece exponencialmente. Para extraer conocimiento útil (patrones, tendencias, valores atípicos) de esos datos, no basta con aplicar algoritmos: también es crucial presentar los resultados de forma clara, comprensible y significativa. Por ello, las **técnicas de visualización y representación de información** juegan un papel fundamental en el proceso de extracción de conocimiento. Este reporte investiga dichas técnicas, explica para qué sirven, cómo se usan y por qué son esenciales en proyectos de análisis de datos.

2. Qué es la visualización de datos

La visualización de datos es la **representación gráfica de información y datos** mediante gráficos, mapas, diagramas y otros recursos visuales. Su propósito principal es facilitar la comprensión de conjuntos de datos complejos, revelar tendencias, patrones, correlaciones o valores atípicos, y comunicar hallazgos de forma efectiva.

Más allá de ser simplemente decoración, la visualización es una herramienta clave para la **toma de decisiones basada en datos**, para comunicar resultados a audiencias técnicas o no técnicas, y para guiar el análisis exploratorio en procesos de extracción de conocimiento.

3. Principales técnicas de visualización y representación de información

A continuación se describen varias técnicas comunes, con explicación y ejemplos de cuándo conviene usarlas.

3.1 Gráficos básicos

Estos son los gráficos más usados por su simplicidad y eficacia.

- **Gráfico de barras / columnas:** útil para comparar valores entre categorías. Por ejemplo, ventas por producto, número de clientes por región, etc.
- **Gráfico de líneas:** ideal para mostrar evoluciones o tendencias en el tiempo, como evolución de ventas mensuales, temperatura a lo largo de días, crecimiento de usuarios, etc.
- **Gráfico de dispersión (scatter plot):** para explorar relaciones entre dos variables cuantitativas — por ejemplo, edad vs ingreso, glucosa vs presión, etc. Permite ver correlaciones, agrupamientos, outliers.
- **Histograma:** para ver la distribución de una variable numérica — por ejemplo distribución de edades, ingresos, valores de una medición. Útil para ver sesgos, moda, dispersión.

Cuándo usar cada uno depende de qué quieras mostrar: comparaciones discretas → barras; tendencias → líneas; relaciones → dispersión; distribuciones → histograma.

3.2 Mapas de calor (Heatmaps)

Los **heatmaps** usan colores para representar valores en una matriz o en coordenadas.

Son útiles cuando se tienen muchas variables o mediciones, y se quiere visualizar intensidad, densidad, correlaciones, o valores altos/bajos en una matriz.

Ejemplo de uso: mostrar la correlación entre múltiples variables, identificar valores atípicos o zonas de concentración en datos espaciales o de densidad.

"heatmap ejemplo correlación variables",

"mapa de calor densidad datos ejemplo",

"heatmap uso big data visualización",

"heatmap correlation matrix seaborn example"

],"num_per_query":1}?

3.3 Infografías y Dashboards

Cuando se quiere presentar resultados de manera integrada —gráficos, KPIs, resúmenes, textos explicativos— las **infografías** o **dashboards** son ideales. Permiten condensar información clave, combinar distintos gráficos, mostrar indicadores, mapas, tablas, etc.

Por ejemplo, un dashboard de ventas incluye una gráfica de tendencias, un mapa de regiones, una tabla de top productos, indicadores de totales y filtros interactivos para explorar por periodo o categoría.

3.4 Visualización multivariante / alta dimensión

Cuando los datos tienen muchas variables, se pueden usar técnicas como:

- **Scatter-matrix** (matriz de dispersión): múltiples gráficas de dispersión entre pares de variables para ver relaciones.
- **Coordenadas paralelas**: cada variable es un eje vertical paralelo; cada dato se representa como una línea que atraviesa los ejes — útil para visualizar multivariables.
- **Gráficos de burbujas, radar, treemaps, network graphs**, etc., dependiendo del tipo de dato.

Este tipo de visualización permite explorar relaciones complejas, clusters, correlaciones múltiples, outliers, y estructuras en datos de muchas dimensiones.

4. Relación entre visualización y extracción de conocimiento (KDD / Data Mining)

El proceso de extracción de conocimiento (KDD — *Knowledge Discovery in Databases*) o minería de datos consta de varias fases: recolección, limpieza, transformación, modelado y presentación de resultados. La **visualización de datos** interviene tanto en la fase exploratoria como en la fase de interpretación y comunicación:

- En la **fase exploratoria**: permite identificar patrones, correlaciones, valores atípicos, distribución, agrupamientos, etc. antes de aplicar modelos. Así puedes decidir qué variables transformar, descartar, agrupar o tratar.
- En la **fase de presentación / comunicación**: una visualización clara facilita que quienes reciben los resultados (compañeros, profesores, directivos) comprendan los hallazgos, decisiones, recomendaciones.
- En la **fase de interpretación de modelos**: tras modelar (clustering, clasificación, regresión...), se usan gráficos y dashboards para mostrar resultados, comparar modelos, explicar decisiones.

Por lo tanto, la visualización no es un “extra” decorativo: es una parte esencial del flujo de extracción de conocimiento.

5. Buenas prácticas para visualización de datos

Para que una visualización cumpla su función de comunicar correctamente, conviene seguir ciertas buenas prácticas:

- Elegir el tipo de gráfico adecuado según la pregunta de análisis.
- Etiquetar claramente los ejes, incluir leyendas cuando haya más de un grupo, usar títulos descriptivos.
- Evitar distorsionar los datos: no manipular ejes de forma engañosa, no alterar proporciones, evitar “chartjunk” (elementos decorativos innecesarios que distraen).

- Mantener simplicidad cuando sea posible, priorizando claridad sobre complejidad gráfica.
- Utilizar colores y estilos consistentes, tener contraste suficiente, considerar accesibilidad visual.

6. Conclusiones

- La visualización de datos es una herramienta central en la extracción de conocimiento: ayuda a explorar, interpretar y comunicar datos de forma intuitiva.
- Existen múltiples técnicas de visualización —desde gráficos básicos hasta dashboards e infografías—, cada una adecuada para distintos tipos de datos y objetivos.
- Una buena visualización permite revelar patrones, tendencias, correlaciones o anomalías que serían difíciles de detectar en tablas de datos.
- Las buenas prácticas en diseño son fundamentales: un gráfico mal diseñado puede inducir a errores de interpretación o incluso engañar.
- Para trabajos académicos o profesionales de minería de datos, combinar visualización + análisis + explicación narrative es muy valioso: facilita la comprensión, la toma de decisiones y la comunicación de hallazgos.

7. Referencias

- DataCamp. (2024). *11 técnicas de visualización de datos para cada caso de uso con ejemplos*.
- Atlassian. (2022). *Essential Charts & Graphs for Data Visualization*.
- Tableau. (s.f.). ¿Qué es la visualización de datos?
- IBM. (s.f.). *Visualización de datos: definición y ejemplos*.
- Analytics Vidhya. (2023). *Top 20 Data Visualization Examples*.
- GeeksforGeeks. (2025). *Charts and Graphs for Data Visualization*.
- ES EID. (2023). *Mejores técnicas de visualización de datos*.
- OpenAI. (2025). *ChatGPT* [Modelo de lenguaje]. OpenAI. <https://chat.openai.com/>