

Rainbows Revisited: Modelado de un diseño de mapa de colores eficaz para Inferencia gráfica

Khairi Reda y Danielle Albers Szafir

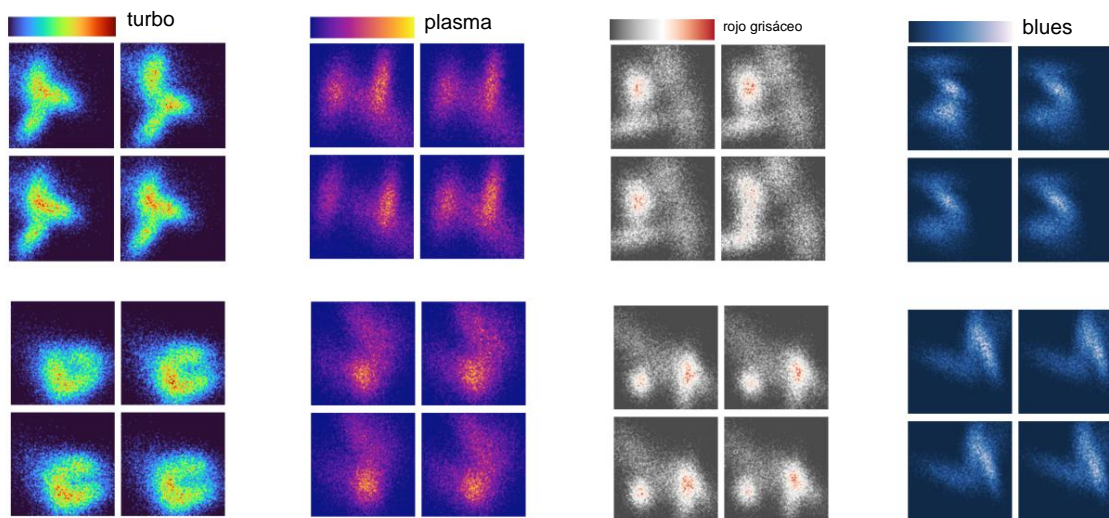


Fig. 1. Ocho estímulos de ejemplo del Experimento 1. Un solo estímulo consiste en una alineación de cuatro campos escalares codificados por colores que se muestran en una cuadrícula de 2x2. Para cada alineación, ¿cuál de las cuatro tramas se destaca como diferente? Las respuestas se encuentran en la Sección 10. Esta prueba de inferencia gráfica nos permite determinar el poder de discriminación de los diseños de mapas de colores de la competencia. Nuestros resultados dan lugar a un nuevo modelo para predecir la utilidad de un mapa de colores, particularmente para tareas que involucran juicio e inferencia basados en modelos.

Resumen: el mapeo de colores es una técnica fundamental para visualizar datos escalares. La literatura anterior ofrece pautas para el diseño efectivo de mapas de colores, como enfatizar la variación de luminancia y limitar los cambios en el tono. Sin embargo, los estudios empíricos del color se centran en gran medida en las tareas de percepción. Este enfoque limitado inhibe nuestra comprensión de cuán generalizables son estas pautas, particularmente para tareas como la inferencia visual que requiere síntesis y juicio a través de múltiples percepciones. Además, el énfasis en los diseños de rampa tradicionales (p. ej., secuenciales o divergentes) puede dejar de lado otras métricas clave o estrategias de diseño. Estudiamos cómo una métrica cognitiva, la variación del nombre del color, afecta la capacidad de las personas para hacer juicios basados en modelos. En dos experimentos de inferencia gráfica, los participantes vieron una serie de campos escalares codificados por colores muestreados de diferentes modelos y evaluaron las relaciones entre estos modelos. Al contrario de las pautas convencionales, los participantes fueron más precisos al ver mapas de colores que cruzan una variedad de colores identificables de forma única.

Modelamos el rendimiento de los participantes utilizando esta métrica y descubrimos que proporciona un mejor ajuste a los datos experimentales que los principios de diseño existentes. Nuestros hallazgos indican ventajas cognitivas para mapas coloridos como el arcoíris, que exhiben una alta categorización de colores, a pesar de sus propiedades perceptivas tradicionalmente indeseables. Tampoco encontramos evidencia de que la categorización del color llevaría a los observadores a inferir características de datos falsos. Nuestros resultados proporcionan métricas empíricamente fundamentadas para predecir el rendimiento de un mapa de colores y sugieren pautas alternativas para diseñar nuevos mapas de colores cuantitativos para respaldar la inferencia. Los datos y materiales para este documento están disponibles en: <https://osf.io/tck2r/>

Términos de **índice**: color, percepción, inferencia gráfica, datos escalares

1 INTRODUCCIÓN

El color es uno de los canales más utilizados en visualización. Los diseñadores utilizan el color para codificar una variedad de información cuantitativa, como la velocidad del viento de un huracán o el nivel de ruido creado por una simulación de motor a reacción. En estas y otras visualizaciones similares, el color no solo comunica valores de datos individuales, sino que también ayuda a transmitir

formas y patrones [59]. Luego, un observador puede estudiar estos patrones para inferir algo sobre el proceso físico subyacente o el modelo que generó los datos.

El trabajo anterior ofrece una variedad de pautas para el diseño efectivo de mapas de colores (ver Bujack et al. [11] o Zhou & Hansen [68] para encuestas). Por ejemplo, los expertos abogan por secuencias de colores que aumenten gradualmente en luminancia para variables continuas [43, 51]. La idea detrás de esta recomendación es que al controlar cuidadosamente la luminancia, ayudamos a establecer la ordinalidad. Además, la luminancia tiene una mayor capacidad para transmitir detalles espaciales sutiles en comparación con el tono o el croma [14, 61]. Los expertos han desaconsejado durante mucho tiempo el uso del arco iris y otras secuencias espectrales, que tienden a variar predominantemente en tonalidad [5, 35, 45]. Los mapas de colores del arco iris en particular se han señalado como un ejemplo de diseño de visualización ineficaz, o incluso engañoso [5, 45].

Las recomendaciones anteriores se han incorporado al canon de visualización de datos [36]. Sin embargo, los estudios empíricos de conducción de color

• Khairi Reda trabaja en la Universidad de Indiana-Universidad de Purdue en Indianápolis.
Correo electrónico: redak@iu.edu.

• Danielle Albers Szafir está en la Universidad de Colorado Boulder. Correo electrónico:
danielle.szafir@colorado.edu.

Manuscrito recibido xx xxx. 201x; aceptado xx xxx. 201x. Fecha de publicación xx xxx. 201x; fecha de la versión actual xx xxx. 201x. Para obtener información sobre cómo obtener reimpresiones de este artículo, envíe un correo electrónico a: reprints@ieee.org.
Identificador de objeto digital: [xx.xxxx/TVCG.201x.xxxxxx](https://doi.org/10.1109/TVCG.201x.xxxxxx)

estas pautas han enfatizado predominantemente las tareas de percepción, como estimar valores en ubicaciones específicas del mapa [41, 59] o comparar la distancia entre parches de color [32]. Aunque informativos, estos estudios no siempre reflejan las tareas más complejas e interpretativas que las personas realizan con las visualizaciones, lo que hace que la aplicabilidad de las pautas sea incierta. Tales tareas a menudo requieren que los analistas integren múltiples patrones y estadísticas para hacer inferencias sobre los datos visualizados. Considere a un científico computacional que analiza un conjunto de gráficos codificados por colores que representan diferentes resultados del modelo. En este contexto, el científico puede estar interesado en comprender los modelos detrás de cada visualización, comparar estos modelos y emitir un juicio sobre las relaciones entre los modelos. El científico puede concluir, por ejemplo, que los modelos concuerdan en gran medida a pesar de la aparente variación en el resultado o puede recurrir a la experiencia previa para inferir que un modelo específico muestra un resultado aparentemente inusual. Al decidir qué mapa de colores usar para este tipo de análisis, podríamos suponer que las tareas perceptivas del componente, como comparar valores clave, se apoyan mejor en una rampa secuencial o divergente (p. ej., viridis o gris-rojo). Sin embargo, no está claro si las operaciones más cognitivas de inferencia y juicio del modelo, como distinguir diferencias significativas del ruido, se ven afectadas de manera similar. Es posible que los determinantes cognitivos, como la capacidad de distinguir categóricamente entre colores o asociar colores con distintos nombres y conceptos [20, 31, 49], desempeñen un papel importante en la inferencia visual. Una diferencia en el procesamiento del mapa de colores entre las tareas más cognitivas y perceptivas puede requerir diferentes pautas para un diseño de visualización efectivo, según la naturaleza del análisis.

Nuestro objetivo en este trabajo es doble. En primer lugar, nuestro objetivo es identificar los principios de diseño de mapas de colores para mejorar la inferencia gráfica a partir de visualizaciones cuantitativas. En segundo lugar, probamos si ciertos mapas de colores pueden hacer que las personas vean diferencias falsas que no están presentes en los datos subyacentes (es decir, falsos positivos). Con ese fin, comparamos los diseños de mapas de colores tradicionales con una métrica cognitiva no convencional: **la variación del nombre del color**. Estudiamos el impacto de esta métrica en la capacidad de las personas para hacer juicios gráficos sobre visualizaciones de conjunto. Los participantes vieron una alineación de campos escalares (Figura 1) y se les pidió que identificaran una gráfica 'extraña' que pertenece a un modelo incompatible. Descubrimos que las características del mapa de colores afectan de manera predecible el rendimiento en esta tarea, y los esquemas de arco iris brindan la mayor precisión. Específicamente, la precisión se correlacionó positivamente con el nivel de variación del nombre de un mapa de colores, lo que proporcionó una ventaja significativa a los mapas que atraviesan una variedad de colores identificables. En un segundo experimento, medimos el impacto de esta métrica en la especificidad para probar la hipótesis de que la categorización de colores conduce a inferencias falsas. Los resultados mostraron que, a pesar de la mayor sensibilidad, los mapas de arco iris no conducían necesariamente a más falsos positivos. Nuestros hallazgos sugieren colectivamente beneficios para maximizar el color categórico y la variación de nombres en mapas cuantitativos. Discutimos los resultados, especulamos sobre los mecanismos perceptivos y cognitivos subyacentes y sugerimos estrategias alternativas de diseño de mapas de colores.

2 TRABAJO RELACIONADO

Revisamos los principios básicos de codificación de colores y discutimos los enfoques actuales para el diseño de mapas de colores. Luego examinamos estudios empíricos recientes del color en la visualización, centrándonos en mapas de color cuantitativos (en oposición a categóricos).

2.1 Directrices para el mapeo de colores

Hay una gran cantidad de directrices para elegir secuencias de colores para datos cuantitativos [46, 51]. Si bien el vocabulario varía de una fuente a otra [11], las pautas generalmente coinciden en tres principios [51]: 1) Orden: una buena secuencia de mapas de colores debe ser naturalmente ordenada (por ejemplo, de un color azul frío a un rojo cálido); 2) Continuidad: el mapa de colores solo debe reflejar las diferencias reales en los datos sin crear límites artificiales; y 3) Uniformidad de percepción: los colores adyacentes deben reflejar distancias de percepción uniformes a lo largo de la secuencia, de modo que un paso en la magnitud de los datos se corresponda con un paso de percepción equivalente en el color.

