

Resumen Esquemático de la Teoría para el Examen de Visualización de Datos

Este informe presenta un resumen esquemático y detallado de los conceptos fundamentales en la visualización de datos, organizados para facilitar la comprensión y la recuperación rápida de información clave para fines de estudio. Se abordan las técnicas de procesamiento de datos, la visualización multivariable, los principios del color y la precisión visual, la honestidad en la representación, el uso de mapas, la distinción entre visualizaciones 2D y 3D, las directrices para la visualización científica y el arte del *storytelling* con datos, así como la interactividad y la experiencia de usuario.

I. Reducción de Dimensionalidad y Procesamiento de Datos

La reducción de dimensionalidad (DR) es un proceso esencial en la visualización de datos, que busca transformar conjuntos de datos de alta dimensión a un espacio de menor dimensión, conservando la mayoría de los conocimientos significativos de los datos originales.¹ El objetivo primordial de la DR es preservar la estructura significativa de un conjunto de datos, utilizando menos atributos para representar los elementos.¹ Esta estrategia es fundamental para gestionar la complejidad en la visualización, permitiendo la exploración de datos y el aprendizaje automático.¹ Al reducir las dimensiones a 2D o 3D, se facilita la visualización de las muestras, lo que a su vez permite detectar clústeres e identificar valores atípicos.¹

Las técnicas generales de DR se dividen en varias categorías: la **Selección de Características** (*Feature Selection*), que implica elegir un subconjunto de características relevantes para representar el conjunto de datos de manera óptima; la **Eliminación de Características** (*Feature Elimination*), que reduce el espacio de características descartando variables; y la **Ingeniería de Características** (*Feature Engineering*), que transforma los datos brutos en características más representativas.¹ Dentro de la selección de características, se incluyen métodos como el Missing Value Ratio, Low Variance Filter, High Correlation Filter, Random Forest, Backward Feature

Extraction y Forward Feature Selection.¹ Otras técnicas de DR se basan en componentes o factores (como Factor Analysis, Principal Component Analysis - PCA, e Independent Component Analysis) o en proyección (como ISOMAP, t-SNE y UMAP).¹

Un aspecto importante a considerar es la compensación inherente en la reducción de dimensionalidad para la visualización. Las técnicas de DR buscan preservar la información, pero la forma en que lo hacen varía. Por ejemplo, PCA prioriza la varianza global, mientras que t-SNE se enfoca en la estructura local. Esta diferencia implica que optimizar una de estas propiedades puede comprometer la otra. Si el objetivo es identificar patrones amplios y valores atípicos, PCA podría ser más adecuado. Sin embargo, para encontrar clústeres densos y relaciones de vecindad, t-SNE es superior, aunque con la advertencia de que las relaciones globales entre clústeres en t-SNE pueden carecer de significado. La elección de la técnica de DR, por lo tanto, no es arbitraria y debe alinearse con la pregunta analítica específica que se busca responder, lo cual impacta directamente en la honestidad y precisión de la visualización resultante.

Análisis de Correlación y Correlogramas

Cuando se trabaja con más de 3 o 4 variables cuantitativas, la Scatterplot Matrix (SPLOM) se vuelve difícil de manejar.¹ En estos escenarios, resulta más útil cuantificar la asociación entre pares de variables y visualizar estas cantidades en lugar de los datos brutos.¹ Para ello, se utiliza el

Coefficiente de Correlación (R), que se calcula mediante la fórmula $R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$.¹ El valor de R oscila entre -1 y 1, indicando el grado de correlación entre dos variables.¹ Un R igual a 0 significa que no hay asociación, mientras que un R de 1 o -1 indica una asociación perfecta.¹ Si R es mayor que 0, las variables están correlacionadas positivamente (valores mayores en una coinciden con mayores en la otra); si R es menor que 0, están anticorrelacionadas (valores mayores en una coinciden con menores en la otra).¹

Los **Correlogramas** son una herramienta visual para representar la estructura de correlación de las variables de entrada.¹ La intensidad del color y la forma de la elipse en un correlograma están directamente vinculadas al coeficiente de correlación.¹ Por convención, las correlaciones positivas ($R > 0$) se muestran en azul, y las anticorrelaciones ($R < 0$) se representan en rojo.¹ Un color más oscuro o intenso y una

elipse más estrecha indican una correlación más fuerte, ya sea positiva o negativa.¹

Varianza, Covarianza y Matriz de Covarianza

Antes de profundizar en las técnicas de reducción de dimensionalidad, es crucial comprender los conceptos de varianza y covarianza. La **Varianza** representa la variación de los valores en una única variable, indicando cuán dispersos están los valores entre sí.¹ Su fórmula es

$\text{var}(x) = \frac{1}{N-1} \sum (x_i - \bar{x})^2$.¹ A diferencia de la varianza, la

Covarianza se calcula entre dos variables diferentes y su propósito es indicar cómo estas dos variables varían juntas.¹ La fórmula de la covarianza es

$\text{cov}(x,y) = \frac{1}{N-1} \sum (x_i - \bar{x})(y_i - \bar{y})$.¹ La covarianza puede variar entre $-\infty$ y $+\infty$ y su signo indica la dirección de la relación lineal entre las variables.¹ Es importante destacar que la covarianza es cero para variables independientes.¹

La distinción entre covarianza y correlación es fundamental: la covarianza mide la extensión y dirección en que dos variables cambian en conjunto y es afectada por el cambio de escala, mientras que la correlación mide tanto la fuerza como la dirección de la relación lineal y no es influenciada por el cambio de escala, con un rango entre -1 y 1.¹

La **Matriz de Covarianza** se utiliza para representar los valores de covarianza de cada par de variables en datos multivariados.¹ En esta matriz simétrica, la diagonal muestra la varianza de cada variable (la covarianza entre la misma variable consigo misma).¹ Los valores dentro de la matriz de covarianza proporcionan información sobre la magnitud y dirección de la distribución de datos multivariados en un espacio multidimensional.¹ Una covarianza positiva indica que los valores de las variables tienden a aumentar o disminuir juntos; una covarianza negativa, que una variable tiende a aumentar mientras la otra disminuye; y una covarianza cero, que no hay relación lineal.¹

La comprensión de estos conceptos estadísticos es un puente fundamental hacia la codificación visual efectiva. La visualización de datos no es meramente una cuestión estética; está profundamente arraigada en la comprensión estadística. La forma en que se codifican visualmente las relaciones, como el color y la forma en los

correlogramas o los ejes en los gráficos PCA, son representaciones directas de estas propiedades estadísticas subyacentes. Una sólida comprensión de estos fundamentos estadísticos es, por lo tanto, crucial tanto para crear visualizaciones precisas como para interpretarlas correctamente, ya que una mala interpretación de la varianza o la correlación puede llevar a una lectura errónea de los patrones visuales, lo que subraya la importancia de la precisión visual.

Análisis de Componentes Principales (PCA)

El **Análisis de Componentes Principales (PCA)** es una técnica ampliamente utilizada para examinar datos de alta dimensión, operando bajo un paradigma de aprendizaje no supervisado.¹ Se recomienda usar PCA cuando se necesita reducir el número de variables sin poder identificar cuáles eliminar por completo, cuando se desea asegurar que las variables sean independientes entre sí, o cuando se está cómodo con que la variable independiente resultante sea menos interpretable.¹

El mecanismo de PCA consiste en introducir un nuevo conjunto de variables, denominadas **Componentes Principales (PCs)**, que son combinaciones lineales de las variables originales, estandarizadas a media cero y varianza unitaria.¹ Estos PCs se ordenan por varianza: el PC 1 representa la dirección de la mayor varianza de los datos, mientras que el PC 2 representa la mayor parte de la varianza restante ortogonal al PC 1. Este proceso se extiende para obtener el número de componentes necesarios para cubrir la cantidad de varianza deseada.¹

La relación de PCA con la matriz de covarianza es directa: los autovalores (*eigenvalues*) de la matriz de covarianza representan la magnitud de la dispersión en la dirección de los componentes principales, y los autovectores (*eigenvectors*) muestran la dirección.¹ En casos donde la covarianza es cero (variables independientes), los autovalores son iguales a los valores de varianza.¹

El proceso de PCA sigue una serie de pasos definidos:

1. Tomar la matriz de características $X \in \mathbb{R}^{N \times P}$, donde $N > P$.¹
2. Calcular el vector de medias para cada dimensión.¹
3. Calcular la matriz de covarianza.¹
4. Calcular los autovectores y autovalores correspondientes para cada dimensión.¹
5. Ordenar los autovectores por autovalores decrecientes y elegir P autovectores con los mayores autovalores para formar una nueva matriz $W \in \mathbb{R}^{P \times P}$.¹

6. Usar esta matriz de autovectores para transformar las muestras al nuevo subespacio: $Z=XW$.¹

Entre las fortalezas de PCA, se destaca que es un método muy interpretable, donde cada PC está bien definido y es ortogonal a las otras dimensiones. Además, permite obtener la varianza explicada por cada PC para seleccionar un número apropiado de dimensiones.¹ Sin embargo, PCA tiene debilidades importantes: tiende a ser altamente afectado por valores atípicos (

outliers) y funciona mejor solo con datos continuos.¹

Análisis Discriminante Lineal (LDA)

El **Análisis Discriminante Lineal (LDA)** es una técnica de aprendizaje supervisado cuyo propósito es separar o discriminar de la mejor manera posible las muestras en el conjunto de entrenamiento según su valor de clase.¹ La idea fundamental detrás de LDA, propuesta por Ronald Fisher, es encontrar un nuevo espacio de características para proyectar los datos de manera que se maximice la separabilidad de las clases.¹ Esto se logra maximizando la diferencia entre las medias de las clases (centroides de clase), normalizada por una medida de la variabilidad dentro de la clase, y minimizando la separación de las muestras dentro de cada clase.¹ LDA asume una distribución Gaussiana de los datos.¹

Los pasos para implementar LDA son los siguientes:

1. Calcular el vector de medias d-dimensional para las diferentes clases del conjunto de datos.¹
2. Calcular la matriz de dispersión (*Scatter matrix*), tanto entre clases como dentro de la clase.¹
3. Ordenar los autovectores por autovalores decrecientes y elegir k autovectores con los mayores autovalores para formar una matriz W de dimensión $d \times k$.¹
4. Usar esta matriz de autovectores para transformar las muestras al nuevo subespacio mediante la multiplicación de matrices: $Y=XW$.¹

Al comparar PCA y LDA, se observa que ambos son técnicas de transformación lineal comúnmente utilizadas para la reducción de dimensionalidad.¹ Sin embargo, PCA es un algoritmo no supervisado que busca los ejes de componentes ortogonales de máxima varianza en un conjunto de datos, mientras que LDA es un algoritmo

supervisado cuyo objetivo es encontrar el subespacio de características que optimiza la separabilidad de clases.¹

A pesar de sus ventajas, LDA presenta varias debilidades: no funciona bien si el diseño no está balanceado (es decir, si el número de objetos en las distintas clases es muy diferente).¹ También puede tener un rendimiento deficiente si la distribución de los datos es significativamente no Gaussiana.¹ Además, es sensible al sobreajuste (*overfit*) y no es aplicable (o es inferior) para problemas no lineales.¹

T-Distributed Stochastic Neighbor Embedding (t-SNE)

T-Distributed Stochastic Neighbor Embedding (t-SNE) es una técnica no lineal de reducción de dimensionalidad, utilizada principalmente para visualizar conjuntos de datos de alta dimensión.¹ Su mecanismo clave consiste en visualizar datos de alta dimensión asignando a cada punto de datos una ubicación en un mapa 2D o 3D, con el beneficio fundamental de intentar preservar la estructura local de los datos.¹ Esto significa que las distancias y agrupaciones de los puntos observadas en dimensiones bajas (como un gráfico de dispersión 2D) son lo más cercanas posible a las distancias en el espacio de alta dimensión, ya que t-SNE optimiza las distancias teniendo en cuenta la estructura local.¹

La principal diferencia entre t-SNE y PCA radica en que t-SNE intenta desconvolucionar las relaciones entre vecinos en datos de alta dimensión, siendo particularmente efectivo para conjuntos de datos altamente no lineales, donde PCA podría distorsionar las distancias (ejemplo del "swiss roll").¹ Mientras que PCA es una técnica matemática, t-SNE es probabilística.¹ Los algoritmos lineales de reducción de dimensionalidad como PCA se concentran en colocar puntos de datos disímiles lejos unos de otros en una representación de menor dimensión. En contraste, para representar datos de alta dimensión en un manifold no lineal de baja dimensión, es esencial que los puntos de datos similares se representen cerca, lo cual t-SNE logra y PCA no.¹

Sin embargo, t-SNE presenta varias debilidades: es computacionalmente costoso y puede tardar varias horas en conjuntos de datos de millones de muestras, a diferencia de PCA que se completa en segundos o minutos.¹ Además, debido a su naturaleza probabilística, diferentes ejecuciones con los mismos hiperparámetros pueden

producir resultados distintos, lo que requiere observar múltiples gráficos antes de realizar cualquier evaluación.¹ Otra limitación es que t-SNE no logra preservar la estructura global entre clústeres; las distancias de incrustación entre clústeres carecen de significado y la distribución global de los clústeres es aleatoria.¹ A pesar de esto, t-SNE está diseñado para capturar relaciones polinómicas complejas entre características, algo que PCA no puede hacer debido a su linealidad.¹

Visualización Tomográfica (Isocontornos, Superficies, Renderizado de Volumen)

La **visualización tomográfica** se refiere a las técnicas utilizadas para medir y segmentar estructuras internas de objetos 3D, como se ve en el análisis de estructuras craneofaciales en embriones de ratón.¹ Dentro de este campo, los

isocontornos o contornos de mapas tomográficos son líneas que conectan puntos de igual valor. Cuando las líneas de contorno están muy separadas, indican una pendiente más suave, mientras que si están muy juntas, muestran una pendiente más pronunciada.¹ Las elevaciones de puntos, por su parte, representan las alturas entre estas líneas de contorno.¹

Existen métodos que reducen la dimensionalidad de los datos de 3D a 2D para su visualización. Un **gráfico de superficie** (*surface plot*) muestra el límite del conjunto de datos, pero no revela su parte interior.¹ Un

plano de corte (*slice plane*) exhibe una sección 2D en una coordenada específica, como una misma coordenada Y.¹ Por otro lado, una

isosuperficie (*isosurface*) representa un valor escalar específico, ignorando todos los puntos de volumen que no cumplen con ese valor.¹

El **renderizado de volumen** (*volume rendering*) es una técnica cuyo objetivo principal es crear una imagen 2D que refleje, en cada píxel, los datos escalares dentro de un conjunto de datos 3D dado.¹ El desafío principal en este proceso radica en la elección de la función de mapeo que transforma un conjunto completo de valores escalares (para los vóxeles a lo largo del rayo de visión) a un solo píxel en la imagen 2D resultante.¹

Finalmente, la **Tomografía Computarizada (CT)** utiliza rayos X y un modelo matemático, la Transformada de Radon, para reconstruir una imagen 3D de un objeto

a partir de sus proyecciones.¹ El objetivo de la CT es recuperar una imagen detallada del cuerpo (valores de densidad) a partir de los datos de rayos X sobre un número finito de líneas.¹

La representación de información 3D en visualizaciones 2D presenta un desafío inherente. Dado que muchas de estas técnicas implican una reducción de la dimensionalidad de 3D a 2D, se produce una pérdida o transformación de información. Esta limitación subraya la dificultad de transmitir eficazmente estructuras 3D complejas en un medio 2D, como una pantalla o papel. La elección de la técnica de visualización en este contexto se convierte en una decisión sobre qué aspecto de los datos 3D es más crítico comunicar en 2D, sentando las bases para una discusión más profunda sobre las ventajas y desventajas de las visualizaciones 2D frente a las 3D.

Tabla 1: Comparativa de Técnicas de Reducción de Dimensionalidad

Característica	PCA (Análisis de Componentes Principales)	LDA (Análisis Discriminante Lineal)	t-SNE (t-Distributed Stochastic Neighbor Embedding)
Propósito	Reducir variables, encontrar ejes de máxima varianza.	Separar (discriminar) clases, maximizar separabilidad entre grupos.	Visualizar conjuntos de datos de alta dimensión, preservar estructura local.
Linealidad	Lineal	Lineal	No Lineal
Supervisión	No Supervisado	Supervisado (requiere etiquetas de clase)	No Supervisado
Objetivo Principal	Maximizar la varianza global de los datos.	Optimizar la separabilidad de clases.	Preservar las relaciones de vecindad (estructura local).
Fortalezas	Muy interpretable, PCs ortogonales, varianza explicada.	Maximiza la diferencia entre medias de clases, minimiza variabilidad intra-clase.	Efectivo para identificar clústeres locales, captura relaciones complejas.
Debilidades	Sensible a outliers,	Sensible a datos no	Computacionalmente

	funciona mejor con datos continuos.	balanceados, distribuciones no Gaussianas, <i>overfit</i> , no apto para problemas no lineales.	costoso, resultados variables entre ejecuciones, no preserva estructura global entre clústeres.
Costo Computacional	Bajo	Moderado	Alto
Naturaleza	Matemática	Matemática	Probabilística
Consistencia	Consistente (mismos resultados con mismos datos/parámetros)	Consistente (mismos resultados con mismos datos/parámetros)	Puede variar entre ejecuciones (requiere múltiples observaciones)

II. Visualización Multivariable y Avanzada

Los conjuntos de datos grandes presentan un desafío fundamental en la visualización: contienen mucha más información de la que puede mostrarse eficazmente en un solo gráfico.¹ Para abordar esta complejidad, se recurre a soluciones como la agrupación de variables en un único panel o, para conjuntos de datos más complejos, la creación de figuras con múltiples paneles, donde cada uno muestra un subconjunto específico de los datos.¹ Esta necesidad de manejar la "maldición de la dimensionalidad" se extiende más allá del procesamiento de datos y afecta directamente el diseño de la visualización. Incluso después de la reducción de dimensionalidad, el gran número de variables restantes o la complejidad de sus relaciones sigue siendo un desafío significativo para la visualización directa, imponiendo una carga cognitiva considerable al observador. Esto impulsa el desarrollo de técnicas de visualización multivariable especializadas que abstractan u organizan los datos de manera más digerible, incluso si se pierde algún detalle. La visualización, en este sentido, se convierte en un acto de cuidadosa selección y presentación, no solo de descarga de datos brutos.

