

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA
TECNOLOGÍAS DE LA INFORMACIÓN



Extracción de Conocimiento en Bases de Datos

DOCENTE:

Enrique Mascote

Reporte de investigación de técnicas de visualización

PRESENTA:

Iván Eduardo Martínez Martínez

Grupo:

IDGS91N

Introducción

La visualización de datos es un componente fundamental dentro del análisis moderno de información, ya que permite transformar datos complejos, numerosos o abstractos en representaciones gráficas comprensibles para el usuario. En la actualidad, la toma de decisiones dentro de organizaciones públicas y privadas depende en gran medida de la capacidad de presentar información de manera clara, visual y significativa. Por ello, comprender las técnicas, herramientas y principios de visualización se ha convertido en una competencia esencial dentro del campo del análisis de datos y la ciencia de datos.

El objetivo de este reporte es investigar, describir y analizar las principales técnicas de visualización de datos, sus fundamentos teóricos, las herramientas más utilizadas en la industria y las bibliotecas de programación para crear visualizaciones. Además, se expone el proceso de *Storytelling* con datos, un enfoque indispensable para comunicar hallazgos de manera narrativa y persuasiva.

El alcance de esta investigación abarca conceptos técnicos, fundamentos visuales, ejemplos gráficos, análisis comparativos, ventajas y desventajas de herramientas, así como buenas prácticas para la presentación profesional de datos.

Fundamentos de Visualización de Datos

Conceptos básicos

Sistemas de coordenadas

Los sistemas de coordenadas son estructuras matemáticas que permiten representar datos en un espacio gráfico. Los más utilizados en visualización son:

- **Sistema cartesiano:** Representa datos mediante los ejes X y Y; se usa en la mayoría de las gráficas tradicionales.
- **Sistema polar:** Representa la información mediante ángulos y radios; se utiliza en gráficos circulares.
- **Sistemas geoespaciales:** Basados en coordenadas geográficas (latitud-longitud) para mapas y análisis territoriales.

Tipos de ejes

Los ejes permiten interpretar la escala y el comportamiento de los datos. Entre los principales tipos se incluyen:

- **Eje lineal:** Incrementos constantes (ej. 1, 2, 3...).
- **Eje logarítmico:** Escalas exponenciales, útil para variaciones muy amplias.
- **Eje categórico:** Representa categorías o grupos no numéricos.
- **Eje temporal:** Representa fechas, periodos o timestamps.

Esquemas de colores y su importancia psicológica

El color cumple una función cognitiva clave: guía la atención, transmite estados, agrupa elementos y resalta relaciones.

La psicología del color indica, por ejemplo:

- Azul: confianza, estabilidad.
- Rojo: alerta, urgencia.
- Verde: crecimiento, equilibrio.
- Amarillo: atención, energía.

Existen también escalas:

- **Cualitativas:** Para categorías (nombres, regiones).
- **Secuenciales:** Para valores crecientes.
- **Divergentes:** Para contrastar valores positivos vs. negativos.

Principios de diseño visual

Para una visualización efectiva deben aplicarse:

- Claridad y simplicidad
- Minimizar ruido visual
- Resaltar patrones relevantes
- Usar tipografías legibles
- Mantener coherencia en colores y estilos

El objetivo final es **facilitar la comprensión inmediata del lector**

Tipos de representación gráfica

A. Visualización de cantidad

1. Gráficos de barras

Definición: Representan cantidades mediante la altura o longitud de barras.

Cuándo usarlo: Comparaciones entre categorías.

2. Gráficos de columnas

Variante vertical de las barras.

Utilizados para comparar grupos en distintos periodos.

3. Pictogramas

Usan íconos para representar cantidades.

Eficientes en presentaciones para público general.

B. Visualización de distribución

1. Histogramas

Muestran la frecuencia de datos dentro de intervalos.

Útiles para identificar patrones o sesgos.

2. Box plots

Resaltan mediana, cuartiles y valores atípicos.

Utilizados en estadística y análisis exploratorio.

3. Gráficos de violín

Combinan boxplot con densidad de distribución.

Útiles para comparar múltiples grupos.

C. Visualización de proporción

1. Gráficos de pastel

Representan partes de un todo.

Recomendables solo cuando hay pocas categorías.

2. Gráficos donut

Variante moderna del pastel.

Permiten un espacio central para un dato clave.

3. Treemaps

Representación jerárquica mediante rectángulos.

Adecuados para grandes volúmenes con proporciones.

D. Visualización de relación XY

1. Diagramas de dispersión

Muestran correlaciones entre dos variables.

2. Gráficos de burbujas

Añaden una tercera variable representada por el tamaño de cada punto.

E. Visualización geoespacial

1. Mapas de calor

Representan densidades o intensidades mediante colores.