Teniendo en cuenta estos principios, la mayoría de los investigadores abogan por rampas con aumento monótono de la luminancia [5] mientras desaconsejan el uso de esquemas 'espectrales' (p. ej., arco iris) [45]. libros de texto fundamentales

también fomenta el uso de mapas secuenciales y divergentes para datos numéricos, pero recomienda esquemas espectrales para datos categóricos [36]. Se cree que la tendencia de los arco iris a crear límites entre los tonos (a veces denominado efecto de "bandas de tonos") engaña a los espectadores [51]. Sin embargo, el impacto de este tipo de bandas en la interpretación de los datos aún no se comprende bien [37, 39].

A pesar de las recomendaciones constantes de los investigadores, los mapas de colores del arco iris continúan siendo populares entre los profesionales [35]. No está claro por qué los practicantes continúan adoptando diseños aparentemente inferiores. Los investigadores especulan que las personas encuentran atractivos los mapas de colores del arco iris [4], lo que a su vez impulsa sus elecciones de codificación de colores. Sin embargo, la evidencia también sugiere que las personas prefieren combinaciones de colores que tengan un matiz armonioso [48], de lo que carecen la mayoría de los arco iris. ¿Podrían las preferencias de los profesionales reflejar una noción utilitaria que de algún modo se pasa por alto en las pautas convencionales? Nuestro trabajo considera esta pregunta en el contexto de una tarea de inferencia gráfica.

2.2 Enfoques del diseño de mapas de colores Las

herramientas para seleccionar o crear mapas de colores dictan en gran medida los tipos de codificaciones de colores que se utilizan en las visualizaciones. Podría decirse que la herramienta de selección de colores más popular es ColorBrewer [19], que proporciona un conjunto de mapas de colores hechos a mano para variables cuantitativas o categóricas. Estos mapas de colores se basan en una extensa investigación empírica [7, 8] y, en general, se adhieren a las pautas establecidas. Sin embargo, las paletas de colores de Brewer et al están dirigidas principalmente a mapas temáticos (es decir, coropletas) [6, 38]. No está claro si estas recomendaciones se generalizan a representaciones espaciales continuas, como campos escalares donde los valores de datos suelen crear gradaciones de color suaves.

Los mapas continuos aparecen con frecuencia en dominios como la ciencia computacional [66], imágenes médicas [4], astrofísica [29] y sensores remotos para aplicaciones críticas (p. ej., visualización de datos de huracanes). El impacto analítico y comunicativo potencialmente grande de estas visualizaciones destaca la necesidad de principios de codificación de colores validados. Aunque muchos profesionales continúan usando mapas de colores del arco iris, algunos sistemas de visualización han adoptado alternativas minuciosamente diseñadas como opciones predeterminadas. Por ejemplo, cool-warm en Paraview [34] y viridis en Matplotlib [57] tienen como objetivo reemplazar el arco iris con alternativas más basadas en la percepción (p. ej., asegurando la uniformidad de la percepción). Sin embargo, estos sistemas ofrecen una pequeña biblioteca de opciones de mapas de colores fijos, lo que dificulta que los diseñadores los personalicen [31].

Algunas herramientas permiten a los diseñadores construir sus propios mapas de colores guiados por convenciones. Los diseñadores pueden crear rampas especificando un puñado de puntos clave, que luego se asignan a una ruta geométrica (por ejemplo, una línea o una curva simple) y se muestrean de manera uniforme en un espacio de color perceptivo. Por ejemplo, Wijnelaars et al. observó que las rampas ColorBrewer tendían a atravesar curvas cúbicas [65]. Su herramienta permite a las personas manipular puntos de control clave en estas curvas, generando mapas que imitan a ColorBrewer. Otros enfoques se basan en tareas específicas para impulsar la selección del mapa de colores [2, 56]. Por ejemplo, PRAVDAColor [2] aplica convenciones de diseño basadas en características de datos y tipos de tareas para generar mapas de colores personalizados. Sin embargo, trabajos más recientes sugieren que las suposiciones sobre el diseño de mapas de colores, como la afirmación de que CIELAB coincide con las diferencias percibidas en las marcas de datos [52, 54], pueden no ser válidas en la práctica. Un análisis reciente de las prácticas también encontró que los mapas de colores de los diseñadores a menudo no obedecen las pautas convencionales [53]. Las evaluaciones empíricas del rendimiento del mapa de colores apuntan a resultados contradictorios que a menudo no se pueden reconciliar con las pautas [41, 42]. Los datos empíricos limitados, junto con la confianza casi universal en las herramientas de diseño estándar, resaltan la importancia de validar las pautas de codificación de colores en tareas realistas.

2.3 Estudios empíricos de mapas de colores cuantitativos Históricamente,

las pautas de diseño de mapas de colores se han basado en gran medida en la intuición. Los estudios empíricos posteriores a veces confirman [32, 60] ya veces desafían [41, 62] esas directrices. En un estudio anterior, Ware abogó por mapas de color que aumentaran la luminancia para ayudar a la comprensión de formas [59]. Sin embargo, el mismo estudio también encontró variaciones no monótonas (p. ej., en el arco iris) para reducir los errores de contraste simultáneos, mejorando efectivamente la estimación del valor. Reda et al. confirmado

esta hipótesis, pero descubrió que las características de frecuencia espacial de los datos dictan en gran medida el diseño óptimo del mapa de colores [41]. Ware et al. también evaluó qué tan bien los espectadores pueden encontrar características de alta frecuencia espacial con diferentes mapas de colores. Descubrieron que el umbral de detección no se puede atribuir únicamente a la luminancia, sino que se suma a la variación de croma y tono [62]. Algunos trabajos anteriores respaldan la efectividad de los esquemas de arco iris en la interpretación de mapas temáticos y superficies continuas [7, 23, 30]. Sin embargo, estudios recientes informan hallazgos que están en línea con las preferencias tradicionales hacia los mapas de color orientados a la luminancia. Por ejemplo, en un estudio con médicos, Borkin et al. encontró que los mapas de colores del arco iris son sustancialmente menos precisos que los esquemas divergentes para diagnosticar enfermedades del corazón [4]. Los autores plantearon la hipótesis de que las rampas divergentes son mejores para encontrar regiones de bajo valor, que fueron clave para completar esa tarea con precisión. Liu & Heer también encontraron que el azabache (una variante del arcoiris) es generalmente inexacto para juzgar las distancias perceptivas entre los parches de color [32]. Dasgupta et al. descubrió que los científicos del clima pueden estimar con mayor precisión los valores medios del mapa con un esquema secuencial que con un arco iris [13]. Sin embargo, informaron una mayor utilidad de los mapas de arco iris en la estimación de la varianza. Usando métodos tomados de la psicofísica, Reda y Papka también encontraron que las personas exhiben un JND más bajo (es decir, una sensibilidad más alta) al estimar la varianza espacial con el arco iris [42].

En este trabajo, estudiamos una tarea que enfatiza la inferencia gráfica [10]. Este tipo de tarea nos permite probar qué tan bien los espectadores pueden discriminar entre modelos subyacentes a un conjunto de visualizaciones codificadas por colores. Sospechamos que tal tarea es importante en muchos contextos científicos, donde uno no solo puede estar interesado en una sola característica, sino también en interpretar las relaciones entre múltiples conjuntos de datos.

También anticipamos que tales tareas interpretativas requieren características de mapas de colores diferentes a las necesarias para tareas de percepción, como la estimación de valores, según la evidencia de estudios previos. En este artículo, comparamos específicamente las recomendaciones de diseño basadas en la percepción con una métrica cognitiva de categorización de colores [20].

3 MÉTODOS

El análisis visual de 'conjuntos' surge con frecuencia en la ciencia. Por ejemplo, los científicos del clima a menudo analizan datos de campo escalares para pronosticar el aumento de la temperatura global bajo diferentes modelos climáticos [66]. Un ingeniero aeroespacial podría estar interesado en comparar niveles de ruido simulados generados por diferentes prototipos de motores a reacción [27]. Más allá de los datos simulados, los investigadores de enfermedades infecciosas a menudo estudian histogramas 2D codificados por colores para comprender las características de los receptores de las poblaciones de células inmunitarias sometidas a técnicas de citometría de flujo [50]. Estas tareas no se limitan a decodificar valores brutos a partir del color. Más bien, los analistas a menudo necesitan inferir propiedades del modelo detrás de las visualizaciones y evaluar potencialmente las relaciones entre múltiples modelos alternativos. La inferencia se complica aún más por el hecho de que una visualización generalmente representa una sola muestra de lo que podría ser una gran cantidad (o infinita) de muestras potenciales de un modelo o del mundo real. Se puede obtener una visualización visualmente distinta del mismo fenómeno simplemente volviendo a muestrear el modelo o tomando un nuevo conjunto de observaciones. Por lo tanto, un analista tiene que pensar qué características de la imagen se deben a las propiedades del modelo real y cuáles se deben a las variaciones aleatorias que surgen del proceso de muestreo. La complejidad involucrada en este tipo de tarea brinda la oportunidad de probar nuestras suposiciones acerca de cómo funciona el color en la visualización, al mismo tiempo que examina una clase importante de análisis relevante para muchos profesionales (por ejemplo, científicos profesionales).

3.1 Tarea modelo

Para simular experimentalmente los análisis anteriores, ideamos una tarea basada en el concepto de inferencia gráfica de Buja et al [10, 63]. La tarea evalúa qué tan bien un espectador puede discriminar visualmente entre dos distribuciones. La idea es ocultar un gráfico de los datos "reales", que normalmente contiene un patrón de interés, en una alineación que también contiene varias visualizaciones "señuelo". Los gráficos de señuelo representan conjuntos de datos que podrían haber surgido por casualidad, normalmente muestreados a partir de una distribución nula. La tarea del espectador es identificar los datos reales a partir del valor nulo infiriendo las verdaderas diferencias entre las dos distribuciones.

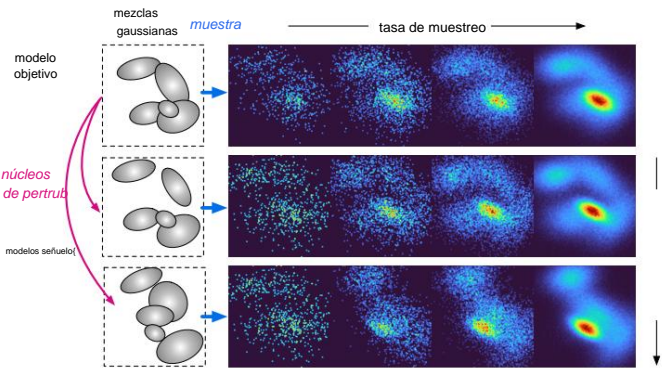


Fig. 2. Los campos escalares de estímulo se sintetizan mediante un muestreo iterativo de un modelo de mezcla gaussiana. La primera fila corresponde a un modelo objetivo. La segunda y tercera filas muestran dos modelos de señuelo de ejemplo obtenidos al perturbar el objetivo. La tercera fila corresponde a un señuelo con mayor perturbación (mayor divergencia KL) para una prueba más fácil. Las columnas representan tasas de muestreo crecientes, y la columna más a la derecha recrea casi a la perfección la densidad de probabilidad real. En nuestros experimentos, el muestreo se detiene cerca de la Columna 3 para brindar a los participantes una representación imperfecta del modelo.

