

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

TECNOLOGÍAS DE LA INFORMACIÓN



EXTRACCION DE CONOCIMIENTO EN BASES DE DATOS

**V.1. REPORTE DE INVESTIGACIÓN DE TÉCNICAS DE
VISUALIZACIÓN**

IDGS91N

PRESENTA:

SEBASTIÁN ACOSTA ORTIZ

DOCENTE:

**LUIS ENRIQUE MASCOTE
CANO**

Chihuahua, Chih., 30 de noviembre de 2025

Introducción

La cantidad de datos generados diariamente en organizaciones, gobiernos y aplicaciones digitales hace que la visualización de datos sea una herramienta esencial para transformar números en información comprensible. La visualización permite representar datos mediante gráficos, mapas, diagramas y otros recursos visuales, facilitando la identificación de patrones, tendencias y anomalías de manera más rápida que con tablas numéricas tradicionales.

El objetivo de este reporte es presentar los fundamentos teóricos y prácticos de la visualización de datos, describiendo conceptos básicos, tipos de gráficos, herramientas y bibliotecas especializadas, así como el proceso de *storytelling* con datos. El alcance de la investigación abarca desde los principios de diseño visual y elección de gráficos según el tipo de variable, hasta un análisis comparativo entre herramientas de Business Intelligence y bibliotecas de programación.

Fundamentos de Visualización de Datos

Conceptos básicos

Sistemas de coordenadas

Un sistema de coordenadas es el marco de referencia en el que se representan los datos. Los más utilizados son:

- **Coordenadas cartesianas:** utilizan ejes perpendiculares (x , y , y en ocasiones z) para ubicar puntos numéricos. Es el sistema más común en gráficos de barras, líneas y dispersión.
- **Coordenadas polares:** representan puntos mediante un ángulo y una distancia radial (r , θ). Se usan en gráficos circulares, de radar o de rosas de viento.

- **Coordenadas geográficas:** basadas generalmente en latitud y longitud para representar datos geoespaciales en mapas.

Tipos de ejes

- **Ejes lineales:** el incremento entre marcas es constante; adecuados para la mayoría de las variables numéricas.
- **Ejes logarítmicos:** la escala crece de forma multiplicativa; útiles cuando los datos varían en varios órdenes de magnitud (ej. datos financieros o científicos).
- **Ejes categóricos:** representan categorías discretas (países, productos, grupos de edad).
- **Ejes de tiempo (series temporales):** muestran fechas u horas en el eje horizontal para analizar tendencias.

Esquemas de colores y su importancia psicológica

Los colores no solo diferencian categorías, también transmiten significados emocionales y jerárquicos:

- **Esquemas secuenciales:** van de claro a oscuro para representar valores crecientes (ej. intensidad de una métrica).
- **Esquemas divergentes:** utilizan dos colores extremos con un punto medio neutro, útiles para mostrar desviaciones respecto a un valor de referencia (por ejemplo, ganancias vs pérdidas).
- **Esquemas categóricos:** emplean colores distintos para categorías sin orden.

Además, es importante considerar la **percepción humana** y el **daltonismo**; muchas guías recomiendan evitar combinaciones problemáticas (como rojo-verde) y utilizar paletas diseñadas para ser perceptualmente uniformes.

Principios de diseño visual

Algunos principios clave:

- **Claridad:** evitar elementos innecesarios (chartjunk) y priorizar la legibilidad.
- **Jerarquía visual:** usar tamaño, color y posición para guiar la atención hacia lo más importante.
- **Consistencia:** mantener estilos, tipografías y escalas coherentes.
- **Proporcionalidad:** no distorsionar escalas ni truncar ejes sin justificación.
- **Contexto:** incluir títulos, etiquetas, unidades y fuentes de datos.

Tipos de representación gráfica

Para cada tipo se indica qué es, cuándo usarlo y un ejemplo típico.

Visualización de cantidad

a) Gráficos de barras y columnas

- **Definición:** representan valores mediante la longitud de barras (horizontales) o columnas (verticales).
- **Cuando usarlo:** comparar cantidades entre categorías (ventas por producto, usuarios por país).
- **Ejemplo visual:** columnas verticales que muestran el número de clientes por segmento.

b) Pictogramas

- **Definición:** usan íconos o símbolos repetidos en lugar de barras sólidas.
- **Cuándo usarlo:** cuando se quiere comunicar de manera más “amigable” o divulgativa.

- **Ejemplo:** íconos de personas para mostrar población.