Visualización de Distribuciones Múltiples

En escenarios donde se desea visualizar múltiples distribuciones simultáneamente, es útil conceptualizar los datos en términos de una **variable de respuesta** (la variable cuyas distribuciones se quieren mostrar, como lifeExp o el value de métricas de Iris) y **variables de agrupación** (que definen subconjuntos de datos con distribuciones distintas de la variable de respuesta, como continent/year o species/petal/sepal).¹

Para este propósito, se emplean diversas técnicas de visualización ¹:

- **Boxplots:** Muestran la distribución de la variable de respuesta para cada grupo.¹
- **Violin Plots:** Similares a los boxplots, pero exhiben la densidad de la distribución.
- **Strip Charts:** Muestran cada punto de datos individualmente.
- **Stacked Histograms:** Histogramas apilados para comparar distribuciones.
- **Overlapping Densities:** Presentan densidades superpuestas para comparar distribuciones.¹
- **Ridgeline Plot:** Muestran múltiples distribuciones de densidad en un formato apilado.
- **Sina Plots:** Una variación de los strip charts que ajusta los puntos para visualizar la densidad.
- Las distribuciones también pueden organizarse a lo largo de los ejes vertical u horizontal para facilitar la comparación.¹

Visualización de Proporciones Múltiples

El desafío de visualizar proporciones se acentúa cuando estas se especifican según múltiples variables de agrupación.¹ Para ello, existen varias técnicas de visualización ¹:

- **Multiple Pie Charts:** Tienden a ser ineficientes en espacio y a menudo oscurecen las relaciones entre las proporciones.¹
- **Grouped Bars:** Funcionan bien siempre que el número de condiciones comparadas sea moderado.¹
- **Stacked Bars:** Pueden ser eficaces para un gran número de condiciones.¹ Permiten ajustes de posición como dodge (lado a lado), fill (apilar y normalizar altura), jitter (añadir ruido para evitar superposición), nudge (mover etiquetas) y stack (apilar).¹
- **Stacked Densities:** Son apropiadas cuando las proporciones cambian a lo largo de una variable continua.¹ El apilamiento desglosa los datos por más de una

variable categórica que componen el todo ¹, como se ilustra con el conjunto de datos

diamonds usando `geom_density(position="fill")`.¹ El parámetro `adjust` en `geom_density` permite ajustar el ancho de banda.¹

- **Mosaic Plots:** Asumen que cada nivel de una variable de agrupación puede combinarse con cada nivel de otra.¹ Son similares a los gráficos de barras apiladas, pero tanto las alturas como los anchos de las áreas sombreadas varían, siendo proporcionales al número de casos de cada combinación posible de variables categóricas.¹ Se construyen subdividiendo sucesivamente los ejes X e Y.¹
- **Treemaps:** Funcionan bien incluso si las subdivisiones de un grupo son completamente distintas de las subdivisiones de otro, a diferencia de los mosaic plots.¹ Anidan recursivamente rectángulos, donde el área de cada uno es proporcional a la cantidad de casos.¹ Requieren una variable numérica continua para el área (`value`), una para el color de relleno (`parent`), una para la etiqueta (`id`) y el grupo padre (`parent`).¹ Su implementación en R se realiza con el paquete `treemapify`, utilizando funciones como `geom_treemap()`, `geom_treemap_subgroup_border()`, `geom_treemap_subgroup_text()` y `geom_treemap_text()`.¹
- **Parallel Sets (PSP):** Son más efectivos que los mosaic plots o treemaps cuando hay más de dos variables de agrupación.¹ Muestran cómo el conjunto de datos total se desglosa por cada variable categórica individual, utilizando bandas sombreadas para ilustrar las relaciones entre los subgrupos.¹ No son adecuados para datos numéricos (para los cuales se utilizan los PCP) y las barras horizontales en los PSP muestran la frecuencia absoluta de cada categoría.¹

Visualización de Relaciones y Correlaciones

Para entender cómo se relacionan dos o más variables cuantitativas ¹, se emplean técnicas como la

Scatterplot Matrix (SPLOM) y el **Bubble Plot**. La SPLOM utiliza múltiples diagramas de dispersión organizados en una matriz para determinar y visualizar la correlación entre una serie de variables.¹ Permite identificar relaciones por pares, su naturaleza, la presencia de valores atípicos y el agrupamiento por grupos.¹ Sin embargo, se vuelve difícil de manejar con más de 3 o 4 variables cuantitativas.¹ El Bubble Plot, por su parte, utiliza círculos u otras formas para representar datos, donde el tamaño y el

color del punto pueden codificar variables adicionales simultáneamente.¹

Figuras Multi-Panel (Small Multiples, Figuras Compuestas)

Las figuras multi-panel se clasifican en dos categorías principales ¹:

- **Small Multiples** (también conocidos como "trellis plot" o "faceting"): Son gráficos que consisten en múltiples paneles dispuestos en una cuadrícula regular. Cada panel muestra un subconjunto diferente de los datos, pero todos los paneles utilizan el mismo tipo de visualización.¹ La idea clave es segmentar los datos según una o más dimensiones, visualizar cada segmento por separado y organizar las visualizaciones en una cuadrícula, etiquetando columnas, filas o paneles individuales con los valores de las dimensiones de los datos.¹ Es fundamental evitar el uso de diferentes escalas en los ejes en paneles separados, o al menos llamar la atención del lector sobre ello si es necesario, y organizar los paneles en un orden lógico y significativo.¹
- **Compound Figures**: Combinan varios paneles independientes ensamblados en una disposición arbitraria. Estos paneles pueden mostrar visualizaciones completamente diferentes o incluso conjuntos de datos distintos, con el objetivo de transmitir un punto general unificador.¹ Para asegurar la claridad, es crucial mantener la consistencia entre colores, símbolos y fuentes, y los ejes deben estar alineados para evitar malas interpretaciones.¹

Gráficos de Coordenadas Paralelas (PCP)

Los **Gráficos de Coordenadas Paralelas (PCP)** son una técnica de visualización utilizada para analizar datos numéricos multivariados, permitiendo trazar elementos de datos individuales a través de muchas dimensiones.¹ A diferencia del plano cartesiano clásico, cada variable numérica tiene su propio eje, dispuesto en paralelo, verticalmente y equidistante.¹ Cada elemento de datos se representa mediante segmentos de línea conectados, derivados de un conjunto conectado de puntos, uno en cada eje.¹

La fortaleza de los PCP reside en su capacidad para comparar variables que pueden ser completamente diferentes, con rangos e incluso unidades distintas.¹ En general,

muchas líneas paralelas indican una relación positiva, mientras que muchas líneas que se cruzan (formas de X) sugieren una asociación negativa.¹ Sin embargo, los PCP suelen ser difíciles de entender para audiencias no técnicas debido al típico "embrollo de líneas".¹ No se recomienda mostrar demasiadas variables numéricas a la vez, y los PCP no son adecuados para datos categóricos; en su lugar, se deben usar los Parallel Sets Plots (PSPs), donde las barras horizontales muestran la frecuencia absoluta de cada categoría.¹

Para mejorar la legibilidad de los PCP, se aplican transformaciones típicas ¹:

- **Escalado (Scaling):** Transforma los datos brutos a una nueva escala común, crucial para comparar variables con diferentes unidades.¹
- **Orden de los Ejes (Axis Order):** Optimizar el orden de los ejes verticales puede disminuir el desorden (*clutter*) al minimizar el número de cruces entre series.¹
- **Resaltado (Highlighting):** Resaltar una muestra o grupo de interés específico puede evitar la superposición excesiva de líneas.¹

La importancia de seleccionar la técnica de visualización adecuada según el tipo de variable y la pregunta analítica es un principio rector. La efectividad de una visualización depende en gran medida de su alineación con el tipo de datos (numéricos, categóricos, proporcionales) y la tarea analítica específica (comparar distribuciones, mostrar partes de un todo, identificar relaciones). Utilizar un tipo de gráfico incorrecto, como un PCP para datos categóricos, puede llevar a una interpretación errónea o a una falta de eficiencia. Esto refuerza el concepto de "precisión visual", sugiriendo que la visualización de datos es una ciencia de decisiones deliberadas para comunicar el mensaje correcto.

Tabla 2: Tipos de Paletas de Color y sus Aplicaciones

Tipo de Paleta de Color	Tipo de Datos Ideal	Propósito/Aplicación	Características Clave
Categórica/Cualitativa	Datos nominales (categorías sin orden)	Diferenciar categorías discretas.	Grandes diferencias en tono (hue), saturación y/o luminosidad para facilitar la distinción. Evitar gradientes continuos que sugieran orden. Máximo 12 colores.

Cuantitativa Secuencial	Datos ordinales o cuantitativos que progresan de bajo a alto.	Mostrar una progresión ordenada de valores.	Variación consistente en luminosidad y/o saturación de un mismo tono o tonos adyacentes. Puede ser continua o discretizada.
Cuantitativa Divergente	Datos cuantitativos con un punto central significativo (ej. cero, promedio).	Resaltar desviaciones positivas y negativas desde un valor central.	El color central (normalmente blanco o neutro) debe representar el valor central. Los colores divergen hacia tonos distintos en cada extremo.
Agrupada	Múltiples categorías principales con subcategorías.	Organizar visualmente datos jerárquicos o agrupados.	Utiliza diferentes tonos para las categorías principales y variaciones en saturación y luminosidad dentro de cada tono para las subcategorías.
Mapas Coropléticos Bivariados	Dos variables cuantitativas superpuestas espacialmente.	Mostrar la interacción o relación entre dos distribuciones espaciales.	Esquema de color con dos escalas superpuestas en dos direcciones, a menudo formando una matriz (ej. 3x3) de combinaciones de color.

III. Principios de Color y Precisión Visual

La elección del color en la visualización de datos va más allá de la estética; es un componente crítico para la integridad y accesibilidad de la información. La forma en que los humanos percibimos el color difiere fundamentalmente de cómo las computadoras lo procesan.¹ Los humanos tenemos tres tipos de conos que

reaccionan a diferentes espectros de luz, y nuestra percepción se basa en la luminosidad, el tono y la saturación.¹ En contraste, las computadoras procesan bandas de frecuencia muy estrechas de rojo, verde y azul, y calculan la luz de manera lineal.¹ La Teoría del Proceso Oponente de Ewald Hering explica cómo el cerebro combina las señales por sustracción para crear canales de diferencia de color (rojo-verde, amarillo-azul) y un canal de luminancia (blanco y negro).¹ Esta teoría también explica por qué ciertos colores, como el amarillo, pueden percibirse con mayor intensidad debido a la sobreexcitación de los receptores.¹ Además, nuestra percepción es inherentemente relativa; los juicios de color no son absolutos y el contexto que rodea a los elementos visuales puede modular la discriminabilidad y la precisión.¹

Para garantizar la precisión y la coherencia, se recomienda utilizar **modelos de color perceptualmente correctos** como HSV (Hue, Saturation, Value), HCL (Hue, Chroma, Luminance) y CIELab (L^* , a^* , b^*).¹ Estos modelos son preferibles porque se acercan más a la forma en que el ojo humano percibe el color, y se deben usar escalas de color con interpolación no lineal para evitar distorsiones.¹

Canales de Color y su Efectividad

El color es uno de los canales visuales más potentes para codificar datos.¹ Se clasifica en:

- **Canales de Magnitud** (para atributos ordenados o datos cuantitativos): Incluyen la posición en una escala común (la más efectiva), la posición en una escala no alineada, la longitud, la inclinación/ángulo, el área, la profundidad, la luminancia de color, la saturación de color, la curvatura y el volumen (la menos efectiva).¹
- **Canales de Identidad** (para atributos categóricos): Incluyen la región espacial, el tono de color (hue), el movimiento y la forma.¹

La luminancia y la saturación son canales de magnitud, mientras que el tono es un canal de identidad.¹ Estos tienen poca o ninguna interferencia con los canales de posición.¹ La jerarquía de los canales visuales implica que el color debe usarse juiciosamente, complementando en lugar de reemplazar codificaciones más efectivas. Aunque el color es un canal poderoso, no es el más efectivo para comparaciones cuantitativas precisas. Una excesiva dependencia del color para juicios cuantitativos exactos, cuando la posición o la longitud serían más claras, constituye una deficiencia

en el diseño. Por lo tanto, el color debe emplearse para apoyar la codificación principal, por ejemplo, para resaltar categorías o añadir una capa adicional de información, en lugar de ser el único medio para transmitir diferencias cuantitativas precisas, especialmente en datos ordenados.