El protocolo de Buja et al fue concebido como un método para protegerse contra falsos descubrimientos [10]. Sin embargo, la prueba también es útil para evaluar el poder de las técnicas de visualización competidoras [12, 22, 64]: si un método de visualización en particular hace que sea más fácil distinguir las diferencias más pequeñas del modelo, entonces se puede decir que esa visualización tiene un mayor poder estadístico. Por ejemplo, Hofmann et al. empleó inferencia gráfica para comparar dos diseños de visualización diferentes para series temporales [22]. Usamos un enfoque similar para medir el poder discriminativo de diferentes diseños de mapas de colores para la inferencia. Nos enfocamos en mapas de color continuos para datos escalares e histogramas 2D, ya que esas visualizaciones surgen con frecuencia en aplicaciones científicas, donde a menudo vemos la necesidad de inferencia. En nuestros experimentos, el espectador ve una alineación de cuatro campos escalares codificados por colores (consulte la Figura 1 para ver ejemplos). Los campos se muestrean a partir de uno de dos modelos: un objetivo y un señuelo. Específicamente, una de las cuatro visualizaciones se muestrea del modelo objetivo, mientras que las otras tres se muestrean del modelo señuelo. Según el método original [63], el espectador debe identificar la trama que "no pertenece": en este caso, la visualización que proviene del modelo de destino.

La inferencia gráfica puede verse como análoga a una prueba de hipótesis estadística en el sentido de que ambas están destinadas a discriminar diferencias verdaderas [10,63]. Sin embargo, es importante reconocer que las visualizaciones admiten una variedad de estilos inferenciales, y la tarea que adoptamos aquí representa un ejemplo de una gran clase de tareas inferenciales. En este artículo, usamos el término 'inferencia' como abreviatura de inferencia gráfica.

3.2 Estímulos

Empleamos un procedimiento de generación de datos sintéticos para generar campos 2D con pares de modelo objetivo-señuelo conocidos. El procedimiento crea un modelo de destino y posteriormente lo perturba para producir señuelos (consulte la Figura 2 para ver una ilustración de este proceso). El objetivo se sintetiza primero a partir de una combinación de núcleos gaussianos 2D. Los núcleos se centran aleatoriamente dentro del dominio espacial y sus parámetros (valor esperado, desviación estándar y covarianza) varían para generar configuraciones pseudoaleatorias. Luego, los núcleos se integran para producir una distribución de probabilidad 2D conjunta. Para generar un modelo señuelo, los parámetros del kernel individuales del modelo de destino se perturban aleatoriamente dentro de un rango fijo (determinado a través del pilotaje). Una perturbación más grande da como resultado una divergencia objetivo-señuelo más obvia y, por lo tanto, un juicio más fácil.

Todo este proceso se repite por separado para cada ensayo.

El procedimiento sintético anterior produce modelos que exhiben una variedad de arreglos y patrones espaciales, como se muestra en las Figuras 1 y 2. Cualitativamente, nuestra implementación genera tanto modelos dominados por grandes densidades como aquellos que están compuestos por características más pequeñas.

Los pares objetivo-señuelo emergentes también difieren en una variedad de formas a través de los estímulos; observamos diferencias caracterizadas por el cambio de las densidades o "espacios vacíos", la adición o eliminación de "puntos calientes" y cambios generales en el patrón 2D. Esta amplia diversidad ayuda a garantizar que nuestros resultados experimentales no estén sesgados por formas o patrones específicos.

Generamos un estímulo de alineación (Figura 1) muestreando los modelos objetivo y señuelo para producir cuatro campos (uno del objetivo y tres del señuelo). Debido a que los campos representan sorteos aleatorios de cualquier distribución, una alineación siempre exhibirá variaciones entre campos, incluso para campos extraídos de la misma distribución. La visualización de muestras, a diferencia de las densidades de probabilidad reales, proporciona a los participantes representaciones imperfectas del modelo, cuya calidad se ve afectada por el proceso de muestreo. Esta incertidumbre es clave para operacionalizar la inferencia gráfica. Primero, la aleatoriedad introducida por el muestreo evita que los participantes basen sus juicios en pequeñas variaciones en una parte de la imagen. En cambio, deberán comparar las cuatro visualizaciones de manera integral e integrar múltiples percepciones [40] para tomar una determinación resumida. En segundo lugar, los participantes deberán distinguir las características visuales que reflejan las diferencias sistemáticas del modelo de las fluctuaciones debidas al muestreo aleatorio. La combinación de estos dos factores da lugar a una tarea más interpretativa que en experimentos anteriores, que enfatizan las percepciones estadísticas o localizadas [32, 41, 59, 62]. Esta tarea, a su vez, nos permite cambiar el enfoque de los problemas de percepción del color a las cuestiones de cómo los códigos de color afectan la inferencia y el juicio basado en modelos.

4 HIPÓTESIS Y MÉTRICAS

Las pautas de asignación de colores, como garantizar distancias de percepción iguales y minimizar las discontinuidades de color, están inspiradas en los espacios de color tradicionales (por ejemplo, CIELAB o LUV). Estos modelos de color se aproximan a la apariencia relativa de un pequeño número de parches de color aislados. Por el contrario, las visualizaciones a menudo comprenden una cantidad significativa de marcas o gradaciones de color (como en campos escalares o mapas de calor). Los diseños y las condiciones de visualización en estas pantallas de datos son mucho más complejos que los que se suponen en los modelos de color convencionales, lo que hace que estos modelos (y las pautas en las que se basan) sean menos confiables en la práctica [54]. Además, mientras que las métricas de la distancia perceptiva del color pueden ser importantes en las tareas de percepción (p. ej., estimar un valor de color), esas mismas métricas podrían ser menos relevantes para la inferencia gráfica, donde las diferencias cuantitativas precisas pueden no importar tanto como la calidad general de la información. esas diferencias.

Presumimos que las características cognitivas accesibles del color [15], como la capacidad de reconocer fácilmente los colores por su nombre y distinguir diferentes tonos categóricamente, juegan un papel más importante en la inferencia gráfica. En particular, esperamos que la nomenclatura del color ayude a las personas a razonar sobre diferencias de grano grueso a pesar de una posible pérdida de precisión al comparar variaciones perceptivas más pequeñas. El trabajo empírico reciente proporciona evidencia para corroborar esta conjetura. Por ejemplo, Reda y Papka descubrieron que los observadores estimaban eficientemente las propiedades estructurales (p. ej., gradientes) en campos escalares utilizando arcoíris [42]. Los participantes parecieran aprovechar las bandas de colores discretos emergentes (por ejemplo, parches azules, verdes y rojos en el arcoíris). Al juzgar visualmente el tamaño aparente, la aritmética y el agrupamiento de estas características, se pueden obtener estimaciones bastante confiables sobre varias propiedades estadísticas. La categorización por colores puede parecer una mala idea para los conjuntos de datos cuantitativos [5], pero esta discretización puede ayudar en la práctica al proporcionar una representación "destacada" de los datos. Las características segmentadas, a su vez, pueden facilitar la extracción de ciertas propiedades del conjunto de un vistazo [16].

La adhesión dogmática a los mapas de colores "malos" por parte de los profesionales de la visualización también puede sugerir alguna utilidad no observada. Muchos científicos profesionales siguen evitando las directrices de los expertos en favor de mapas de arcoíris aparentemente inferiores [35, 45]. Las justificaciones comunes para esta preferencia incluyen que los profesionales dudan en cambiar sus convenciones [34] o encuentran que las visualizaciones coloridas son estéticamente atractivas [4]. Sin embargo, también es posible que los profesionales encuentren útiles los arco iris para algunas tareas, probablemente debido a los beneficios cognitivos que no se pueden discernir de los modelos de apariencia de color tradicionales, una hipótesis que probamos. También consideramos un contraargumento de que los mapas más coloridos pueden hacer que las personas "alucinen" con características que no están en los datos reales [28]. Este último podría manifestarse como una reducción en la especificidad de la inferencia.

Postulamos dos hipótesis basadas en las observaciones anteriores: H1:

argumentamos que las métricas de mapas de colores basadas en la percepción, como la uniformidad, el orden y la suavidad [11], no son adecuadas para tareas inferenciales. En cambio, planteamos la hipótesis de que el rendimiento estará vinculado a la capacidad de las personas para razonar categóricamente sobre el color en una visualización.

Esto último puede verse facilitado por mapas que combinan una variedad más amplia de colores identificables. Debido a que no existe una forma establecida de medir el grado de separación categórica en mapas de colores continuos, proponemos una nueva métrica, variación de nombre de color (ver §4.1) basada en un modelo popular de distancia de nombre [20].

H2—Esperamos que los mapas de colores con una gran variación de nombres provoquen que las personas para detectar diferencias falsas entre visualizaciones del mismo modelo.

Se cree que la categorización por colores es engañosa para los tipos de datos que varían continuamente [5]. Por ejemplo, la transición de verde a amarillo en un arcoíris hace que las personas perciban un límite nítido, que podría confundirse con una característica de datos. Por lo tanto, si bien anticipamos que las personas serán más sensibles con los mapas de colores que varían mucho en el nombre, sospechamos que estas rampas también aumentarán la tasa de juicios falsos positivos. Este último se manifiesta como una mayor probabilidad de informar diferencias entre visualizaciones cuando los modelos subyacentes son, de hecho, idénticos.

4.1 Variación del nombre del color

Los nombres de colores se refieren a las asociaciones lingüísticas básicas que hacemos con el color (p. ej., "rojo", "verde" y "azul") [3, 9, 67]. La denominación de colores también puede ser una herramienta cognitiva, que proporciona a las personas una forma de categorizar, discutir y razonar sobre el color. En las visualizaciones, la categorización de colores puede ayudar a las personas a pensar en patrones complejos al permitir la atención selectiva a características de datos específicas, que de otro modo podrían ser demasiado borrosas. La categorización también puede ayudar a reducir una gran cantidad de puntos de datos a un conjunto más pequeño de "contenedores" con los que se puede razonar más fácilmente. Por lo tanto, la medida en que un mapa de colores facilita la categorización puede correlacionarse con el desempeño en tareas de inferencia gráfica (H1).