Visualización de distribución

a) Histogramas

- **Definición:** agrupan datos numéricos en intervalos (bins) y muestran la frecuencia de valores.
- **Cuando usarlo:** analizar la distribución de una variable continua (ej. ingresos, edades).

b) Box plots (diagramas de caja)

- **Definición:** resumen distribución mediante mediana, cuartiles y posibles valores atípicos.
- **Cuando usarlo:** comparar distribuciones de varias categorías en una sola gráfica.

c) Gráficos de violín

- **Definición:** combinan box plot con una estimación de densidad, mostrando la forma de la distribución.
- **Cuando usarlo:** cuando se desea ver diferencias en la forma de la distribución entre grupos.

Visualización de proporción

a) Gráficos de pastel y donut

- **Definición:** muestran la proporción de cada categoría como porciones de un círculo.
- **Cuando usarlo:** pocas categorías y diferencias grandes; casos sencillos.
- **Limitación:** difícil comparar muchas categorías o proporciones similares.

b) Treemaps

- **Definición:** representan proporciones mediante rectángulos anidados.
- **Cuando usarlo:** mostrar composición jerárquica (por ejemplo, participación de mercado por línea y sublínea).

Visualización de relación XY

a) Diagramas de dispersión (scatter plots)

- **Definición:** ubican cada observación como un punto en función de dos variables numéricas (x, y).
- **Cuando usarlo:** descubrir relaciones, correlaciones o clusters entre variables (ej. edad vs ingreso).

b) Gráficos de burbujas

- **Definición:** extensión del scatter plot donde el tamaño del punto representa una tercera variable.
- **Cuando usarlo:** comparar tres variables simultáneamente (ej. población, PIB y esperanza de vida por país).

Visualización de datos geoespaciales

a) Mapas de calor geográficos (heatmaps)

- **Definición:** muestran intensidad de un valor sobre un mapa mediante colores.
- **Cuando usarlo:** densidad de eventos, concentración de ventas, tráfico, etc.

b) Mapas de coropletas

- **Definición:** colorean áreas geográficas (países, estados, municipios) según valores numéricos.

- **Cuando usarlo:** comparar indicadores entre regiones (ej. tasa de desempleo).

c) Cartogramas

- **Definición:** distorsionan la forma o tamaño de las regiones para representar una variable (por ejemplo, tamaño proporcional a población).
- **Cuándo usarlo:** visualizaciones más conceptuales o divulgativas donde se quiere enfatizar la magnitud de una variable sobre la geografía real.

Visualización de incertidumbre

a) Gráficos de error (error bars)

- **Definición:** agregan líneas de error a una estimación (media) para indicar variabilidad o confianza.
- **Cuando usarlo:** resultados de experimentos, intervalos en modelos estadísticos.

b) Intervalos de confianza (confidence intervals)

- **Definición:** se muestran como bandas sombreadas alrededor de una línea o como barras; indican el rango probable de un parámetro.
- **Cuándo usarlo:** comunicar la incertidumbre en predicciones o estimaciones (ej. pronósticos).

Proceso de Storytelling con Datos

Definición y componentes

Storytelling con datos es la práctica de combinar **análisis cuantitativo, visualizaciones y narrativa** para comunicar hallazgos de forma clara, persuasiva y adaptada a la audiencia. No se trata solo de mostrar gráficos, sino de construir una historia con inicio, desarrollo y cierre, donde los datos respaldan el mensaje principal.

Elementos clave:

- **Contexto:** problema, preguntas, marco de referencia.
- **Narrativa:** hilo conductor que guía de un punto A (situación actual) a un punto B (conclusión o decisión).
- **Visualización:** gráficos que refuerzan el mensaje y hacen los datos comprensibles.

Etapas del proceso

1. **Comprensión de la audiencia**

- ¿Quiénes son? ¿Técnicos, directivos, público general?
- ¿Qué nivel de detalle y tecnicismo es apropiado?

2. **Definición del mensaje clave**

- Resumir en una frase lo que se quiere comunicar (ej. "Las ventas crecieron en el norte, pero cayeron en el sur").

3. **Selección de datos relevantes**

- Elegir solo las variables y periodos que apoyan el mensaje.
- Limpiar y preparar los datos.

4. **Diseño de la narrativa visual**

- Ordenar los gráficos en secuencia lógica.
- Usar anotaciones, resaltados y colores para guiar la atención.

5. **Presentación efectiva**

- Ajustar el nivel de detalle al tiempo disponible.
- Practicar la explicación del "por qué" detrás de cada visualización.