Problemas de Escalas Arcoíris

El uso de una paleta de color incorrecta, como las escalas arcoíris, puede mostrar efectos en la visualización que no están presentes en los datos, lo que representa un riesgo significativo de desinformación.¹ Las escalas arcoíris no tienen en cuenta cómo percibimos el color y carecen de un orden inherente, lo que las hace inadecuadas para representar datos ordenados.¹ Su variación no lineal rompe la linealidad de la escala debido a la sobreexcitación de receptores en ciertos puntos, lo que puede generar la aparición de límites o "bandas" perceptuales que no existen en la realidad.¹ Esto dificulta la distinción de regiones y lleva a una pérdida de detalle.¹ Por todas estas razones, se recomienda encarecidamente evitar el uso de la escala arcoíris.¹

Propiedades del Canal de Color (Contraste, Luminosidad, Saturación, Tono)

Las propiedades del canal de color son cruciales para una visualización efectiva. El **tono** (*hue*), la **saturación** y la **luminancia/claridad/brillo** son componentes fundamentales.¹ El

contraste facilita la detección: un mayor contraste hace que los elementos sean más fáciles de percibir, mientras que un menor contraste los dificulta.¹ Un contraste homogéneo en el tono, pero con variación en el tono y la luminosidad, facilita la identificación.¹ Es importante tener en cuenta el

contraste simultáneo, donde el contexto que rodea a los elementos modula la percepción del color.¹

Paletas y Esquemas de Color (Categoricos, Secuenciales, Divergentes,

Agrupados)

Una directriz general para la selección de colores es evitar intentar seleccionar colores individualmente; en su lugar, se recomienda utilizar paquetes de R como grDevices, colorRamps, RcolorBrewer y colorspace.¹ El paquete

colorspace, basado en los modelos HCL y HSV, es particularmente recomendado por su atractivo estético y sus opciones de personalización.¹

Los tipos de paletas y esquemas de color se eligen en función del tipo de datos ¹:

- **Categórica/Cualitativa:** Se utiliza para diferenciar categorías discretas.¹ Es importante evitar colores que formen un gradiente continuo, ya que podrían confundirse con valores ordinales.¹ Se prefieren grandes diferencias en tono, saturación y/o luminosidad, y se recomienda no usar más de 12 colores.¹
- **Cuantitativo Secuencial:** Ideal para datos que progresan de bajo a alto, pudiendo ser continuos o discretizados (ordinales).¹
- **Cuantitativo Divergente:** Diseñado para datos con un punto central significativo (por ejemplo, el cero).¹ El color central (normalmente blanco) debe representar este valor central, y los colores divergen hacia tonos distintos en cada extremo. Puede ser continuo o discretizado.¹
- **Agrupado:** Un esquema de color para múltiples categorías principales (diferentes tonos) con pasos de saturación y luminosidad dentro de cada tono.¹
- **Mapas Coropléticos Bivariados:** Utilizan un esquema de color con dos escalas superpuestas en dos direcciones, a menudo resultando en una matriz de 3x3.¹

Las escalas perceptualmente lineales son aquellas que varían de forma consistente a través del rango de valores.¹ Se recomienda usar escalas perceptualmente correctas, como las de la familia Viridis (Viridis, Magma, Plasma, Inferno, Cividis, Mako, Rocket, Turbo).¹

La elección del color no es meramente estética; es un componente crítico para la integridad y la accesibilidad de los datos. Si el color se elige incorrectamente, puede distorsionar activamente los datos, creando patrones falsos u ocultando los reales, y además puede excluir a segmentos de la audiencia. Esto eleva la selección del color de una decisión de diseño a una cuestión de integridad de los datos, lo que implica una responsabilidad ética para garantizar la accesibilidad.

Diseño para Deficiencias de Color

Es fundamental diseñar visualizaciones que sean accesibles para personas con daltonismo.¹ Para ello, se pueden utilizar herramientas y simuladores como

Colorblindness para R, Coblis (www.color-blindness.com) y Color Brewer 2.0 (colorbrewer2.org).¹

Semántica del Color

La asignación de color debe basarse en su significado inherente para evitar confusiones y mejorar la comprensión. Por ejemplo, se pueden usar paletas cálidas para altas temperaturas, una escala azul-roja para temperaturas negativas a positivas, o el verde para datos de vegetación.¹ Es crucial evitar usos engañosos, como emplear una escala continua como si fuera categórica, usar más de 12 categorías en una misma paleta, o presentar rankings sin un orden claro.¹ Una buena práctica es seguir el principio de "No data, no color", es decir, no asignar color a regiones que carecen de datos.¹

IV. Honestidad y Precisión en la Visualización

La precisión visual es un pilar fundamental en la comunicación de datos. Las representaciones visuales, incluso si se basan en datos precisos, pueden inherentemente distorsionar la verdad si se ignoran los principios de diseño. Las decisiones de diseño en sí mismas conllevan mensajes implícitos y pueden alterar activamente la realidad. Incluso con datos correctos, una visualización mal diseñada puede "mentir". Esto establece un vínculo causal directo entre el diseño y la interpretación, lo que subraya la dimensión ética de la visualización de datos.

Manipulación de Ejes (Escala, Línea Base Cero, Etiquetas, Ejes Duales)

La manipulación de los ejes es una fuente común de deshonestidad visual. La elección de la escala puede enfatizar la variación o la tendencia de manera engañosa.¹ Un

eje Y truncado, que no comienza en cero o por debajo del valor mínimo significativo, puede magnificar pequeños cambios, haciéndolos parecer más dramáticos de lo que realmente son.¹ La directriz es mostrar el eje completo para no magnificar descensos o ascensos.¹

Para las **gráficas de barras**, que codifican el **tamaño**, es **imprescindible** que tengan una **línea base cero** para representar con precisión las cantidades.¹ No hacerlo es incorrecto y puede generar una falsa impresión de las magnitudes relativas.¹ En contraste, los gráficos de líneas, que codifican la

posición de los puntos, no necesariamente necesitan una línea base cero.¹ Las

etiquetas en los ejes son **imprescindibles** para la claridad y comprensión de los datos.¹

Se recomienda **evitar los ejes duales** ya que son problemáticos y pueden llevar fácilmente a conclusiones equivocadas.¹ Esto se debe a que sugieren una correlación implícita entre las líneas que puede no existir, y solo son aceptables en casos muy específicos.¹

Agregaciones Equivalentes y Datos Faltantes

Para asegurar la precisión, las **agregaciones** deben ser **equivalentes** al comparar datos, evitando mensajes engañosos.¹ Por ejemplo, al comparar inversiones de la industria antes y después de regulaciones, los periodos de tiempo deben ser comparables.¹

En cuanto a los **datos faltantes**, es crucial **marcarlos explícitamente** en la visualización.¹ Esto implica mantener el espacio para las marcas ausentes o usar marcas distintas para indicar su ausencia, en lugar de simplemente omitirlas, lo que podría generar una falsa sensación de continuidad o completitud.¹

Precisión en Gráficas de Área y Barras Apiladas

Las **barras apiladas no alineadas** son problemáticas porque dificultan la comparación de porcentajes y la identificación de puntos de referencia como el 50%.¹ Se recomienda alinearlas a una base común o usar otras representaciones.

Las **gráficas de área** (*area charts*) también pueden ser engañosas.¹ Es muy difícil identificar qué categoría es responsable de un aumento, y los picos en las capas superiores tienden a magnificarse, distorsionando la percepción del cambio.¹ En muchos casos, se recomienda no usar gráficos de área y, en su lugar, calcular índices o utilizar gráficos de líneas separados para cada categoría.¹

Un aspecto importante a considerar es que las configuraciones predeterminadas del software de visualización pueden ser una fuente de deshonestidad. Muchos programas pueden establecer ejes automáticamente o usar gráficos de área sin el contexto adecuado, lo que puede llevar a representaciones engañosas. Confiar únicamente en estas configuraciones sin un pensamiento crítico sobre el diseño puede resultar en visualizaciones que, sin intención maliciosa, distorsionan la verdad. Por lo tanto, es fundamental que los usuarios evalúen críticamente la salida visual de sus herramientas y anulen las configuraciones predeterminadas cuando sea necesario para garantizar la precisión y la honestidad.