Para aproximar la categorización, usamos el modelo de distancia de nombres de Heer & Stone [20]: dados dos colores CIELAB, el modelo genera la probabilidad de que el par tenga nombres diferentes. Extendemos esta métrica a un mapa de colores continuo al medir la probabilidad de que un par de colores adyacentes estén asociados con nombres distintos. La suma de estas habilidades de probabilidad nos da una aproximación de la cantidad de colores distintos contenidos en el mapa de colores. Nos referimos a esta medida como variación del nombre del color (CNV). Formalmente:

$$\text{CNV}(C) = \sum_{y=1}^n \bar{y}(C(i\bar{y})/n, C_i/n) \quad (1)$$

donde C_i/n es un color muestreado de una rampa continua C en la i -ésima posición, n es el número de muestras que se tomarán (usamos $n = 8$), y \bar{y} es el coseno nombre-distancia de Heer & Stone [20], calculado como:

$$\bar{y}(C(i\bar{y})/n, C_i/n) = 1/\cos(\text{TC}(i\bar{y})/n, \text{TC}(i)/n) \quad (2)$$

donde TC es la matriz de conteo de términos de color derivada de cinco millones de muestras a través de una encuesta en línea. 1 Una variación de nombre más alta indica

una rampa que combina una variedad de colores con nombres distintos, lo que a su vez se presenta como una mayor probabilidad de categorización de color. Por ejemplo, los mapas de colores del arco iris que recorren un rango de tonos saturados obtienen una puntuación más alta en esta métrica, mientras que las rampas de un solo tono generalmente muestran una variación de nombre mucho menor. Las rampas divergentes y de varios tonos tienden a superponerse y se encuentran en algún lugar entre los arcoíris y las rampas de un solo tono (consulte la Figura 3). Aunque asociamos intuitivamente los nombres de los colores con diferentes matices, la variación de los nombres no está ligada únicamente a ningún factor de percepción único. A través de un corpus de 235 mapas de color (ver §4.2), los componentes CIE L exhibieron una correlación comparable con la variación del nombre (r de Pearson = 0.754, 0.727 y 0.718, respectivamente). Por lo tanto, la variación del nombre parece estar impulsada casi por igual por los cambios en la luminosidad, el croma y el matiz.

<http://blog.xkcd.com/2010/05/03/color-survey-results/>

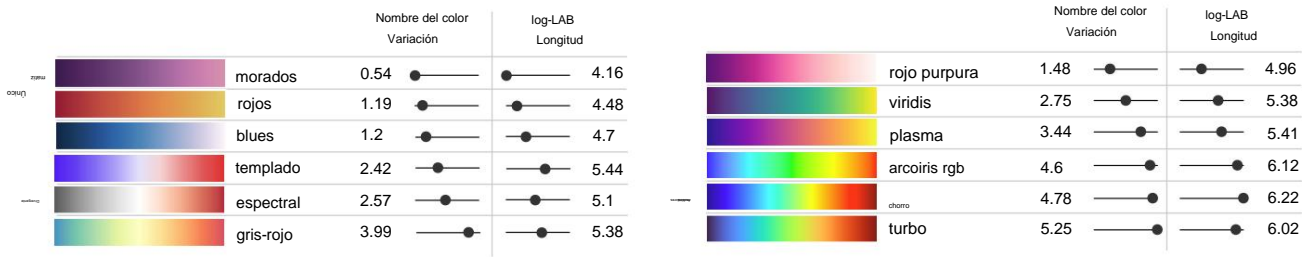


Fig. 3. Seleccionamos un conjunto de 12 rampas de color que representan cuatro familias de diseño convencionales: rampas de un solo tono, de varios tonos, divergentes y arcoíris. Estas rampas exhiben un rango de puntajes de variación de nombres de colores, la métrica principal que usamos para modelar el rendimiento del sujeto. También consideramos una métrica CIELAB transformada logaritmicamente como una aproximación de la variación del nombre.

4.2 Selección de rampa

Para comprender qué propiedades del mapa de colores facilitan la inferencia gráfica, tomamos muestras de un conjunto de 12 mapas de colores de un corpus de 235 diseños únicos. El corpus original incluía una variedad de secuencias recopiladas de ColorBrewer [19], ColourLovers, 2 MATLAB, Matplotlib [57], Google [33] y Tableau. La muestra resultante de 12 rampas representa un continuo de niveles de variación de nombres en cuatro familias de diseño diferentes (secuencial de un solo tono, secuencial de múltiples tonos, divergente y arcoíris). La Figura 3 ilustra las rampas seleccionadas.

Además de representar ampliamente los mapas de color de uso común, las rampas seleccionadas ofrecen comparaciones interesantes. Por ejemplo, el rojo-púrpura de Color Brewer tiene una variación de nombre más alta que los púrpuras de ColourLover pero más baja que el plasma, aunque las tres rampas comparten una gama similar de tonos púrpura. De manera similar, incluimos tres variantes de arco iris: turbo, jet y un arco iris RGB estándar. Esta selección representa una variación de nombre de color relativamente alta, pero también diferencias notables en las características de percepción. Por ejemplo, el turbo de Google se presenta como una alternativa al jet perceptualmente mejorada debido a su perfil de luminancia casi uniforme.

5 EXPERIMENTO 1

El Experimento 1 prueba la hipótesis de que la utilidad de un mapa de colores para la inferencia gráfica se correlaciona con su variación de nombre (H1). Esperamos que los participantes sean más precisos con las rampas que combinan una mayor variedad de colores identificables. Para probar esta hipótesis, llevamos a cabo un estudio colaborativo con el mapa de colores como factor entre sujetos.

Analizamos los resultados utilizando tres modelos, cada uno de los cuales proporciona una explicación contradictoria de los resultados. El primer modelo predice que los mapas de color de la misma familia de diseño mostrarán una precisión comparable, pero que las diferentes familias (tono único, multitono, divergente y arcoíris) exhibirán un rendimiento variable. Por el contrario, el segundo modelo es más parsimonioso y predice el rendimiento utilizando una sola métrica cuantitativa: la variación del nombre del mapa de colores. El modelo es capaz de distinguir de forma más granular entre mapas de colores, incluso para aquellos dentro de la misma familia de diseño (p. ej., viridis y plasma). El tercer modelo utiliza la longitud transformada logarítmica de un mapa de colores en el espacio CIELAB como una aproximación para la variación del nombre.

Al comparar estos modelos alternativos, evaluamos la idoneidad de nuestras dos métricas (variación del nombre del color y longitud del log-LAB) para predecir la composición efectiva del mapa de colores. Además, al contrastar con el modelo de familia de diseño, probamos si estas métricas proporcionan un indicador más útil que las pautas convencionales (p. ej., uniformidad perceptual y monotonicidad de luminancia). Usamos el criterio de información bayesiano (BIC) para comparar el ajuste del modelo. La elección de BIC refleja nuestro objetivo de descubrir los principios generales de diseño de mapas de colores en lugar de predecir con precisión el rendimiento por sujeto. BIC favorece modelos más simples penalizando el número de parámetros; sin embargo, todavía se favorecen los modelos más complejos si mejoran el ajuste a los datos.

5.1 Tarea

Empleamos la tarea de inferencia gráfica descrita en §3.1. Cada estímulo comprende una alineación de cuatro campos escalares (consulte la Figura 1) que se muestrean de uno de dos modelos: un modelo objetivo y un modelo señuelo. Todos los campos generados midieron 200x200 píxeles, subtendiendo aproximadamente 4° de ángulo visual cuando se vieron desde 30 pulgadas de distancia a 96 DPI (la densidad de píxeles estándar para navegadores web [58]). Se mostró una escala de colores a la derecha de la alineación como referencia. Se instruyó a los participantes para que identificaran la visualización que "no pertenece". Indicaron su elección haciendo clic en una de las cuatro imágenes y confirmaron la selección presionando Enter para pasar a la siguiente prueba.

Limitamos las alineaciones a cuatro visualizaciones, a diferencia de las diez gráficas más típicas [63], para reducir el tiempo de respuesta por prueba y permitir una mayor cantidad de estímulos por sujeto. Si bien esta configuración aumenta el nivel alfa para una prueba de alineación individual de 0,1 a 0,25, nuestro diseño experimental no depende de alcanzar la significación estadística en cada prueba. En cambio, para acomodar el poder estadístico reducido, aumentamos nuestra tasa de muestreo y recopilamos 88 ensayos por participante. También empleamos un conjunto diverso de distribuciones para garantizar que los resultados no estén sesgados por características específicas del modelo (p. ej., conglomerados o puntos críticos).

5.2 Diseño experimental Evaluamos

12 mapas de colores (Figura 3, §4.2) en un diseño entre sujetos. Cada participante vio un mapa de colores de cada una de las cuatro familias de diseño para un total de cuatro mapas de colores por participante.

El experimento se bloqueó mediante un mapa de colores con un orden de bloques aleatorio. Los participantes completaron 22 pruebas con cada mapa de colores, lo que resultó en 88 pruebas por participante. Cada ensayo consistió en una alineación recién generada (como se describe en §3.2). La dificultad se controló variando sistemáticamente la divergencia entre los modelos objetivo y señuelo. Medimos la divergencia usando Kullback-Leibler (KL), una métrica teórica de información popular para cuantificar la distancia entre dos distribuciones [21]. Muestreamos estímulos de manera uniforme con divergencia KL en un rango de 10 a 35%. Una divergencia más baja corresponde a diferencias de modelo más pequeñas (es decir, juicios más difíciles). Este rango se determinó a través de pruebas piloto para reflejar las probabilidades de éxito esperadas que son ligeramente mayores que el azar (P = 0,25) hasta casi perfectas. Como referencia visual, la fila superior de la Figura 1 representa estímulos con un 27 % de divergencia (algo fácil) mientras que la fila inferior tiene un 15 % (diferencias menos obvias).

Las pruebas individuales dentro de cada bloque (correspondientes a diferentes niveles de dificultad) se mostraron en orden aleatorio. Además de las pruebas reales, insertamos aleatoriamente 2 comprobaciones de compromiso por bloque para un total de 8 comprobaciones a lo largo del experimento. Estas comprobaciones consistieron en estímulos muy fáciles (40% de divergencia, o cuatro veces más fácil que la dificultad inicial).

5.3 Procedimiento

Los participantes fueron evaluados primero por deficiencia de visión del color utilizando 14 paneles Ishihara. Luego vieron un tutorial y completaron 24 pruebas de práctica. Estos ensayos de práctica incluían comentarios que informaban a los participantes si habían adivinado correctamente o indicaban la respuesta correcta en caso contrario. Durante el entrenamiento, cada participante vio a sus cuatro asignados

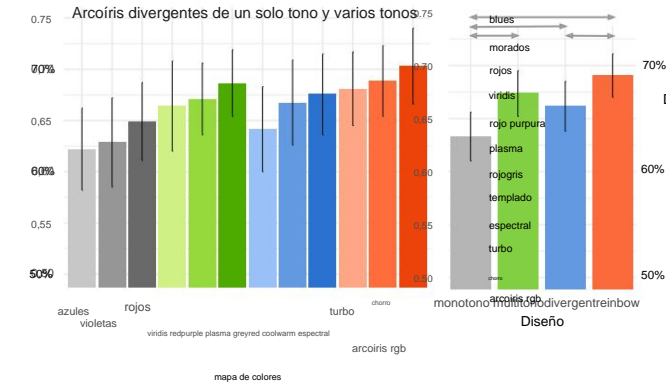


Fig. 4. Exactitud media observada para los 12 mapas de color (izquierda). Los mismos datos también se agrupan por familia de diseño (derecha). La barra de errores son intervalos de confianza del 95%. Las flechas indican diferencias significativas en la precisión entre las familias de diseño (p < 0,05 con el ajuste de Tukey).

mapas de colores en orden aleatorio. Después de la práctica, los participantes completaron las pruebas reales, en las que no se dieron comentarios. Al final del experimento, les pedimos a los participantes que proporcionaran una breve descripción cualitativa de la estrategia que siguieron en la tarea, incluidas las "características o características visuales" en las que basaron sus juicios. Los participantes concluyeron completando una breve encuesta demográfica.