2. Mapas coropléticos

Asignan colores a regiones geográficas según un valor.

3. Cartogramas

Alteran el tamaño de regiones según una variable (población, PIB).

F. Visualización de incertidumbre

1. Barras de error

Muestran rangos de desviación estándar.

2. Intervalos de confianza

Representan el rango probable donde se ubica un valor.

Proceso de Storytelling con Datos

Definición y componentes

El *Storytelling* con datos consiste en combinar narrativa, visualizaciones y contexto para comunicar mensajes de manera clara y persuasiva. Su objetivo es transformar análisis técnicos en historias entendibles y memorables.

Componentes clave:

- **Contexto**
- **Mensaje central**
- **Narrativa**
- **Visualizaciones**
- **Conclusión accionable**

Etapas del proceso

1. Comprensión de la audiencia

Identificar nivel de conocimiento, intereses y objetivos.

2. Definición del mensaje clave

Debe ser claro, directo y sustentado en datos.

3. Selección de datos relevantes

Solo se debe mostrar información útil para el mensaje.

4. Diseño de la narrativa visual

Organizar información en una secuencia lógica.

5. Presentación efectiva

Debe ser simple, visual, atractiva y basada en evidencias.

Mejores prácticas

- Mantener una estructura narrativa: inicio, desarrollo, cierre.
- No saturar de gráficas: menos es más.
- Evitar gráficos innecesarios (pie charts excesivos).
- Usar colores consistentes y legibles.
- Presentar conclusiones accionables.

Ejemplos de casos exitosos:

Visualizaciones del New York Times, Our World in Data, Gapminder.

Herramientas de Visualización

A continuación se describen tres herramientas de Business Intelligence.

Tableau

Descripción general:

Herramienta de BI especializada en visualizaciones interactivas.

Características:

- Dashboards interactivos
- Arrastrar y soltar
- Conexiones múltiples a bases de datos

Ventajas:

- Excelente diseño
- Muchas opciones de visualización

Desventajas:

- Licencia costosa

Tipos de gráficas:

Mapas, barras, líneas, dispersión, treemaps, etc.

Caso de uso:

Empresas que requieren dashboards ejecutivos atractivos.

Power BI

Descripción general: Herramienta de Microsoft orientada a análisis empresarial.

Características: Integración con Excel, automatización, dashboards.

Ventajas: Económico, potente, accesible.

Desventajas: Algunas limitaciones gráficas comparado con Tableau.

Caso de uso: Empresas que trabajan con ecosistema Microsoft.

Google Data Studio (Looker Studio)

Ventajas: Gratis, fácil de usar, conexión con Google Analytics.

Desventajas: Menos variedad de gráficas.

Bibliotecas de Visualización

A continuación se presentan 4 bibliotecas.

Matplotlib (Python)

Filosofía: Base de la mayoría de visualizaciones en Python.

Características: Flexible, altamente personalizable.

Complejidad: Alta para principiantes.

Gráficas: Barras, líneas, dispersión, histogramas.

Ejemplo de código:

```
import matplotlib.pyplot as plt

plt.plot([1,2,3], [4,5,6])

plt.show()
```

Seaborn (Python)

Biblioteca basada en Matplotlib con diseño moderno.

Excelente para análisis estadístico.

```
import seaborn as sns  
  
sns.histplot(data)
```

Plotly (Python/Javascript)

Especializada en visualizaciones interactivas.

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

D3.js (JavaScript)

Biblioteca más poderosa para visualizaciones web.

Características: Total control del DOM, animaciones, interactividad.

Complejidad: Avanzada.

Conclusión

En conclusión, la visualización de datos constituye una herramienta esencial en el análisis moderno, ya que permite comunicar información de manera rápida y precisa. Elegir el tipo adecuado de gráfica facilita la identificación de patrones, tendencias, anomalías y comportamientos relevantes. Además, las herramientas y bibliotecas revisadas muestran que existe una amplia gama de opciones tanto para usuarios profesionales como para desarrolladores.

El *Storytelling* con datos reafirma la importancia de no solo mostrar gráficos, sino construir una historia clara y orientada a un mensaje específico. En lo personal, esta investigación permitió comprender la importancia de combinar técnica, diseño y narrativa para generar visualizaciones con alto impacto comunicativo.

8. Referencias (formato APA)

Cairo, A. (2016). *The Truthful Art: Data, Charts, and Maps for Communication*. New Riders.

Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs for Enlightening*. Analytics Press.

Microsoft. (2024). *Power BI Documentation*. <https://learn.microsoft.com/power-bi>

Tableau. (2024). *Tableau Online Help*. <https://help.tableau.com>

Python Software Foundation. (2024). *Matplotlib Documentation*. <https://matplotlib.org>