Selección de Métricas Relevantes (Tasas vs. Totales)

Al visualizar datos, es esencial elegir **métricas que sean verdaderamente relevantes** para el mensaje.¹ Un error común es mostrar totales (por ejemplo, el número total de asesinatos en ciudades) en un mapa geográfico, lo que a menudo se correlaciona con el tamaño de la población y puede llevar a la conclusión errónea de que las ciudades más grandes son intrínsecamente más peligrosas.¹ La directriz es usar

tasas (por ejemplo, la tasa de asesinatos por cada 100,000 personas) en lugar de totales para comparaciones significativas entre entidades de diferentes tamaños.¹ Los mapas de perfil geográfico que son esencialmente solo mapas de población se

consideran una "molestia" común (*pet peeve*).¹

Representación de Incertidumbre y Error (Alternativas a Boxplots)

La representación de la incertidumbre y el error es crucial para una visualización honesta. Los **Box and Whisker Plots (Boxplots)**, aunque útiles para mostrar la distribución, tienen limitaciones¹: la caja se percibe como una unidad, no muestran toda la distribución, y la asociación de tamaño a cantidad puede ser engañosa (una caja pequeña puede representar una cantidad grande).¹ Además, pueden hacer que grupos distintos parezcan casi idénticos.¹

Como alternativas a los boxplots, se sugieren¹:

- **Violin Plots:** Muestran la densidad de la distribución.
- **Gradient Plots.**
- **Bean Plots:** Similares a los violin plots pero muestran los puntos individuales.
- **Beeswarm Plots:** Muestran cada punto de datos individualmente, evitando la superposición.

La inclusión de **estimaciones de incertidumbre o barras de error** es una práctica fundamental que mejora la toma de decisiones y proporciona un mejor reflejo de la comprensión científica.¹ Para los gráficos de líneas, se recomiendan las

bandas de error (Error Bands) y los **fancharts** como alternativas para representar la incertidumbre.¹

Tabla 3: Errores Comunes en Precisión Visual y Cómo Evitarlos

Error Común / Práctica Engañosa	Por qué es Engañoso	Práctica Correcta / Alternativa
Eje Y Truncado (no empieza en cero o valor mínimo)	Magnifica pequeños cambios, distorsiona la magnitud de las variaciones.	Mostrar el eje completo , incluyendo el cero o el rango completo de datos.
Línea Base No Cero para	Las barras codifican tamaño;	Las barras deben tener una

Barras	sin línea base cero, la magnitud relativa es incorrecta.	línea base cero. (Los gráficos de líneas, que codifican posición, no la requieren necesariamente).
Ejes Duales	Sugieren una correlación implícita que puede no existir; son difíciles de interpretar.	Evitar ejes duales en la mayoría de los casos. Si es indispensable, usar con extrema precaución y explicaciones claras.
Barras Apiladas No Alineadas	Dificulta la comparación de porcentajes y la identificación de proporciones (ej. 50%).	Alinear las barras apiladas a una base común o usar alternativas que faciliten la comparación de proporciones.
Magnificación en Gráficos de Área	Los picos en capas superiores se magnifican, haciendo difícil identificar la contribución real de cada categoría.	No usar gráficos de área en muchos casos. Considerar gráficos de líneas separados o índices para cada categoría.
Totales en Mapas Geográficos (en lugar de tasas)	Los mapas de totales a menudo reflejan la densidad de población, no la tasa real del fenómeno, llevando a conclusiones erróneas.	Usar tasas (ej. por 100,000 habitantes) para comparaciones significativas entre entidades de diferente tamaño.
Boxplots sin contexto completo	No muestran toda la distribución, pueden hacer que grupos parezcan idénticos o que el tamaño de la caja se asocie erróneamente a la cantidad.	Usar Violin Plots, Bean Plots, Beeswarm Plots o incluir la distribución completa para una representación más precisa.
Ausencia de Etiquetas en Ejes	Impide la comprensión clara de los datos y sus unidades.	Las etiquetas en los ejes son imprescindibles.
Datos Faltantes no indicados	Crea una falsa sensación de continuidad o completitud en los datos.	Marcar explícitamente dónde faltan datos, manteniendo el espacio o usando marcas distintas.

V. Visualización con Mapas

Los mapas son herramientas poderosas para la visualización de datos geoespaciales.¹ Permiten que audiencias no técnicas comprendan y analicen datos de manera más sencilla.¹ Sin embargo, crear un buen mapa no es una tarea sencilla, ya que las personas a menudo recuerdan la apariencia del mapa más que los datos que representa.¹ Antes de diseñar un mapa, es crucial responder a dos preguntas fundamentales: ¿Cuál es el propósito del mapa? y ¿Cuál es el costo/impacto esperado del resultado?.¹

La saliencia visual del área geográfica en un mapa puede inducir a error en la interpretación si no se normaliza mediante métricas relevantes. Por ejemplo, los mapas coropléticos, aunque útiles para mostrar patrones, pueden ser engañosos si el tamaño de una región no se correlaciona con los datos que se le atribuyen. Esto se debe a una tendencia cognitiva común: las áreas más grandes atraen naturalmente más atención y se perciben implícitamente como "más" de lo que se está mapeando, incluso si los datos subyacentes (como la densidad o la tasa) son menores. Esto establece un vínculo causal directo entre la codificación visual (tamaño del área) y la posible mala interpretación. Para evitar esto, es crucial distinguir cuándo un mapa coroplético es apropiado (por ejemplo, para patrones de votación política donde el área es relevante) frente a cuándo se necesita un mapa de símbolos proporcionales (para recuentos brutos) o un mapa basado en tasas (para comparaciones normalizadas).

Fundamentos y Tipos de Mapas

Existen diversos tipos de mapas para diferentes propósitos ¹:

- **Mapa de Puntos** (*Point Map*): Es la representación más simple, colocando un punto en cada ubicación correspondiente a la variable.¹ Son útiles para mostrar patrones de distribución y densidad, pero requieren datos de ubicación precisos (geocodificación) y los puntos pueden superponerse según el nivel de zoom.¹
- **Mapas de Símbolos Proporcionales** (*Proportional Symbol Maps / Bubble Maps*): Utilizan círculos u otras formas cuyo tamaño y color representan variables de datos adicionales.¹ También pueden sufrir de superposición y requieren geocodificación precisa.¹
- **Mapas de Clúster** (*Cluster Maps*): Similares a los mapas de símbolos

proporcionales, pero al hacer zoom, pueden revelar puntos más pequeños representados por los puntos más grandes.¹

- **Diagramas de Voronoi** (*Voronoi Diagrams*): Se emplean para encontrar la distancia mínima necesaria para alcanzar un punto o hito, ayudando a comprender la proximidad y las características de distancia.¹ Son útiles, por ejemplo, para elegir una nueva ubicación para una escuela lejos de las existentes.¹
- **Mapas Coropléticos** (*Choropleth Maps*): Dividen un área en secciones (por ejemplo, límites políticos o geográficos) y rellenan cada sección con un tono o sombra diferente para representar una variable o un rango de valores.¹ Son excelentes para patrones y tendencias generales, pero se debe tener precaución al usarlos con áreas que difieren mucho en tamaño, ya que el tamaño de una región puede no tener relación con los datos.¹
- **Mapa de Puntos** (*Dot Map*): Cada punto representa una instancia o cantidad específica.¹
- **Mapa de Isopleas** (*Isopleth Map*): Muestra datos continuos utilizando líneas que conectan puntos de igual valor.¹
- **Mapas Esquemáticos** (*Schematic Maps*): Se utilizan para simplificar navegaciones y sistemas complejos, como metros o redes eléctricas.¹
- Existen muchos otros tipos de mapas, como los mapas de fondo con R y ggplot2, mapas de burbujas interactivos, cartogramas (donde la geometría de las regiones se distorsiona para representar una variable), mapas hexbin (regiones representadas como hexágonos o densidad 2D) y mapas de conexión (que muestran rutas entre ubicaciones).¹

Consideraciones para Mapas Interactivos

Las APIs de Google Maps (Google Earth, Google Maps Images, Google Places) ofrecen herramientas para construir programas de mapeo visual interactivos.¹ Los mapas son herramientas poderosas para el

storytelling, como se observa en ejemplos como "The desirability of Boston streets" o "Pinellas County is the worst place in Florida to be black and go to public school".¹ Sin embargo, su impacto puede verse sesgado por el hecho de que el usuario a menudo se enfoca en la apariencia del mapa en sí, más que en los datos que representa.¹ Esta dualidad sugiere que, si bien los mapas son atractivos y pueden transmitir narrativas potentes, su fuerte atractivo visual puede, en ocasiones, eclipsar la información

cuantitativa que se pretende comunicar. El medio puede convertirse en el mensaje, desviando potencialmente la atención de los conocimientos numéricos subyacentes. Por ello, las visualizaciones basadas en mapas requieren un diseño cuidadoso para garantizar que los datos sigan siendo lo más importante. Los diseñadores deben equilibrar el atractivo estético con la claridad de los datos, posiblemente utilizando elementos interactivos o anotaciones para guiar la atención del usuario hacia la información cuantitativa.

VI. Visualización 2D vs. 3D

La elección entre visualizaciones 2D y 3D es una decisión crítica que debe basarse en la tarea del espectador y las capacidades perceptuales humanas, no en la mera disponibilidad de herramientas 3D.

Desafíos Perceptuales de 3D (Profundidad, Oclusión, Distorsión)

La intuición común de que experimentamos el mundo en 3D es engañosa; en realidad, los humanos vemos en **2.05D**, y la inexactitud de nuestros juicios de profundidad no es sorprendente.¹ Solo podemos percibir la capa exterior de un mundo 3D, viendo un único punto a lo largo de cada rayo.¹

La **oclusión** es un desafío inherente a las visualizaciones 3D, donde partes de la escena pueden quedar ocultas por otras.¹ Aunque la navegación interactiva puede ayudar a resolver la estructura 3D de las partes ocluidas, esto requiere tiempo e impone una carga cognitiva considerable al usuario.¹

La **distorsión de perspectiva** es otro fenómeno problemático: los objetos distantes parecen más pequeños y cambian su posición planar en el plano de la imagen.¹ Esta distorsión, junto con la oclusión, hace que los gráficos de barras 3D sean intrínsecamente más difíciles de interpretar que sus equivalentes 2D.¹

El sistema perceptual humano está fundamentalmente mal equipado para realizar juicios cuantitativos precisos en 3D sobre una pantalla 2D. La proyección de datos 3D en una pantalla 2D para análisis cuantitativos introduce sesgos perceptuales y

precisiones inherentes que son difíciles de superar. El cerebro lucha por comparar con precisión longitudes, áreas o posiciones en un espacio 3D distorsionado. Esto constituye un fuerte argumento contra el uso gratuito de 3D para datos que son inherentemente 2D o que podrían representarse eficazmente en 2D. Subraya que el 3D debe reservarse únicamente cuando la estructura geométrica es la principal información que se necesita.