5.4 Participantes

Reclutamos a 180 participantes (115 hombres, 64 mujeres y otros 2) de Amazon Mechanical Turk, compensándolos con \$3 cada uno. Los participantes tenían una edad media autoinformada de 37,4 años (ST D = 11,7). Se excluyó del análisis a cualquier participante que no pasara la prueba de visión cromática o juzgara mal la mayoría de las comprobaciones de compromiso (el rendimiento medio en las comprobaciones de compromiso fue del 95,6 %). Luego reclutamos nuevos participantes para reemplazar a los excluidos hasta que alcanzamos el tamaño de muestra anterior. Para mitigar los efectos de la variación entre participantes, cada mapa de colores fue visto por exactamente 60 personas diferentes, y cada mapa de colores de una familia de diseño en particular se probó al menos dos veces con todos los demás mapas de colores de una familia diferente.

5.5 Resultados

Los participantes completaron el experimento en 20,4 minutos en promedio (ST D = 8,2). En total, obtuvimos 15.840 juicios binarios (es decir, si el participante había identificado correctamente la distribución objetivo).

La precisión media en la tarea fue del 66,5 % (la probabilidad de rendimiento es del 25 %).

La figura 4 traza la precisión en función del mapa de colores y la familia de diseño.

Analizamos los resultados utilizando tres modelos de regresión logística: 1) un modelo que predice la probabilidad de inferencia correcta por separado para cada familia de diseño; 2) un modelo que predice el rendimiento en función de la variación del nombre de un mapa de colores; y 3) un modelo que representa el refinamiento post-hoc del segundo modelo con la longitud log-LAB que sirve como una aproximación a la variación del nombre (la distancia LAB sin procesar dio como resultado un ajuste deficiente que no pudimos obtener un modelo bien formado). Los tres modelos incluyeron un parámetro adicional de efectos fijos correspondiente a la divergencia entre el objetivo y el señuelo (es decir, la dificultad de la prueba) y una intercepción aleatoria para tener en cuenta la variación individual entre los sujetos. Primero analizamos cada modelo por separado y luego comparamos su ajuste relativo. La Figura 6 representa las respuestas del modelo frente a los datos observados. La Tabla 2 proporciona los parámetros ajustados para cada modelo.

5.5.1 Familia de diseño Ei

modelo de familia de diseño contiene cuatro variables discretas, una para cada uno de los cuatro grupos de diseño (secuencial de un solo tono, secuencial de múltiples tonos, divergente y arcoíris). En promedio, los participantes fueron más precisos cuando vieron mapas de colores del arco iris (Precisión media: 69,1 %, IC del 95 %: 67,0–71,1). Los tonos múltiples fueron los siguientes más efectivos (media: 67,4 %, IC: 65,2–69,5), seguidos de las rampas divergentes (media: 66,2 %, IC: 63,8–68,5) y de un solo tono (media: 63,3 %, IC: 61,0–65,6).

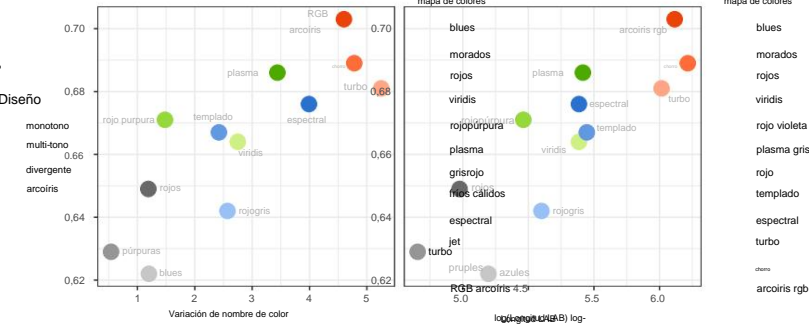


Fig. 5. Precisión media del mapa de colores en función de la variación del nombre del color (izquierda) y la longitud del log-LAB. Ambas métricas muestran una correlación significativa con el rendimiento de inferencia observado en el Experimento 1.

Realizamos comparaciones por pares entre las cuatro familias de diseño utilizando la Z de Wald y empleamos el ajuste de Tukey para comparaciones múltiples (consulte la Figura 4 a la derecha para obtener un resumen de las diferencias significativas).

Los mapas de colores del arco iris fueron significativamente más precisos que los de un solo tono (Z = 6,020, p < 0,001) o las rampas divergentes (Z = 3,107, p < 0,05). Los mapas de colores múltiples y divergentes fueron más precisos que los de un solo tono (Zmh = 4,153, pmh < 0,001; Zd = 2,922, pd < 0,05).

5.5.2 Variación del nombre del color

Nuestro segundo modelo captura el rendimiento únicamente en función de la variación del nombre del color, sin hacer distinción entre familias de diseño. La prueba de Wald indica que la variación del nombre es un predictor significativo del rendimiento (Z = 5,526, p < 0,001). Un aumento de paso en este parámetro mejora las probabilidades de inferencia correcta por un factor de 1,07. En otras palabras, la variación del nombre del color se correlaciona positivamente con la precisión. Esto se traduce en una ventaja considerable para las rampas que combinan una variedad de colores distintivos. A modo de ejemplo, la diferencia de variación de nombre entre viridis y jet se sitúa en 2,03 a favor de este último. En consecuencia, la probabilidad de inferir el modelo correcto es 1,15 veces mayor con jet que con viridis, siendo iguales todos los demás factores.

5.5.3 Longitud log-LAB (post-hoc)

La variación del nombre del color puede ser engorrosa de calcular: no es compatible con la mayoría de las herramientas de diseño y requiere acceso a un modelo de nombre empírico. Por lo tanto, buscamos una métrica estrechamente relacionada que se base en espacios de color más estándar. Descubrimos que al tomar la longitud completa de la curva de un mapa de colores en el espacio LAB y transformar esa medida en registros, obtenemos una aproximación cercana a la variación de nombre. Para los mapas de color incluidos en este estudio, las dos medidas muestran una alta correlación (r de Pearson = 0,946). La Tabla 2-C enumera los coeficientes para un modelo entrenado con esta métrica. El modelo indica la longitud de log-LAB como un predictor significativo de la precisión (Z = 5,99, p < 0,001): un aumento gradual en este parámetro mejora las probabilidades de una inferencia correcta en un factor de 1,2. Por ejemplo, al comparar jet y viridis, el modelo predice probabilidades de inferencia que son 1,17 veces mayores con jet.

5.5.4 Comparación y discusión de modelos Usamos

puntajes BIC para determinar cuál de los modelos se ajusta mejor a los datos (consulte la Tabla 1). Un BIC más bajo indica un modelo más deseable basado en una compensación entre ajuste y parsimonia. Una diferencia superior a 10 proporciona una fuerte evidencia a favor del modelo de puntuación más baja [25]. los

Modelo	Parámetros	AIC	BIC	Desviación logLik
Familia de diseño	5	17209.39	17255.4	-8598.7 17197.4
Nombre del color	3	17213.24	17243.9	-8602.6 17205.2
Longitud log-LAB	3	17207.92	17238.6	-8600.0 17199.9

Tabla 1. Comparación de la bondad de ajuste de los tres modelos. Para seleccionar entre los modelos, utilizamos el criterio BIC (negrita, cuanto más bajo, mejor).

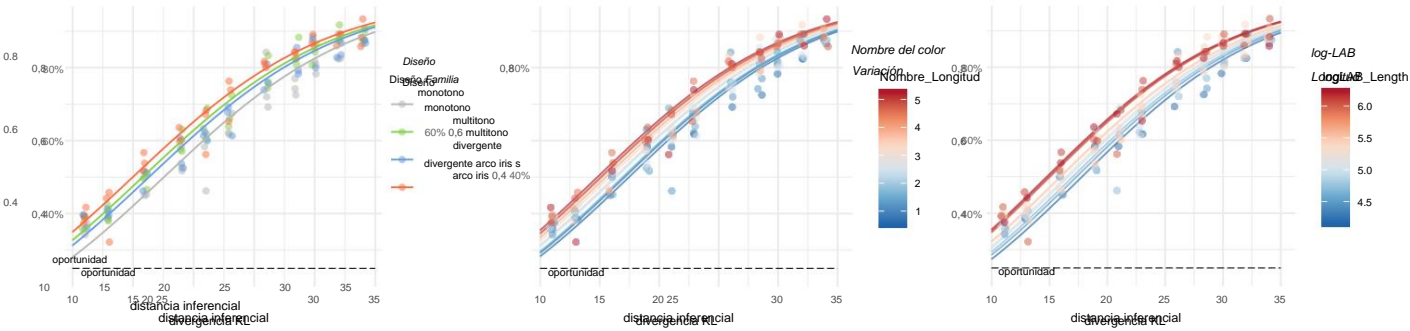


Fig. 6. Los tres modelos evaluados en el Experimento 1 y su ajuste. Las curvas muestran las predicciones del modelo. Los puntos de colores representan la precisión media del sujeto para un nivel de divergencia particular (es decir, dificultad).

Parámetro	Estimación (IC del 95 %)	valor Z	p
(Intercepción)	0,39 (0,34—0,44)	14.18	***
divergencia	1,13 (1,13—1,14)	45.33	***
Multi-tono	1,25 (1,12—1,38)	4.15	***
Divergente	1,17 (1,05—1,29)	2.92	**
arcoiris	1,38 (1,24—1,53)	6.02	***

(A) Modelo de familia de diseño

Parámetro	Estimación (IC del 95 %)	del valor Z	p
(Interceptar)	0,39 (0,35—0,45)	14.31	***
divergencia	1,13 (1,13—1,14)	45.33	***
Nombre del color Variación	1,07 (1,05—1,1)	5.53	***

(B) Color Nombre modelo

Parámetro	Estimación (IC del 95 %)	del valor Z	p
(Intercepción)	0,38 (0,33—0,43)	14.42	***
divergencia	1,13 (1,13—1,14)	45.33	***
log-LAB Longitud	1,20 (1,13—1,28)	5.99	***

(C) modelo de longitud log-LAB

Tabla 2. Parámetros para los tres modelos. Las estimaciones que se muestran corresponden a los coeficientes del modelo exponenciados para reflejar razones impares (intervalos de confianza de $\pm 95\%$). Los asteriscos indican valores de p (***=p < 0,001, **=p < 0,01)

La puntuación BIC para el modelo de nombre de color es más baja que el modelo de familia de diseño ($\bar{y}BIC=11.5$). Esta diferencia corresponde a una fuerte evidencia de que la variación del nombre es una mejor explicación de los resultados empíricos. El modelo de longitud log-LAB tiene una puntuación BIC ligeramente más baja que el modelo de nombre de color original al que se aproxima, lo que sugiere un ajuste ligeramente mejor; sin embargo, la diferencia es pequeña ($\bar{y}BIC=5.3$). La longitud del registro de un mapa de colores es, por lo tanto, una buena aproximación para la variación de su nombre y, por extensión, el rendimiento de la inferencia.