Mejores prácticas

- **Estructurar la historia:** contexto → conflicto o hallazgo → implicaciones → recomendación.
- **Usar un gráfico por idea principal:** evitar sobrecargar visuales con demasiada información.
- **Resaltar lo importante:** colores de énfasis para la serie o categoría clave.
- **Evitar errores comunes:**
 - Gráficos 3D innecesarios que dificultan lectura.
 - Ejes truncados sin indicarlo.
 - Paletas de colores confusas o sin contraste suficiente.
- **Casos exitosos:** dashboards de ventas que resumen KPIs, visualizaciones interactivas en medios de comunicación para explicar elecciones, pandemias o fenómenos económicos.

Herramientas de Visualización (Business Intelligence)

A continuación, se describen tres herramientas de BI: **Tableau**, **Power BI** y **Google Data Studio** (Looker Studio).

Tableau

Descripción general:

Herramienta de Business Intelligence enfocada en la creación de visualizaciones interactivas y dashboards. Permite conectarse a múltiples fuentes de datos y crear vistas mediante operaciones de arrastrar y soltar.

Características principales:

- Conexión a bases de datos locales y en la nube.
- Dashboards interactivos, filtros y acciones.
- Amplio catálogo de gráficos, mapas y tablas.
- Publicación en Tableau Server o Tableau Online.

Ventajas:

- Muy potente para análisis visual exploratorio.
- Gran comunidad y recursos de aprendizaje.

Desventajas:

- Licenciamiento de costo relativamente alto.
- Curva de aprendizaje intermedia para usuarios novatos.

Tipos de gráficos soportados:

Barras, líneas, áreas, mapas, scatter, box plots, histogramas, mapas de calor, treemaps, entre otros.

Casos de uso recomendados:

Análisis de ventas, seguimiento de KPIs, dashboards ejecutivos, análisis geográfico.

Power BI

Descripción general:

Plataforma de análisis de datos de Microsoft que permite conectar, transformar y visualizar información para generar reportes y dashboards interactivos.

Características principales:

- Integración nativa con Excel y otros productos de Microsoft.
- Modelado de datos mediante Power Query y DAX.

- Publicación en la nube (Power BI Service).
- Amplia galería de visualizaciones y visuales personalizados.

Ventajas:

- Versión de escritorio gratuita (Power BI Desktop).
- Integración con ecosistema Microsoft 365 y Azure.

Desventajas:

- Algunas funciones avanzadas requieren licencias Pro o Premium.
- La complejidad crece con modelos de datos grandes.

Tipos de gráficas soportadas:

Gráficos de barras, columnas, líneas, áreas, embudos, mapas, matrices, tarjetas de KPI, treemaps, etc.

Casos de uso recomendados:

Reportes corporativos, paneles de inteligencia de negocio, seguimiento financiero.

Google Data Studio (Looker Studio)

Descripción general:

Herramienta gratuita de Google para crear reportes interactivos y dashboards conectados a fuentes como Google Analytics, BigQuery, Hojas de Cálculo, entre otros.

Características principales:

- Basado en la web, sin instalación.
- Integración con ecosistema Google.
- Compartición sencilla (enlace, permisos).

Ventajas:

- Gratuita.
- Ideal para marketing digital y analítica web.

Desventajas:

- Menos robusta que Tableau/Power BI para modelos complejos.
- Dependencia de la cuenta Google.

Tipos de gráficas soportadas:

Series temporales, barras, pasteles, mapas geográficos, tablas dinámicas, scorecards.

Casos de uso recomendados:

Reportes de campañas de marketing, análisis de tráfico web, monitoreo de redes sociales.

Bibliotecas de Visualización

Se describen cuatro bibliotecas: **Matplotlib**, **Seaborn**, **Plotly** y **D3.js**.

Matplotlib

- **Lenguaje:** Python.
- **Descripción y filosofía:** Biblioteca base de visualización en Python; permite crear gráficos estáticos, animados e interactivos de bajo nivel.
- **Características principales:**
 - Altamente configurable.
 - Permite controlar cada elemento del gráfico (ejes, leyendas, estilos).
- **Nivel de complejidad:** Medio; flexible pero más “verborreico” que otras bibliotecas de alto nivel.