Criterios para Elegir entre 2D y 3D

La decisión de usar 3D debe ser deliberada:

- **Cuándo usar 3D:** Solo cuando la tarea del espectador requiere fundamentalmente comprender la **estructura geométrica tridimensional** de objetos o escenas.¹ Ejemplos incluyen animaciones faciales en tiempo real, donde la combinación de datos 2D y 3D puede reducir el costo computacional.¹ En estos casos, una vista 3D con **controles de navegación interactivos** (para ajustar el punto de vista) permite a los usuarios construir un modelo mental útil de la estructura del conjunto de datos más rápidamente.¹
- **Cuándo NO usar 3D** (para datos 2D): Un gráfico 3D que muestra datos 2D no añade valor y, de hecho, añade ruido (*clutter*), haciendo el gráfico más difícil de leer.¹ Si no es necesario para la comprensión de la estructura 3D, no se debe usar 3D.¹

La navegación interactiva es una solución necesaria, pero insuficiente, para los problemas inherentes de la visualización 3D. Aunque la interactividad puede mitigar algunos de los desafíos de la visualización 3D, como la oclusión, no los elimina por completo. Al contrario, desplaza la carga de la interpretación al usuario, quien debe participar activamente para resolver las ambigüedades visuales. Esto sugiere que, si bien la interactividad permite la exploración, conlleva un costo en términos de carga cognitiva para el usuario. Una visualización 3D verdaderamente efectiva minimiza la necesidad de una interacción extensa para resolver ambigüedades fundamentales, lo que implica que la interactividad en 3D es una herramienta para la exploración, no una solución para las deficiencias perceptuales inherentes.

Se deben aplicar **métricas de calidad** generales, aplicables tanto a 2D como a 3D ¹:

- Usar **tonos de color suaves**, excepto donde se quiera llamar la atención, ya que

los colores brillantes pueden ser agotadores.

- Asegurar la **accesibilidad al color** utilizando herramientas como Coblis para verificar la compatibilidad con el daltonismo.
- **Pivotar el gráfico** (ej., usar gráficos de barras horizontales) cuando las etiquetas de categoría son largas para facilitar la lectura.
- Elegir un **ancho de barra apropiado**: barras más delgadas con suficiente espacio en blanco para que sean ligeramente más gruesas que el espacio entre ellas, evitando que estén demasiado juntas (el cerebro podría intentar evaluar área vs. longitud).
- **Eliminar el encabezado del eje X y añadir etiquetas de datos** si estas últimas ya proporcionan la información necesaria, haciendo el encabezado redundante.

VII. Visualización Científica y Métricas de Calidad

La visualización científica es una forma especializada de comunicación que prioriza la transferencia de información sobre el embellecimiento estético. Las decisiones de diseño se inclinan fuertemente hacia la claridad, la precisión y la eficiencia. Cualquier elemento de diseño que no contribuya directamente a la comprensión se considera "ruido" o "clutter". Esto significa que el contexto de la visualización es primordial; una visualización para una publicación científica tendrá prioridades de diseño diferentes a las de una revista de estilo de vida, lo que subraya la importancia de la audiencia y el propósito en las decisiones de diseño.

Directrices Clave para Publicaciones Científicas

El objetivo principal de cualquier gráfico en el contexto de publicaciones científicas y presentaciones es **transmitir información de manera efectiva**.¹ Para ello, se proponen diez directrices clave ¹:

1. **Simplicidad**: Crear el gráfico más simple que transmita la información deseada.¹
2. **Codificación**: Considerar el tipo de objeto y atributo de codificación utilizados.¹
3. **Foco**: Centrarse en visualizar patrones o detalles según el propósito del gráfico.¹
4. **Rangos de Ejes**: Seleccionar rangos de ejes significativos.¹
5. **Tasas de Cambio**: Utilizar transformaciones de datos y relaciones de aspecto del

gráfico para enfatizar las tasas de cambio en series de tiempo.¹

6. **Puntos Superpuestos:** Trazar puntos superpuestos de manera que las diferencias de densidad sean aparentes en gráficos de dispersión (esta directriz también aplica a mapas).¹
7. **Series de Tiempo:** Usar líneas al conectar datos secuenciales en gráficos de series de tiempo.¹
8. **Agregación:** Agregar conjuntos de datos grandes de manera significativa.¹
9. **Comparación de Variables:** Mantener los rangos de los ejes lo más similares posible al comparar variables.¹
10. **Esquema de Color:** Seleccionar un esquema de color apropiado basado en el tipo de datos (en referencia a la clase de color).¹

Además, la inclusión de **estimaciones de incertidumbre o barras de error** es una práctica fundamental que mejora la toma de decisiones y proporciona un mejor reflejo de la comprensión científica.¹

Reducción de Ruido Visual (Clutter)

El *clutter* se define como una colección desordenada de entidades gráficas que no son necesarias para comprender la información.¹ Su impacto es significativo, ya que el

clutter oscurece la estructura de los datos y dificulta la identificación de patrones, relaciones y estructuras.¹

Para reducir el *clutter*, se pueden aplicar varios métodos:

- **Orden de Dimensiones:** Variar el orden de las dimensiones puede reducir el *clutter* sin disminuir el contenido de información.¹
- **Eliminación de Ítems Comunes de Clutter:** Esto incluye efectos 3D en datos 2D, líneas de cuadrícula oscuras (se deben usar líneas grises suaves o eliminarlas), el uso excesivo de colores brillantes o negritas, y el uso innecesario de texto en mayúsculas.¹

Proceso de Conceptualizar, Crear, Criticar y Cortar (4 Cs)

El proceso de las "4 C" (Conceptualizar, Crear, Criticar y Cortar) representa un enfoque cíclico e iterativo para el diseño de visualizaciones, enfatizando el refinamiento y la centralidad en el usuario.

- **Conceptualizar:** Implica crear un borrador (*rough plot*) de la infografía, definir un título pegadizo y descriptivo, establecer un punto de partida y un punto final (conclusión o llamada a la acción), y definir la ruta visual que se desea que el lector siga (hipótesis).¹
- **Crear:** Consiste en elaborar la historia de los datos, incluyendo información interesante y sorprendente, y personalizando la presentación.¹
- **Criticar:** Se centra en la consistencia de la historia de los datos (tiempo verbal, terminología, diseño) y en pedir a otra persona que revise la visualización antes de publicarla.¹
- **Cortar (Cut):** Implica eliminar información superflua o innecesaria, manteniendo solo lo esencial e interesante. El objetivo es no abrumar al lector con demasiado texto o imágenes, y considerar si la información extra podría ir en otra infografía o si una parte se comunicaría mejor como imagen, reorganizando para mejorar el flujo.¹ Se proporcionan ejemplos que comparan infografías de cerveza, mostrando una con mucho *clutter* frente a una clara y concisa.¹

Este proceso es un flujo de trabajo práctico que enseña que el diseño de visualizaciones efectivas es un proceso de mejora continua y refinamiento, donde la retroalimentación y la simplificación son clave. Esto también se conecta con el tema de la usabilidad al enfatizar las pruebas y la iteración.

VIII. Storytelling en Visualización de Datos

El *storytelling* en la visualización de datos es una elección estratégica para superar la carga cognitiva y la apatía, transformando el consumo pasivo de datos en un compromiso activo. La narración estructurada de una serie de hechos y acciones, llevadas a cabo por uno o más personajes, es una técnica de comunicación que crea un vínculo emocional con la audiencia y ayuda a que el mensaje destaque.¹ Aplicado a la visualización de datos, el

storytelling maximiza el impacto en la comunicación de datos.¹ Su propósito principal es atraer y mantener la atención de la audiencia, transmitir un mensaje de forma

memorable a través de la emoción, dirigir el interés y conectar con la audiencia, y comunicar una idea envuelta en emoción, como lo describe Robert McKee.¹

La información en bruto, incluso si está bien visualizada, puede ser abrumadora o poco atractiva para muchas audiencias. El *storytelling* actúa como un puente cognitivo, proporcionando estructura y ganchos emocionales para hacer que la información compleja sea digerible y relevante. Aborda la tendencia humana inherente a buscar narrativas y significado, lo que significa que la capacidad de elaborar una narrativa convincente en torno a los datos es una habilidad de alto nivel que mejora el impacto práctico de las visualizaciones, especialmente en contextos profesionales donde las partes interesadas necesitan información clara y procesable.

Estructuras Narrativas (Viaje del Héroe, Tres Actos)

Las historias poseen características universales.¹ El

Viaje del Héroe, popularizado por Joseph Campbell como el "monomito", describe una estructura fundamental que comparten los mitos: el héroe inicia su aventura desde el mundo ordinario hacia una región de prodigios, se enfrenta a fuerzas fabulosas, obtiene una victoria decisiva y regresa con dones para su comunidad.¹ Esta estructura ha tenido una gran influencia y ha dado lugar a variaciones como el "Story Circle" de Dan Harmon.¹

Otra estructura narrativa fundamental es la **Estructura Clásica de Tres Actos** de Aristóteles¹:

1. **Planteamiento (Introducción)**: Se presenta al personaje y su mundo, y un incidente incitador conduce al conflicto.
2. **Nudo (Conflicto)**: Se desarrollan los esfuerzos para resolver el conflicto, culminando en un clímax.
3. **Desenlace (Resolución)**: Se presenta el resultado del conflicto.