Nuestros resultados muestran que la variación de nombres y la longitud de log-LAB son mejores predictores de la utilidad del mapa de colores para la inferencia gráfica. Cualquiera de las dos métricas por sí sola parece predecir de cerca el rendimiento esperado de los participantes (consulte la Figura 5). Además, los tres modelos están de acuerdo y muestran que la precisión se ve reforzada por la incorporación de colores distintivos. La correlación entre la variación del nombre y la precisión coincide con una preferencia por rampas más coloridas (p. ej., arcoiris RGB y jet), que permitieron una mayor precisión. Estos resultados van en contra de la sabiduría del diseño convencional, que estipula que los arco iris no son adecuados para datos cuantitativos [5] y, en cambio, deben reservarse para variables categóricas [36]. Más bien, encontramos que los mapas de colores del arco iris superan significativamente a las alternativas perceptualmente más uniformes, como un solo tono (p. ej., azules) y rampas divergentes (p. ej., frío-cálido y gris-rojo). Estos resultados contrarios a la intuición se explican considerando las diferencias en la variación de nombres.

6 EXPERIMENTO 2

Los resultados del Experimento 1 indican que la variación de nombres mejora la inferencia gráfica. Los participantes parecen beneficiarse de los mapas de colores que combinan una variedad de nombres de colores. Sin embargo, las rampas más coloridas (p. ej., chorro y arco iris RGB) también pueden ser engañosas [43, 44], ya que discretan implícitamente las visualizaciones incluso cuando los datos subyacentes son continuos [39]. Los artefactos visuales resultantes pueden malinterpretarse como características de datos [5], lo que lleva a inferencias falsas. El Experimento 2 prueba la hipótesis de que los mapas de colores del arco iris aumentan la tasa de falsos descubrimientos (H2). Específicamente, estudiamos cómo la variación del nombre del color afecta los dos tipos de errores que cometen las personas en la inferencia: errores de tipo I (falsos positivos) y tipo II (falsos negativos).

6.1 Tarea

Modificamos la tarea de inferencia gráfica original (ver §3.1) para modelar una situación en la que un analista puede declarar un positivo (es decir, hay

una diferencia detectable entre un conjunto de visualizaciones) o un resultado negativo (no hay diferencia, las visualizaciones representan el mismo fenómeno). Adaptamos la tarea del Experimento 1 al convertir el formato de respuesta de opción múltiple (es decir, elegir uno de cuatro gráficos) a una respuesta binaria (es decir, estos gráficos representan el mismo modelo). Los participantes indicaron cuál de las cuatro visualizaciones en una alineación parece provenir de un modelo diferente (una inferencia positiva) o, alternativamente, declararon que las cuatro visualizaciones pertenecen al mismo modelo (una negativa). Este diseño es análogo a una elección dicotómica entre rechazar la hipótesis nula al afirmar que una de las visualizaciones en la alineación es especial o aceptar la hipótesis nula y declarar que las visualizaciones corresponden a la misma distribución a pesar de cierto ruido.

Modificamos el procedimiento de muestreo (§3.2) de modo que, en la mitad de los ensayos, las parcelas de alineación se muestrean del mismo modelo objetivo (es decir, sin señuelos). La otra mitad consiste en pruebas que son idénticas a la tarea original, con tres de cuatro visualizaciones muestreadas del modelo señuelo y una del objetivo. Para desalentar las adivinanzas aleatorias, mantenemos un formato de respuesta similar y solicitamos a los participantes que hagan clic en la imagen que "no pertenece". Sin embargo, añadimos una quinta opción etiquetada como "sin diferencia perceptible entre las imágenes". Nuestro análisis trata la respuesta como binaria: positiva si el participante selecciona una de las cuatro imágenes o negativa si declara que "no hay diferencia". Los falsos positivos ocurren cuando los participantes informan falsamente una diferencia, como los causados por las características alucinadas de las bandas del arcoiris, mientras que los falsos negativos ocurren cuando los participantes no informan las diferencias, como cuando no se detectan las características diferenciadoras del modelo.

6.2 Diseño y procedimientos experimentales

Seleccionamos un subconjunto de cuatro mapas de colores para este experimento: blues, viridis, cool-warm y RGB rainbow. Estas rampas corresponden a diferentes niveles de variación de nombres (blues-low, viridis & cool-warm-mid, RGB rainbow-high) al mismo tiempo que representan las cuatro familias de diseño. En el Experimento 1, estas rampas mostraron diferencias de rendimiento casi lineales en función de la variación del nombre (r de Pearson = 0,986), lo que proporciona una buena fuente de variación del error. Empleamos un diseño dentro de los sujetos, con todos los participantes experimentando los cuatro mapas de colores.

Los procedimientos fueron similares al Experimento 1 (ver §5.3 para más detalles). Los participantes fueron evaluados primero por deficiencia de visión del color. Luego completaron una sesión de práctica con retroalimentación seguida de cuatro bloques de ensayos analizados. El experimento fue bloqueado por mapa de colores con 24

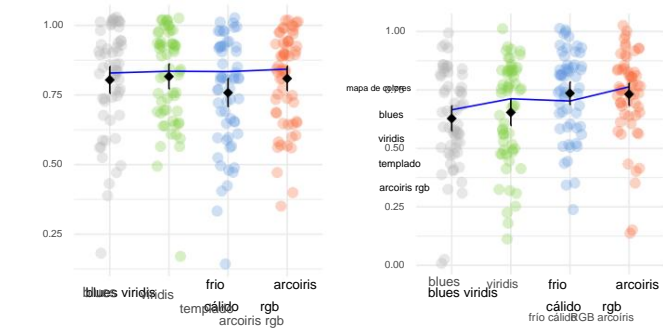


Fig. 7. Especificidad y sensibilidad por mapa de colores. Los puntos de colores representan las líneas para los participantes individuales. Los rombos representan las medias de los grupos (IC del 95 %). Las líneas azules representan las predicciones del modelo. Una línea plana (izquierda) indica que la variación del nombre del color no predice la especificidad; sin embargo, se correlaciona con la sensibilidad.

ensayos por bloque (la mitad de verdaderos positivos y la mitad de verdaderos negativos), para un total de 96 ensayos más 8 comprobaciones de compromiso. Los ensayos y el orden de los bloques fueron aleatorios. Para ayudar a aclarar las instrucciones de la tarea, les dijimos explícitamente a los participantes que la mitad de los estímulos no mostrarían una diferencia perceptible entre las imágenes. Esta información se enfatizó para reducir el posible sesgo de respuesta debido a antecedentes incorrectos (p. ej., es más probable que los participantes asuman erróneamente los estímulos negativos).

6.3 Resultados

Reclutamos a 60 participantes (edad media: 37,4 años, ETS: 10) de Amazon Mechanical Turk, y los compensamos con \$3 cada uno. Usamos los mismos criterios de exclusión que en el Experimento 1 (ver \$5.4). Los participantes completaron el experimento en 22,4 minutos en promedio (STD: 10,7). La precisión media en la tarea fue del 74,2%, con una sensibilidad del 68,8% (tasa de verdaderos positivos) y una especificidad del 79,7% (tasa de verdaderos negativos). La figura 7 traza los resultados por mapa de colores. Empleamos regresión logística para analizar por separado los resultados de especificidad y sensibilidad. Como en el Experimento 1, ajustamos dos modelos: un modelo de nombre de color y un modelo de familia de diseño. Omitimos el modelo de longitud log-LAB por limitaciones de espacio y dada su similitud con el modelo de nombre.

6.3.1 Especificidad

La variación del nombre del color no predice la especificidad (estimación exponencial: 1,03, IC: 0,95–1,12, Z de Wald = 0,732, $p = 0,464$). La estimación, centrada aproximadamente alrededor de uno, sugiere que los mapas coloridos no necesariamente aumentan las probabilidades de cometer un error falso positivo. Sin embargo, el modelo de familia de diseño muestra algunas diferencias de rendimiento: la rampa fría-cálida divergente mostró una especificidad significativamente menor (ee: 0,74, IC: 0,57–0,97, Z = 2,212, $p < 0,05$). Las comparaciones por pares con el ajuste de Tukey muestran que frío-tibio es peor que viridis (Z = 2,857, $p < 0,05$). Todas las demás comparaciones no fueron significativas.

6.3.2 Sensibilidad La

La variación del nombre del color predice significativamente la sensibilidad (ee: 1,15, IC: 1,07–1,24, Z = 3,951, $p < 0,001$): un aumento gradual en la distancia del nombre aumenta las probabilidades de resolver los verdaderos positivos en 1,15 veces. De manera similar, el modelo de familia de diseño indica una preferencia por el arco iris y los diseños divergentes: tanto el arco iris RGB como el frío-cálido generaron una mayor sensibilidad que viridis (Zr = 3,451, $pr < 0,01$; Zcw = 3,579, $pcw < 0,01$) y blues (Zr = 4,428, $pr < 0,001$; Zcw = 4,554, $pcw < 0,001$).

En resumen, los resultados no proporcionan evidencia de que los mapas más coloridos (p. ej., arcoiris RGB) aumenten la tasa de falsos positivos. En cambio, encontramos evidencia que confirma que la variación del nombre respalda la inferencia al aumentar la tasa positiva verdadera. La mayor sensibilidad para el arcoiris RGB, combinada con su especificidad de línea de base, puede explicar por qué este mapa de colores funcionó tan bien en el Experimento 1. Las personas parecen discriminar mejor entre modelos cuando ven un mapa de arcoiris sin ser necesariamente engañados para ver diferencias extrínsecas del modelo falsas.

7 DISCUSIÓN

El color es una de las propiedades visuales más utilizadas para comunicar datos cuantitativos. Sin embargo, todavía estamos desarrollando un entendimiento gráfico de cómo funciona el color en la visualización. Buscamos comprender cómo las características del mapa de colores afectan la inferencia gráfica. Presumimos que los determinantes cognitivos, como la capacidad de nombrar y pensar categóricamente sobre los colores, afectan la capacidad de las personas para extraer inferencias sobre los datos. Además, planteamos la hipótesis de que la precisión inferencial se puede mejorar mediante el uso de mapas de colores que cruzan una variedad de nombres de colores. Para cuantificar esto último, propusimos una nueva métrica, la variación del nombre del color, y estudiamos su impacto en dos experimentos colaborativos.

