

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

TECNOLOGÍAS DE LA INFORMACIÓN



EXTRACCIÓN DE CONOCIMIENTO EN BASES DE DATOS

V.1. REPORTE DE INVESTIGACIÓN DE TÉCNICAS DE VISUALIZACIÓN

IDGS91N

PRESENTA:

REGINA CHÁVEZ TAMAYO - 6521110019

DOCENTE:

LUIS ENRIQUE MASCOTE CANO

Chihuahua, Chih., 30 de noviembre de 2025

Introducción

La visualización de datos es una disciplina fundamental dentro del análisis de información y la ciencia de datos. Su principal función es transformar datos complejos en representaciones visuales comprensibles que faciliten la interpretación, la toma de decisiones y la comunicación de hallazgos. A diferencia de las tablas o el análisis exclusivamente numérico, las visualizaciones permiten identificar patrones, tendencias y anomalías de manera inmediata.

Objetivo del reporte:

Investigar y documentar las principales técnicas, herramientas y bibliotecas de visualización de datos, así como explicar el proceso de storytelling aplicado al análisis de datos. El reporte abarca fundamentos teóricos, tipos de gráficos, buenas prácticas, herramientas profesionales y tecnologías de programación para visualización.

Alcance:

Este documento cubre conceptos esenciales, tipos de representación, herramientas de Business Intelligence, bibliotecas de programación en Python y JavaScript, y el proceso completo de narración con datos desde el contexto hasta la presentación final.

Fundamentos de Visualización de Datos

Conceptos básicos

Sistemas de coordenadas

Los sistemas de coordenadas son la base sobre la cual se construyen los gráficos. Los más utilizados son:

- **Sistema cartesiano:** Representa datos mediante los ejes X y Y. Es el sistema más común para líneas, barras y dispersión.
- **Sistema polar:** Utiliza un ángulo y una distancia radial. Se usa en gráficos como los de pastel o radar.
- **Coordenadas geográficas:** Basadas en latitud y longitud, empleadas en mapas.

Tipos de ejes

- **Eje numérico:** Representa cantidades continuas (ventas, temperaturas).
- **Eje categórico:** Separa grupos o categorías discretas (meses, productos).
- **Eje logarítmico:** Usado para representar datos con crecimiento exponencial o grandes rangos de valores.

Esquemas de colores y psicología del color

El color influye en la percepción del usuario. Su importancia radica en:

- Diferenciar categorías
- Indicar intensidad o magnitud
- Guiar la atención del lector

Buenas prácticas:

- Usar paletas consistentes
- Evitar colores demasiado saturados
- Elegir colores accesibles para personas con daltonismo (paletas colorblind-friendly)

Principios de diseño visual

1. **Claridad:** El gráfico debe ser fácil de interpretar.
2. **Jerarquía visual:** Lo más importante debe ser lo más visible.
3. **Minimización del ruido:** Evitar elementos innecesarios.
4. **Proporcionalidad:** Mantener escalas correctas y evitar distorsión.
5. **Consistencia:** Mismos colores y estilos para elementos equivalentes.

Tipos de representación gráfica

Visualización de cantidad

Gráfico de barras

- **Definición:** Representa cantidades por categoría mediante barras horizontales.
- **Cuándo usarlo:** Comparaciones entre categorías individuales.
- **Ejemplo visual sugerido:** Ventas por producto.

Gráfico de columnas

- **Definición:** Igual que barras, pero vertical.
- **Cuándo usarlo:** Evoluciones con pocos puntos en el tiempo.
- **Ejemplo:** Ingresos trimestrales.

Pictogramas

- **Definición:** Uso de íconos que representan cantidades.
- **Cuándo usarlo:** Datos educativos o públicos.
- **Ejemplo:** 1 ícono = 10 personas.

Visualización de distribución

Histogramas

- **Definición:** Agrupan datos en rangos y muestran su frecuencia.
- **Cuándo usarlo:** Análisis de forma, sesgo y dispersión de un conjunto.
- **Ejemplo:** Distribución de edades.

Box plots

- **Definición:** Gráfico basado en cuartiles, mediana y outliers.
- **Cuándo usarlo:** Comparar dispersión entre grupos.
- **Ejemplo:** Sueldos por departamento.

Gráficos de violín

- **Definición:** Combina boxplot con densidad de kernel.
- **Cuándo usarlo:** Comparar distribuciones complejas.
- **Ejemplo:** Puntajes por categoría.

Visualización de proporción

Gráfico de pastel / pie

- **Definición:** Muestra la proporción de cada categoría respecto al total.
- **Cuándo usarlo:** Datos simples con pocas categorías.

Gráfico donut

- **Definición:** Variante del pie con centro vacío.
- **Cuándo usarlo:** Cuando se quiere incluir un valor total en el centro.

Treemaps

- **Definición:** Representación jerárquica mediante rectángulos proporcionados.
- **Cuándo usarlo:** Proporciones complejas y jerárquicas.

Visualización de relación XY

Diagrama de dispersión (scatterplot)

- **Definición:** Representa puntos con coordenadas x-y.
- **Cuándo usarlo:** Relación entre dos variables.

Gráfico de burbujas

- **Definición:** Variante que agrega un tercer valor como tamaño de la burbuja.
- **Cuándo usarlo:** Comparar correlación con magnitud adicional.