Las transiciones entre estos actos están marcadas por situaciones de conflicto que generan tensión, un elemento clave para mantener la atención.¹ Esta estructura sirve como un modelo general para la comunicación.¹

Tipos de Narrativa (Anotación, Narración, Historia)

No todas las visualizaciones requieren ser una historia completa; existen diversas estrategias para comunicar eficazmente con datos.¹ Para guiar al usuario, se pueden emplear las siguientes estrategias:

1. **Anotación:** Consiste en texto y marcas que ayudan a interpretar la visualización.¹
2. **Narración:** Implica anotaciones conectadas en una secuencia lógica.¹ Puede ser **cerrada** (una estructura secuencial con un mensaje definido) o **abierta** (una visualización exploratoria que sirve como herramienta para buscar narrativas).¹
3. **Historia:** Una narración que incorpora drama o conflicto para provocar una respuesta emocional.¹

La necesidad de contar una historia se acentúa cuando la audiencia no tiene un contexto previo sobre los datos o la visualización, cuando se requiere guiarla para evitar conclusiones erróneas, o cuando no existe un interés previo en el tema, el cual puede generarse a través del *storytelling*.¹

El elemento de "conflicto", tradicionalmente central en el *storytelling*, se traduce en la visualización de datos como la puesta de relieve de discrepancias o problemas, lo que impulsa el interés del usuario. En la narración de datos, el "conflicto" no es necesariamente un drama humano, sino una discrepancia, un problema o un patrón inesperado dentro de los datos. Resaltar estos "conflictos", como la inexactitud de un modelo, un valor atípico o una tendencia divergente, crea una tensión narrativa que impulsa a la audiencia a buscar una resolución o comprensión. Esto proporciona una aplicación concreta de los principios abstractos del *storytelling* a los datos. Los diseñadores pueden identificar "conflictos" en sus conjuntos de datos y utilizarlos como motores narrativos, haciendo que sus visualizaciones sean más atractivas e impactantes al enmarcarlas como problemas que deben entenderse o resolverse.

Planificación de la Historia (Storyboard)

El **storyboard** es el primer paso crucial en la planificación de una historia de datos.¹ Permite organizar las ideas en secciones y subsecciones, reduciendo cada parte a su mínima expresión (una acción, una frase).¹ Esto facilita visualizar el flujo de la historia y realizar cambios. Puede ser multinivel, permitiendo desarrollar el esquema de la

historia en subsecciones.¹ Las herramientas utilizadas son similares a las de planificación de cualquier narrativa, como guiones o tableros.¹ Un ejemplo práctico de aplicación se encuentra en la visualización de datos de una granja eólica, donde se ilustra el proceso desde los prototipos iniciales hasta la imagen final, destacando los desafíos y las soluciones para comunicar la precisión del modelo.¹

IX. Interactividad y Animación

La interactividad transforma al observador en un usuario, otorgándole un papel más prominente y proactivo en la percepción de la información, con menos dependencia de la guía del visualizador.¹ Esta capacidad expande los límites físicos de lo que se puede mostrar en un espacio determinado, abarcando una mayor variedad de análisis para satisfacer diversas necesidades de información.¹ La interactividad facilita la manipulación de datos para responder a diferentes preguntas, amplifica el control y la personalización del observador, y amplía el interés a diversos grupos de usuarios, especialmente a las nuevas generaciones, acostumbradas a un consumo de información más activo.¹

Antes de incorporar técnicas interactivas, es fundamental estudiar las necesidades, los medios disponibles y las opciones utilizables.¹ Los conceptos clave en interactividad incluyen el

Evento (una acción del usuario, como un clic o una pulsación de tecla), el **Control** (la reacción a un evento, como mostrar un pop-up o resaltar una opción) y la **Acción** (la operación a realizar en la aplicación, como filtrar o resaltar información).¹ Los tipos de controles abarcan desde contenedores (Frames, Panels) hasta componentes específicos como botones, casillas de verificación, etiquetas, campos de texto, deslizadores, menús desplegables, barras de búsqueda y selectores de fecha.¹

Factores que Influyen en la Incorporación de Interactividad

Varios factores determinan la incorporación de técnicas interactivas ¹:

- **Restricciones:** La tecnología disponible y las competencias del observador o la

audiencia.¹

- **Objetivos:** La experiencia de usuario deseada y las opciones interactivas disponibles.¹
- **Representación de Datos:** Qué gráficos podrían beneficiarse de la interactividad para facilitar la interpretación.¹
- **Diseño Fiable:** Asegurar que el sistema funcione como se espera y que la integridad de los datos se mantenga.¹
- **Diseño Accesible:** Minimizar las pulsaciones de teclas o clics del ratón para una interactividad discreta y útil.¹
- **Diseño Elegante:** Garantizar una interacción fluida y mejorar la jugabilidad.¹

Categorías de Interacción en Visualización de Datos

Las interacciones se clasifican en varias categorías ¹:

- **Filtrado:** Permite a los usuarios especificar qué datos desean ver o ocultar.¹ Ejemplos incluyen la selección de deportes en resultados olímpicos, diferencias de género en doctorados, especies de árboles en Nueva York y análisis de mercado con deslizadores.¹
- **Selección:** Resalta valores de interés sin eliminarlos, a menudo modificando atributos de color o reordenando datos.¹ Ejemplos son el resaltado de premios Nobel por país, nombres de bebés populares a lo largo de los años, reformas de compensación laboral por estado, histogramas del censo por nación (brushing), muertes por armas en EE. UU. y resultados bancarios (ordenación por columna).¹
- **Participación/Colaboración:** Permite a los usuarios contribuir con sus propios datos para personalizar la representación visual.¹ Ejemplos incluyen "¿Cuántos años tienes?" (comparando la edad del usuario con la de celebridades), "¿Qué tan bien conoces tu área?" (un cuestionario sobre el vecindario), "¿Cuánto azúcar en la comida?" (un cuestionario sobre el contenido de azúcar) y "¿Recuerdas dónde estaba dividida Alemania?" (trazando la frontera).¹
- **Resumen:** Proporciona diferentes niveles de anotaciones de datos bajo demanda del usuario, permitiendo al usuario controlar la percepción de los datos.¹ Ejemplos incluyen un mapa de calor de cruces de personajes de los Vengadores, la retórica de los presidentes de EE. UU. y visualizaciones de Bloomberg con pop-ups de "cómo leer este gráfico".¹
- **Observación y Exploración:** Muestra múltiples vistas o diferentes niveles de detalle de forma dinámica, especialmente cuando no toda la información puede

mostrarse simultáneamente.¹ Ejemplos incluyen "Earth" (patrones de viento, clima, océano con zoom/pan), "Racial Dot Map" (densidad de población con zoom geométrico), "2020 Census Demographic data Map Viewer" (datos demográficos con zoom/pan) y "OECD Better Life Index" (drill-down en países).¹ Esta categoría también incluye la exploración lineal a través de secuencias narrativas, como "Killing the Colorado" (un viaje paso a paso de la degradación del río) y "What's Really Warming the World?" (una secuencia de causas del cambio climático).¹

Animación

La **animación** muestra datos con una dimensión temporal de forma dinámica, permitiendo apreciar patrones que no se pueden ver en imágenes estáticas.¹ Es una secuencia de imágenes con cambios entre imágenes consecutivas a lo largo del espacio y el tiempo, creando la ilusión de movimiento en el cerebro.¹ Aunque genera un fuerte impacto visual en la visualización de datos, no todos los usuarios la aprecian.¹

El propósito de la animación en la visualización de datos es múltiple ¹:

- **Incrementar la Dimensionalidad:** Añadiendo el tiempo como una dimensión a los datos.
- **Proporcionar Información Diferente:** Ofreciendo conocimientos no disponibles en gráficos estáticos.
- **Calidad Estética:** Mejorando el atractivo visual.
- **Conocimiento:** Proporcionando información contextual fuera del área de atención específica.
- **Transición:** Guiando al usuario entre diferentes gráficos.
- **Descripción Funcional:** Relacionando el comportamiento que representa un objeto animado (por ejemplo, el desplazamiento).
- **Énfasis:** Llamando la atención sobre un elemento visible o un proceso específico.
- **Expresión:** Enriqueciendo la experiencia del usuario con la aplicación o tarea.
- **Cambio:** Mostrando cómo un objeto o proceso evoluciona con el tiempo.
- **Visualización Directa:** Estableciendo correspondencias entre atributos de movimiento (fase, frecuencia) y variables del conjunto de datos.
- **Asociación:** Utilizando secuencias de movimiento para establecer relaciones entre grupos de información.

Al considerar la animación, es importante tener en cuenta la duración entre los fotogramas: una animación demasiado lenta puede resultar aburrida, mientras que una demasiado rápida puede provocar la pérdida de detalles.¹ La animación también puede distraer del mensaje.¹ Es crucial que el usuario tenga control para iniciar y detener la animación, a través de eventos y controles como el inicio automático al cargar la página, botones de reproducción/pausa/parada/repetición y deslizadores para controlar los fotogramas.¹ Ejemplos de animación incluyen "Breathing Earth" (simulando zonas de vegetación global), jugadores de la NFL (altura y peso a lo largo del tiempo) y la taquilla global por género (carrera de gráficos de barras animados).¹

Consejos Generales para Interactividad y Animación

La gestión del tiempo es crucial al implementar soluciones interactivas.¹ No se deben invertir tiempo y recursos en técnicas interactivas o innovadoras si no aportan valor a la comprensión.¹ Es fundamental centrarse en lo importante y relevante, cuestionando siempre si la interactividad es necesaria y aporta valor al proyecto.¹

X. Usabilidad y Experiencia de Usuario (UX)

La evaluación de los sistemas de visualización de datos es fundamental, especialmente cuando son interactivos.¹ Los términos

Usabilidad y Experiencia de Usuario (UX) se emplean para evaluar estos sistemas, y aunque a menudo se confunden, UX abarca la usabilidad como uno de sus componentes.¹

Usabilidad

La **Usabilidad** se define como un atributo de calidad que evalúa la facilidad de uso de aplicaciones, sitios web, interfaces y sistemas por parte del usuario.¹ Incluye cinco

componentes clave de calidad ¹:

- **Aprendizaje** (*Learnability*): Facilidad con la que los usuarios aprenden las tareas.
- **Eficiencia**: Rapidez con la que los usuarios realizan las tareas.
- **Memorabilidad**: Facilidad para recordar y reproducir tareas después de un tiempo sin usar el sistema.
- **Errores**: Número y gravedad de los errores, y capacidad de recuperación del sistema.
- **Satisfacción**: Grado de satisfacción general con el uso del sistema.