7.1 Inferencia por color Primero

buscamos modelar el desempeño de los participantes usando la variación del nombre de un mapa de colores como factor explicativo. Los resultados muestran que la variación de nombres es un buen predictor de la utilidad del mapa de colores para la inferencia gráfica. La incorporación de una gama más amplia de colores identificables aumenta significativamente la precisión. Esta correlación positiva se puede ver en la Figura 5-izquierda.

Estos resultados apoyan **H1**.

Nuestros resultados están en desacuerdo con las pautas convencionales, que sugieren que los diseñadores deben limitar los colores a una selección juiciosa de tonos [32] mientras enfatizan la variación de luminancia y la uniformidad perceptiva [26, 43, 62]. En cambio, descubrimos que los mapas de colores de mayor rendimiento eran los arco iris (arco iris RGB, chorro y turbo, en ese orden). Estos hallazgos sugieren que las tareas interpretativas más complejas pueden depender de las características cognitivas de los mapas de colores, en lugar de su apariencia perceptiva.

Si bien nuestros resultados entran en conflicto con las pautas convencionales [51], los hallazgos contrarios pueden explicarse en gran medida por nuestra métrica de variación de nombre. Los mapas de colores del arco iris, por ejemplo, cruzan una amplia gama de colores fácilmente identificables (p. ej., azul, rojo, naranja y amarillo). Esta mezcla crea la apariencia de "bandas" que son fácilmente distinguibles [39].

Que se cree que este tipo de discretización del color es problemático para los datos cuantitativos [5], especulamos que el sistema visual puede aprovechar estas características discretas emergentes. Por ejemplo, un observador puede usar heurísticamente el tamaño aparente, la aritmética y la distribución de parches de color como proxy de varias propiedades estadísticas en los datos (por ejemplo, la varianza [42]). Las personas pueden estimar de manera confiable estas estadísticas con un vistazo rápido, probablemente debido a los procesos de visión de conjunto de acción rápida [1, 47, 55]. Tales estadísticas visuales pueden proporcionar funciones de resumen clave que ayuden a las personas a discriminar entre modelos. Sin embargo, algunos procesos de conjunto operan únicamente en características visuales segmentadas [16], lo que puede explicar la ventaja de discretizar mapas de color como el arco iris. Nuestra métrica de variación del nombre del color parece modelar esta tendencia. Por el contrario, los mapas de colores más suaves pueden complicar la visión del conjunto y, potencialmente, dejar de lado características clave que discriminan el modelo. Además de facilitar la heurística basada en conjuntos, la categorización por colores puede ayudar a los observadores a codificar características de datos específicos en sus memorias visuales o verbales de trabajo [24]. Estas representaciones mentales, a su vez, podrían usarse para comparar visualizaciones, lo que permitiría una evaluación más completa de diferentes modelos. Se necesita trabajo futuro para probar estos conjuntos visuales y las explicaciones de la memoria de trabajo.

La variación de nombre también explica las discrepancias de rendimiento que observamos post-hoc. Por ejemplo, las características de diseño convencionales no explican bien la diferencia entre viridis y plasma. Los dos mapas de color tienen propiedades perceptuales virtualmente idénticas: ambos presentan rampas secuenciales, multitonos y perceptualmente uniformes derivadas con interpolación spline en el espacio CAM02 [57]. Sin embargo, los resultados muestran que el plasma es más preciso (68,6 %, IC: 65,4–71,9) que viridis (66,4 %, IC: 62–70,8). Esta brecha de rendimiento no se puede explicar con las características de diseño tradicionales. Sin embargo, el modelo de nombre de color distingue fácilmente entre estas dos rampas: parece que el plasma combina más colores nombrables (3.44) que viridis (2.75). Intuitivamente, observamos tonos de azul, púrpura, naranja y amarillo en plasma, en comparación con un conjunto de tonos ligeramente más limitado en viridis (azul marino, verde y amarillo). En consecuencia, y de acuerdo con los datos observados, el modelo de nombre predice una precisión de inferencia ligeramente mejor con plasma.

Los criterios BIC proporcionan pruebas estadísticamente sólidas de una correlación entre el rendimiento y la variación del nombre (§5.5.4). La comparación

contra un modelo de familia de diseño convencional sugiere que las propiedades de diseño tradicionales, como si un mapa de colores es secuencial, divergente o del tipo 'arcoíris', pueden no proporcionar una forma óptima de seleccionar mapas de colores. La inferencia gráfica parece menos dependiente de estas características de diseño. En cambio, la variación del nombre por sí sola puede ser suficiente para medir la utilidad del mapa de colores para la inferencia gráfica.

7.2 Categorización y especificidad del color Los resultados

del Experimento 1 indican que la categorización del color puede revelar variaciones sutiles en los datos. Sin embargo, la categorización podría inducir a error a los espectadores y hacer que malinterpreten los límites de color como si fueran características de datos [5] (**H2**). Esta mala interpretación podría hacer que un analista declarara falsamente diferencias entre visualizaciones que, de otro modo, se tomarían como muestra del mismo modelo o fenómeno. Alternativamente, un analista podría concluir erróneamente una discrepancia entre un modelo anterior supuesto y un modelo representado por una visualización observada [17]. Sin embargo, contrariamente a esta hipótesis, el Experimento 2 no encontró relación entre la variación del nombre y la especificidad ($p = 0,464$). Los participantes no tenían más probabilidades de informar diferencias falsas mientras veían mapas de colores con una mayor variación de nombres. No encontramos evidencia de que los mapas de colores no uniformes (p. ej., arcoíris RGB) hagan que las personas sean más susceptibles a los falsos positivos en comparación con las rampas perceptivamente uniformes (p. ej., blues, viridis, coolwarm).

Sin embargo, encontramos una asociación positiva entre la variación del nombre y la sensibilidad (es decir, la tasa de verdaderos positivos), consistente con el Experimento 1.

Los hallazgos de los dos experimentos podrían explicar por qué el arcoíris y otros mapas de color que varían en tono siguen siendo de uso generalizado a pesar de las limitaciones conocidas [35]. Nuestros resultados sugieren que la popularidad de estos esquemas puede deberse a una mayor utilidad para la inferencia gráfica en lugar de preferencias puramente estéticas. Esta observación conduce a una línea guía poco ortodoxa: los diseñadores pueden querer maximizar la gama de colores identificables en una rampa. El modelo basado en LAB sugiere una heurística aún más simple: maximizar la longitud de la curva transformada logarítmicamente de un mapa de colores. Si bien el razonamiento en log-LAB en lugar de LAB puede parecer contradictorio, nuestro análisis indica que la métrica anterior se aproxima mucho a la variación del nombre (r de Pearson = 0,95), lo que proporciona un vínculo entre la apariencia del color y la distancia del nombre [20].

Los esquemas de arco iris no son la única familia de diseño con una amplia variación de nombres. Ciertos mapas de colores de varios tonos, como el plasma, parecen lograr un equilibrio entre la uniformidad perceptiva y la variación de nombres. Las inferencias con plasma tuvieron una precisión superior a la media (68,6 %), aunque ligeramente por debajo del arco iris RGB (70,3 %, IC: 66,6–74,0). En consecuencia, los diseñadores pueden optar por los mapas más coloridos mientras continúan equilibrando otras restricciones de diseño, como la monotonicidad de la luminancia o la uniformidad de la percepción. Aunque aparentemente menos relevantes para la inferencia gráfica, estas propiedades aún juegan un papel clave en muchas tareas, incluida la percepción de formas [43,59]. El trabajo futuro podría investigar cómo optimizar estas restricciones de diseño en competencia, por ejemplo, mediante la síntesis computacional de nuevos mapas de colores bajo demanda (como en Colorgorical [18]). Las dos métricas propuestas en este trabajo (variación del nombre del color y longitud del registro LAB) podrían ayudar en este proceso al proporcionar modelos predictivos de la utilidad del mapa de colores.

8 LIMITACIONES E INVESTIGACIÓN FUTURA

Este estudio proporciona una primera investigación empírica sobre cómo el diseño del mapa de colores afecta la inferencia gráfica. Sin embargo, nuestro enfoque proporciona solo una perspectiva preliminar de esta pregunta y está sujeto a varias limitaciones. Vemos estas limitaciones como oportunidades para el trabajo futuro.

Aunque diversificamos deliberadamente el conjunto de estímulos mediante el uso de una variedad de distribuciones, todos los estímulos finalmente consistieron en una mezcla de gaussianas. Estos modelos se aproximan a una variedad de fenómenos en ciencia e ingeniería, pero no pueden capturar todos los tipos de datos y características. Para explorar más a fondo la inferencia gráfica en otros contextos, los procedimientos empleados en este estudio pueden extenderse a tareas en dominios específicos. Por ejemplo, para evaluar la efectividad del mapa de colores para la visualización de flujo, un experimento podría usar modelos de flujo, aplicando perturbaciones medidas a remolinos y vórtices para identificar los mapas de colores más discriminatorios. Para la astrofotografía en falso color, un experimento podría probar modelos de estructuras galácticas con perturbaciones en la espiralidad o la homogeneidad de las galaxias.

Nuestro estudio limita la inferencia a una clase específica de tareas conocidas en la literatura como "inferencia gráfica" [10, 63]. Los participantes evaluaron las visualizaciones de conjunto de manera holística y emitieron juicios resumidos.

Esta tarea evalúa las interpretaciones generales de los espectadores de un modelo y su capacidad para discriminar entre diferentes modelos. Sin embargo, la tarea no prueba necesariamente si las personas pueden encontrar diferencias sutiles en la imagen. El término 'inferencia' puede abarcar una amplia gama de actividades cognitivas con visualizaciones (p. ej., razonamiento inductivo y prueba de hipótesis), mientras que solo consideramos una interpretación específica de este término.

Aunque nuestro estudio muestra una clara ventaja para las rampas de colores (p. ej., arcoíris), es posible que estas rampas no sean accesibles para las personas con deficiencias en la visión del color. Por lo tanto, los diseñadores deben equilibrar la variación de nombres con métricas de color más accesibles. Una de las rampas que probamos, plasma, parece combinar dos características accesibles (variación de nombre relativamente alta con luminancia monotónica), lo que la convierte en una fuerte candidata para la visualización accesible.

Si bien los resultados del Experimento 2 muestran que los mapas coloridos no necesariamente disminuyen la especificidad de la inferencia, las personas aún pueden percibir artefactos de los mapas de arcoíris en otros contextos (p. ej., al estudiar una sola imagen o al mirar varias imágenes con distribuciones muy diferentes). Comprender cuándo surgen tales artefactos y cómo apoyan o dificultan el análisis es un trabajo futuro clave [39].

Por último, nuestro trabajo considera la variación del nombre del color como la principal métrica cognitiva para modelar el rendimiento en la inferencia gráfica. Sin embargo, encontramos que los resultados también se correlacionan con la longitud de log-LAB. Esta última métrica, que puede proporcionar una medida del poder discriminatorio perceptivo, ofrece una explicación alternativa para nuestros resultados observados.