Visualización de datos geoespaciales

Mapas de calor

- **Definición:** Representan densidad o valores mediante colores.
- **Uso:** Tráfico, densidad poblacional.

Coropletas

- **Definición:** Mapas coloreados por región según una variable.
- **Uso:** Desempleo por estado.

Cartogramas

- **Definición:** Deforman regiones según valores asociados.
- **Uso:** Representación electoral.

Visualización de incertidumbre

Gráficos de error

- **Definición:** Líneas que representan variabilidad o intervalos.
- **Uso:** Resultados científicos.

Intervalos de confianza

- **Definición:** Sombreado que indica el rango probable donde está el verdadero valor.
- **Uso:** Pronósticos.

Proceso de Storytelling con Datos

¿Qué es el Storytelling con datos?

El storytelling con datos combina visualizaciones y narrativa para comunicar insights de forma clara y persuasiva. Su objetivo no es solo mostrar gráficos, sino contar una historia con propósito.

Componentes clave

1. **Contexto:** Situar al lector en el problema.
2. **Narrativa:** Hilo conductor que conecta visualizaciones.
3. **Visualización:** Evidencia visual que apoya la historia.

Etapas del proceso

1. Comprensión de la audiencia

Identificar su nivel técnico, intereses y necesidades.

2. Definición del mensaje clave

Determinar la idea principal que debe recordar el público.

3. Selección de datos relevantes

Filtrar solo los datos que aportan al mensaje.

4. Diseño de la narrativa visual

Ordenar gráficos de forma lógica y coherente.

5. Presentación efectiva

Usar lenguaje claro, énfasis visual y evitar saturación.

Mejores prácticas

Cómo estructurar una historia con datos

- Iniciar con el contexto
- Presentar evidencia visual
- Destacar el insight clave
- Concluir con una acción o recomendación

Errores comunes

- Usar demasiados gráficos
- Elegir visualizaciones incorrectas
- Colores distractores
- Mensajes contradictorios

Casos exitosos

- Dashboards ejecutivos
- Informes de ventas trimestrales
- Visualizaciones COVID-19

Herramientas de Visualización

Tableau

Descripción general

Plataforma líder en visualización interactiva.

Características

- Arrastrar y soltar

- Dashboards dinámicos
- Conexión a múltiples fuentes

Ventajas

- Muy potente
- Ideal para análisis exploratorio

Desventajas

- Costo elevado
- Licencias por usuario

Tipos de gráficas

Barras, líneas, mapas, treemaps, dispersión, etc.

Casos de uso

Reportes ejecutivos, análisis comerciales.

Power BI

Descripción general

Herramienta de Microsoft enfocada en análisis y dashboards.

Características

- Integración con Excel
- Power Query
- Modelado semántico

Ventajas

- Accesible
- Amplia comunidad

Desventajas

- Limitaciones en versión gratuita

Google Data Studio (Looker Studio)

Características

- Online y gratuito
- Conexión directa con Google Analytics
- Dashboards colaborativos

Ventajas

- Fácil de usar
- Integración con otras herramientas Google

Desventajas

- Menos potente que Tableau

Bibliotecas de Visualización

Matplotlib (Python)

Filosofía: Control total sobre cada aspecto del gráfico.

Complejidad: Media–alta.

Tipos: Líneas, barras, scatter, histogramas.

Ejemplo simple:



```
● ● ● Matplotlib
1 import matplotlib.pyplot as plt
2 plt.plot([1,2,3],[4,5,6])
3 plt.show()
```

Seaborn (Python)

Filosofía: Gráficos estadísticos de alto nivel.

Complejidad: Media.

Tipos: Boxplots, violines, heatmaps.



Seaborn

```
1 import seaborn as sns  
2 sns.boxplot(data=df)
```

Plotly (Python/JS)

Filosofía: Gráficos interactivos.

Tipos: 3D, mapas, burbujas.



Plotly

```
1 import plotly.express as px  
2 px.scatter(df, x="A", y="B")
```

D3.js (JavaScript)

Filosofía: Visualizaciones web complejas y personalizadas.

Complejidad: Alta.

Tipos: Todo tipo de gráficos.



Conclusiones

La visualización adecuada permite comprender patrones, comunicar insights y tomar decisiones informadas. Las herramientas como Tableau y Power BI son ideales para dashboards ejecutivos, mientras que bibliotecas como Matplotlib, Seaborn y D3.js permiten visualizaciones personalizadas en análisis técnico.

La selección adecuada depende del propósito, la audiencia y el nivel de interactividad requerido. A nivel personal, este trabajo reafirma la importancia de combinar narrativa, diseño visual y herramientas tecnológicas para lograr presentaciones efectivas y profesionales.

Referencias

- IBM. (2024). *Data visualization explained*. <https://www.ibm.com/topics/data-visualization>
- Tableau. (2024). *Visual analytics overview*. <https://www.tableau.com>
- Power BI. (2024). *What is Power BI?* <https://learn.microsoft.com/power-bi>
- Google. (2024). *Looker Studio documentation*. <https://support.google.com/looker-studio>
- Plotly. (2024). *Plotly Python documentation*. <https://plotly.com/python>
- D3.js. (2024). *Data-Driven Documents*. <https://d3js.org>
- Seaborn. (2024). *Statistical data visualization*. <https://seaborn.pydata.org>
- Matplotlib. (2024). *Matplotlib documentation*. <https://matplotlib.org>