Es importante diferenciar entre **Utilidad**, **Usabilidad** y **Útil** ¹:

- **Utilidad**: Se refiere a la funcionalidad, es decir, si el sistema hace lo que los usuarios esperan.
- **Usabilidad**: Se centra en cuán fácil y satisfactorio es el uso del sistema.
- **Útil**: Es una combinación de Utilidad y Usabilidad.

La usabilidad es crucial porque es una condición necesaria para que los usuarios adopten y continúen utilizando un sistema.¹ Los usuarios tienden a abandonar sistemas complicados, que fallan con frecuencia, no permiten realizar acciones deseadas, causan desorientación o tienen información de ayuda poco clara.¹ La usabilidad está estrechamente ligada a la productividad.¹

En el contexto de la visualización de datos, la "**Usabilidad de Datos**" se refiere a la calidad de los datos y la información, asociada a tres principios ¹:

- **Fiabilidad de Datos**: Calidad del procesamiento de datos, asegurando confianza en los datos con sus intervalos de error.
- **Estabilidad de Datos**: Minimizar el impacto negativo de transformaciones o procesamientos, como la reducción de dimensionalidad.
- **Soporte a la Toma de Decisiones**: La representación debe ser lo suficientemente clara para que los usuarios tomen decisiones informadas.

Para evaluar la usabilidad, el enfoque principal es el usuario, mediante observación y cuestionarios.¹ Dos cuestionarios estándar son:

- **SUS (System Usability Scale)**: Una herramienta rápida para medir la usabilidad de software, sitios web y aplicaciones. Consta de 10 preguntas con una escala Likert de 5 puntos. Es simple, fácil de entender y proporciona resultados fiables, distinguiendo entre sistemas útiles y no útiles. Evalúa la efectividad (si el usuario puede realizar acciones) y la eficiencia (el esfuerzo del usuario).¹ La puntuación se calcula como

$SUS = (X + Y) * 2.5$, donde X es la suma de las puntuaciones de preguntas impares - 5, e Y es 25 - la suma de las puntuaciones de preguntas pares. Una puntuación inferior a 51 indica posibles problemas.¹

- **PSSUQ (Post-Study System Usability Questionnaire):** Ampliamente utilizado para medir la satisfacción del usuario. La versión 3.0 tiene 16 preguntas con una escala Likert de 7 puntos, proporcionando indicadores de satisfacción general, utilidad del sistema (SYSUSE), calidad de la información (INFOQUAL) y calidad de la interfaz (INTERQUAL).¹ La elección entre SUS y PSSUQ depende de los aspectos a evaluar y la fatiga del evaluador.¹

Experiencia de Usuario (UX)

La **Experiencia de Usuario (UX)** estudia la experiencia global de un usuario al interactuar con un sistema.¹ Abarca indicadores objetivos, estéticos y emocionales, incluyendo usabilidad, interfaz de usuario, experiencia de interacción, diseño interactivo y experiencia del cliente.¹ Las definiciones de UX enfatizan que es una consecuencia del estado interno del usuario (predisposiciones, expectativas, necesidades, motivación, estado de ánimo), las características del sistema (complejidad, propósito, usabilidad, funcionalidad) y el contexto de interacción (entorno organizacional/social, significado de la actividad, voluntariedad de uso).¹

Los métodos de evaluación de UX incluyen ¹:

- **Evaluaciones de Expertos:** Realizadas por especialistas.
- **Estudios de Laboratorio:** Similares a las pruebas de usabilidad, útiles en la fase de prototipado.
- **Estudios de Campo:** Los participantes utilizan el sistema en su entorno natural.
- **Cuestionarios:** Los usuarios responden preguntas para formar una opinión.
- **Métodos Mixtos:** Combinación de diferentes métodos para obtener indicadores más completos.
- **Medidas Psicofisiológicas:** Medidas objetivas como la frecuencia cardíaca, la transpiración y los movimientos de los músculos faciales para evaluar emociones positivas o negativas, idealmente con dispositivos no invasivos.¹

Modelos de UX

Los modelos de evaluación de UX se basan en dos aspectos principales: la usabilidad y la experiencia/emotividad.¹ Elementos externos son cruciales para entender la UX como una experiencia subjetiva y dinámica ¹:

- **Usuario:** Sus competencias (novato, experimentado, experto) y características emocionales (estatus social, objetivos personales, necesidades afectivas) pueden influir en la UX.¹
- **Contexto:** El entorno físico (iluminación, conexión a internet) y el contexto sociocultural (valores, actitudes, experiencia previa) desempeñan un papel.¹
- **Aspectos Temporales:** La calidad pragmática (usabilidad) tiende a aumentar con el tiempo, mientras que la calidad hedónica (emotividad) puede disminuir a medida que la novedad desaparece.¹ La evaluación de la UX se realiza típicamente cuando un sistema está terminado, con menos estudios a largo plazo.¹

Los modelos de UX más utilizados incluyen ¹:

- **Hassenzahl UX:** Basado en tres factores (Usuario, Sistema, Contexto), distingue entre atributos pragmáticos (utilidad y funcionalidad) y hedónicos (estética, comportamiento, funcionalidad intuitiva, estimulación, identificación, evocación).¹ El cuestionario Attrakdiff, con 28 criterios, se utiliza para este modelo.¹
- **User Experience Questionnaire (UEQ):** Evalúa productos interactivos con 26 indicadores bipolares en 7 escalas, agrupados en 6 categorías: atractivo, perspicuidad, eficiencia, fiabilidad, estimulación y novedad.¹
- **meCUE:** Una evaluación modular de los componentes clave de la UX, que busca crear una medida estándar. Se basa en el modelo CUE, que define características instrumentales y no instrumentales, similares a los atributos pragmáticos y hedónicos, y que influyen en las respuestas emocionales del usuario.¹ El cuestionario meCUE tiene tres módulos validados (percepciones del producto, emociones y consecuencias de la UX) y es flexible, permitiendo seleccionar módulos según lo que se necesite medir.¹

Modelos de UX para Visualización de Datos

Para los sistemas de visualización de datos, es crucial distinguir entre características instrumentales/pragmáticas y no instrumentales/hedónicas ¹:

- **Instrumentales:** Incluyen utilidad (satisfacer necesidades, datos útiles), completitud (información correcta y suficiente), percepción (información comprensible con mínimo esfuerzo), confianza (información válida y precisa) e intuición (información intuitiva y fácil de entender).¹
- **No Instrumentales:** Abarcan la estética (visualmente agradable) y el atractivo (que atraiga al usuario).¹

El peso de la evaluación entre características instrumentales y no instrumentales depende del uso del sistema.¹ En sistemas basados en acciones, como los de visualización de datos, las cualidades instrumentales son más importantes, ya que el objetivo es optimizar el rendimiento de la tarea.¹

El modelo y cuestionario **meCUE** se considera el más adecuado para evaluar sistemas de visualización de datos ¹ porque:

- Otorga un peso suficiente a las características no instrumentales, al tiempo que enfatiza la usabilidad.
- Su cuestionario es flexible.
- Valora los aspectos visuales y estéticos.
- El modelo y el cuestionario funcionan para diferentes tipos de aplicaciones.

Conclusiones

La visualización de datos es un campo multidisciplinar que exige una comprensión profunda tanto de los principios estadísticos y computacionales como de la percepción humana y el diseño. La reducción de dimensionalidad es un paso fundamental para manejar la complejidad de los datos, pero su elección debe ser estratégica, considerando si se prioriza la estructura global o local de los datos. La precisión visual es un pilar inquebrantable, ya que incluso las visualizaciones basadas en datos correctos pueden distorsionar la verdad si los principios de diseño, como el uso adecuado de los ejes, las líneas base y las paletas de color, no se respetan. La manipulación sutil de los elementos visuales puede llevar a interpretaciones erróneas, lo que subraya la responsabilidad ética del diseñador.

El color, aunque un canal visual potente, debe usarse con juicio, complementando codificaciones más efectivas y considerando siempre la accesibilidad para personas con deficiencias de color. Los mapas, herramientas poderosas para el *storytelling*

geoespacial, requieren una cuidadosa normalización de los datos para evitar que la saliencia visual del área geográfica induzca a error. La visualización 3D, a pesar de su atractivo, presenta desafíos perceptuales inherentes que limitan su precisión para juicios cuantitativos en pantallas 2D, por lo que su uso debe reservarse para cuando la comprensión de la estructura geométrica es esencial.

El *storytelling* emerge como una estrategia vital para transformar el consumo pasivo de datos en un compromiso activo, superando la carga cognitiva y la apatía de la audiencia. Al enmarcar los datos en narrativas estructuradas que incluyan elementos de "conflicto" (discrepancias o problemas en los datos), se genera tensión y se impulsa el interés del usuario. La interactividad y la animación, si bien amplían las posibilidades de exploración y muestran el dinamismo temporal, deben incorporarse con un propósito claro, evitando la distracción y priorizando la comprensión del usuario. Finalmente, la usabilidad y la experiencia de usuario (UX) son métricas críticas para evaluar la eficacia de una visualización. Modelos como meCUE permiten una evaluación holística, equilibrando la funcionalidad instrumental con la estética y la emotividad, asegurando que la visualización no solo sea precisa, sino también atractiva y fácil de usar.

Para la preparación del examen, se recomienda un enfoque esquemático y basado en palabras clave. El dominio de las definiciones, fórmulas y comparativas entre técnicas (ej. PCA vs. LDA vs. t-SNE) es crucial. La capacidad de identificar y corregir errores comunes de precisión visual (ejes truncados, líneas base, uso de totales vs. tasas) será fundamental. Comprender el propósito y las limitaciones de cada tipo de visualización (incluidos los mapas y las representaciones 2D/3D) y la aplicación adecuada de los principios del color garantizará una respuesta precisa. Finalmente, la integración del *storytelling* y la interactividad como herramientas para mejorar la comunicación de datos, en lugar de meros adornos, completará la comprensión experta de la materia.

Obras citadas

1. teoria-all.pdf