- **Tipos de gráficas:** líneas, barras, histogramas, dispersión, mapas de calor, etc.
- **Ejemplo de código básico:**

```
import matplotlib.pyplot as plt
```

```
x = [1, 2, 3, 4]
```

```
y = [3, 5, 2, 6]
```

```
plt.plot(x, y, marker='o')
```

```
plt.title("Ejemplo con Matplotlib")
```

```
plt.xlabel("Eje X")
```

```
plt.ylabel("Eje Y")
```

```
plt.show()
```

Seaborn

- **Lenguaje:** Python.
- **Descripción y filosofía:** Biblioteca de alto nivel construida sobre Matplotlib, orientada a gráficos estadísticos atractivos e informativos; se integra muy bien con pandas.
- **Características principales:**
 - Temas estéticos por defecto.
 - Funciones para distribuciones, relaciones y categorías (histplot, boxplot, violinplot, pairplot, heatmap).
- **Nivel de complejidad:** Bajo-medio; ofrece funciones de alto nivel con pocos parámetros.

- **Tipos de gráficas:** distribuciones, relaciones bivariadas, matrices de correlación, mapas de calor, etc.
- Ejemplo de código básico:

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

```
tips = sns.load_dataset("tips")
```

```
sns.scatterplot(data=tips, x="total_bill", y="tip", hue="time")
```

```
plt.title("Relación entre cuenta total y propina")
```

```
plt.show()
```

Plotly (Python)

- **Lenguaje:** Python (también JS, R).
- **Descripción y filosofía:** Biblioteca para gráficos **interactivos** basados en web; integra bien con notebooks y dashboards.
- **Características principales:**
 - Interactividad (zoom, hover, selección).
 - Capacidad de exportar a HTML.
- **Nivel de complejidad:** Medio; más sencillo con *plotly.express*.
- **Tipos de gráficas:** series, dispersión, mapas, 3D, sunbursts, treemaps, etc.
- Ejemplo de código básico:

```
import plotly.express as px

df = px.data.iris()

fig = px.scatter(df, x="sepal_width", y="sepal_length",
                 color="species", title="Iris con Plotly")

fig.show()
```

5.4 D3.js

- **Lenguaje:** JavaScript.
- **Descripción y filosofía:** Biblioteca de bajo nivel para crear visualizaciones dinámicas e interactivas directamente en el navegador usando SVG, HTML y CSS.
- **Características principales:**
 - Enlaza datos a elementos del DOM (*Data-Driven Documents*).
 - Altísima flexibilidad para visualizaciones personalizadas.
- **Nivel de complejidad:** Alto; curva de aprendizaje considerable, pero muy poderosa.
- **Tipos de gráficos:** prácticamente cualquier visualización (barras, redes, mapas, gráficos de tiempo, animaciones).
- **Ejemplo de código básico (esquemático):**

```
const data = [10, 20, 30, 40];
```

```
d3.select("body")
  .selectAll("div")
  .data(data)
```

```
.enter()  
  
.append("div")  
  
.style("width", d => d + "px")  
  
.text(d => d);
```

6. Conclusiones

La visualización de datos es un componente esencial del análisis moderno, ya que convierte información compleja en representaciones comprensibles que facilitan la toma de decisiones. Elegir la visualización adecuada depende del tipo de dato (cantidad, proporción, distribución, relación, ubicación geográfica o incertidumbre) y del mensaje que se desea transmitir.

Las **herramientas de BI** como Tableau, Power BI o Google Data Studio están orientadas a usuarios de negocio, con interfaces de arrastrar y soltar y capacidades para construir dashboards sin necesidad de programar. Las **bibliotecas de visualización**, por su parte, brindan un mayor nivel de control y flexibilidad a desarrolladores y científicos de datos, integrándose dentro de flujos de trabajo de análisis y modelado.

Personalmente, el aprendizaje de estas técnicas y herramientas resalta la importancia de pensar primero en la **historia que queremos contar con los datos** y después escoger el tipo de gráfico, paleta de colores y herramienta que mejor apoye ese relato, evitando visualizaciones confusas o engañosas.

Referencias

Hunter, J. D. et al. (2025). *Matplotlib: Visualization with Python*. Matplotlib Development Team. Recuperado de [Matplotlib+2Matplotlib+2](#)

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. [Journal of Open Source Software+1](#)

Tableau Software. (s. f.). *¿Qué es la visualización de datos?* Recuperado de [Tableau+1](#)

Microsoft. (2025). *Power BI documentation*. Recuperado de [Microsoft Learn+1](#)

Bostock, M. et al. (2011). D3: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*. Recuperado de [vis.stanford.edu+1](#)