9 CONCLUSIÓN

La codificación de colores es una preocupación fundamental para la visualización de datos. Los mapas de colores efectivos no solo deben ayudar a las personas a percibir características de datos de bajo nivel, sino que también deben facilitar una interpretación e inferencia precisas. Teorizamos que el desempeño en estas tareas está correlacionado con las propiedades cognitivas de los colores. Probamos esta teoría en dos experimentos colaborativos que midieron la capacidad de los participantes para discriminar entre modelos escalares 2D en función del diseño del mapa de colores. Descubrimos que la inferencia gráfica está estrechamente relacionada con la variedad de colores identificables en el mapa. Las rampas de colores, como el jet y el arcoíris RGB, condujeron a inferencias significativamente más precisas que las rampas que comprenden una selección limitada de tonos. En un segundo experimento, medimos la tasa de falsos descubrimientos pero no encontramos evidencia de una relación entre la variación del nombre del color y la tasa de falsos positivos. Nuestros resultados resaltan la necesidad de nuevas pautas de codificación de colores basadas en factores cognitivos en lugar de puramente perceptuales. Propusimos dos nuevas métricas para modelar la utilidad del mapa de colores en la inferencia visual. Estas métricas proporcionan principios generativos para nuevos diseños de mapas de colores y, al mismo tiempo, amplían los principios convencionales sobre el uso del color en la visualización.

10 RESPUESTAS PARA LA FIGURA 1

File superior (comenzando con la alineación más a la izquierda): cuadrante superior derecho, inferior izquierdo, inferior derecho, superior izquierdo.

File inferior: arriba a la izquierda, abajo a la izquierda, abajo a la derecha, abajo a la derecha.

AGRADECIMIENTOS

Este documento se basa en una investigación respaldada por los premios 1942429, 1764092 y 1657599 de la Fundación Nacional de Ciencias.

Instalación para usuarios de la Oficina de Ciencias del Departamento de Energía operada bajo contrato DE-AC02-06CH11357.

REFERENCIAS

- [1] D. Ariely. Ver conjuntos: Representación por propiedades estadísticas. *Psicópata Ciencia lógica*, 12(2):157–162, 2001.
- [2] LD Bergman, BE Rogowitz y LA Treinish. Una herramienta basada en reglas para asistir en la selección de mapas de colores. En *Actas de la 6ª conferencia sobre Visualización'95*, p. 118. Sociedad de Informática del IEEE, 1995.
- [3] KP Berlin B. Términos de color básicos: Su universalidad y evolución.- Berkeley, 2002.

- [4] M. Borkin, K. Gajos, A. Peters, D. Mitsouras, S. Melchionna, F. Rybicki, C. Feldman y H. Pfister. Evaluación de visualizaciones arteriales para el diagnóstico de enfermedades del corazón. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2479–2488, 2011.
- [5] D. Borland y RMT II. El mapa de colores del arcoíris (todavía) se considera dañino. *Aplicaciones y gráficos por computadora de IEEE*, 27(2), 2007.
- [6] Cervecero CA. Pautas para seleccionar colores para esquemas divergentes en mapas. *The Cartographic Journal*, 33(2):79–86, 1996.
- [7] Cervecero CA. Esquemas espectrales: uso de colores controvertidos en mapas. *Cartography and Geographic Information Systems*, 24(4):203–220, 1997.
- [8] CA Brewer, AM MacEachren, LW Pickle y D. Herrmann. Mortalidad de ping de mapas : evaluación de esquemas de color para mapas de coropletas. *Anales de la Asociación de Geógrafos Estadounidenses*, 87(3):411–438, 1997.
- [9] AM Brown, DT Lindsey y KM Guckes. Nombres de color, categorías de color y búsqueda visual con claves de color: a veces, la percepción del color no es categórica. *Revista de visión*, 11(12):2–2, 2011.
- [10] A. Buja, D. Cook, H. Hofmann, M. Lawrence, E.-K. Lee, DF Swayne y H. Wickham. Inferencia estadística para el análisis exploratorio de datos y el diagnóstico de modelos. *Transacciones filosóficas de la Royal Society A: Ciencias matemáticas, físicas y de ingeniería*, 367 (1906): 4361–4383, 2009.
- [11] R. Bujack, TL Turton, F. Samsel, C. Ware, DH Rogers y J. Ahrens. Lo bueno, lo malo y lo feo: un marco teórico para la evaluación de mapas de color continuos. *Transacciones IEEE sobre visualización y gráficos por computadora*, 2017.
- [12] NR Chowdhury, D. Cook, H. Hofmann y M. Majumder. Medir la dificultad de la alineación al hacer coincidir las métricas de distancia con las opciones de tema en datos de fuentes múltiples. *Journal of Computational and Graphical Statistics*, 27(1):132–145, 2018.
- [13] A. Dasgupta, J. Poco, B. Rogowitz, K. Han, E. Bertini y CT Silva. El efecto de las escalas de color en el desempeño objetivo y subjetivo de los científicos del clima en tareas de análisis de datos espaciales. *Transacciones IEEE sobre visualización y gráficos por computadora*, 2018.
- [14] R. De Valois y K. De Valois. *Visión espacial*. Prensa de la Universidad de Oxford, 1990.
- [15] G. Derefeldt, T. Swartling, U. Berggrund y P. Bodrogi. Color cognitivo. *Investigación y aplicación del color*, 29(1):7–19, 2004.
- [16] SL Franconeri, DK Bemis y GA Alvarez. La estimación de números se basa en un conjunto de objetos segmentados. *Cognición*, 113(1):1–13, 2009.
- [17] A. Gelman. Una formulación bayesiana de análisis exploratorio de datos y pruebas de bondad de ajuste. *Revista Estadística Internacional*, 71(2):369–382, 2003.
- [18] CC Gramazio, DH Laidlaw y KB Schloss. Colorgical: Creación de paletas de colores discriminables y preferibles para la visualización de información. *Transacciones IEEE sobre visualización y gráficos por computadora*, 23(1):521–530, 2016.
- [19] M. Harrower y CA Brewer. Cervecero de colores. org: una herramienta en línea para seleccionar esquemas de color para mapas. *The Cartographic Journal*, 40(1):27–37, 2003.
- [20] J. Heer y M. Stone. Modelos de nombres de colores para selección de colores, edición de imágenes y diseño de paletas. En *Actas de la Conferencia SIGCHI sobre factores humanos en sistemas informáticos*, págs. 1007–1016, 2012.
- [21] JR Hershey y PA Olsen. Aproximación de la divergencia kullback leibler entre modelos de mezclas gaussianas. En 2007, Conferencia Internacional IEEE sobre Acústica, Procesamiento de Señales y Habla-ICASSP'07, vol. 4, págs. IV–317. IEEE, 2007.
- [22] H. Hofmann, L. Follett, M. Majumder y D. Cook. Pruebas gráficas para la comparación de potencia de diseños de la competencia. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2441–2448, 2012.
- [23] MD Hyslop. Una comparación del color espectral y la percepción del mapa de tonos continuos en escala de grises . Tesis de maestría, Universidad Estatal de Michigan, 2006.
- [24] T. Ikeda y N. Osaka. ¿Cómo se memorizan los colores en la memoria de trabajo? un estudio de resonancia magnética funcional. *Neuroinforme*, 18(2):111–114, 2007.
- [25] BL Jones, DS Nagin y K. Roeder. Un procedimiento sas basado en modelos mixtos para estimar trayectorias de desarrollo. *Métodos sociológicos e investigación*, 29(3):374–393, 2001.
- [26] AD Kalvin, BE Rogowitz, A. Pelah y A. Cohen. Creación de mapas de colores perceptuales para visualizar datos de intervalos. En *Visión humana e imágenes electrónicas V*, vol. 3959, págs. 323–336. Sociedad Internacional de Óptica y Fotónica, 2000.
- [27] J. Kim, DJ Bodony y JB Freund. Control basado en adjuntos de eventos ruidosos en un chorro turbulento. *Revista de mecánica de fluidos*, 741: 28–59, 2014.
- [28] G. Kindlmann y C. Scheidegger. Un proceso algebraico para el diseño de visualizaciones. *Transacciones IEEE sobre visualización y gráficos por computadora*, 20(12):2181–2190, 2014.
- [29] E. Komatsu, K. Smith, J. Dunkley, C. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Nolte, L. Page, et al. Observaciones de la sonda de anisotropía de microondas wilkinson (wmap) de siete años: interpretación cosmológica. *Serie de suplementos de The Astrophysical Journal*, 192(2):18, 2011.
- [30] MP Kumlir y RE Group. Mapeo de tonos continuos de superficies lisas. *Cartografía y Sistemas de Información Geográfica*, 17(4):279–289, 1990.
- [31] S. Lin, J. Fortuna, C. Kulkarni, M. Stone y J. Heer. Selección de colores semánticamente resonantes para la visualización de datos. En *Computer Graphics Forum*, vol. 32, págs. 401–410. Biblioteca en línea de Wiley, 2013.
- [32] Y. Liu y J. Heer. En algún lugar sobre el arco iris: una evaluación empírica de mapas de color cuantitativos. En *Actas de la Conferencia de CHI de 2018 sobre factores humanos en sistemas informáticos*, pág. 598. ACM, 2018.
- [33] A. Mijailov. Turbo, un mapa de colores del arcoíris mejorado para la visualización. <https://ai.googleblog.com/2019/08/turbo-improved-sequential-color-map.html> (2019, 2020).
- [34] K. Moreland. Mapas de colores divergentes para visualización científica. En *Simposio internacional sobre computación visual*, págs. 92–103. Springer, 2009.
- [35] K. Moreland. Por qué usamos mapas de colores defectuosos y qué puede hacer al respecto. *Imágenes electrónicas*, 2016(16):1–6, 2016.
- [36] T. Münzner. *Análisis y diseño de visualizaciones*. Prensa CRC, 2014.
- [37] L. Padilla, PS Quinan, M. Meyer y SH Creem-Regehr. Evaluación del impacto de agrupar campos escalares 2d. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):431–440, 2017.
- [38] Salmuera LW. Pruebas de usabilidad de diseños de mapas. En *Actas del Simposio sobre la Interfaz de Ciencias de la Computación y Estadística*, págs. 42–56, 2003.
- [39] PS Quinan, L. Padilla, SH Creem-Regehr y M. Meyer. Examinando la discretización implícita en esquemas espectrales. En *Computer Graphics Forum*, vol. 38, págs. 363–374. Biblioteca en línea de Wiley, 2019.
- [40] RM Ratwani, JG Trafton y DA Boehm-Davis. Pensar gráficamente: conectando la visión y la cognición durante la comprensión gráfica. *Revista de Psicología Experimental: Aplicada*, 14(1):36, 2008.
- [41] K. Reda, P. Nalawade y K. Ansah-Koi. Percepción gráfica de mapas cuantitativos continuos: los efectos de la frecuencia espacial y el diseño del mapa de colores. En *Actas de la Conferencia de CHI de 2018 sobre factores humanos en sistemas informáticos*, pág. 272. MCA, 2018.
- [42] K. Reda y ME Papka. Evaluación de la percepción del gradiente en campos escalares codificados por colores . En 2019 IEEE Visualization Conference (VIS), págs. 271–275. IEEE, 2019.
- [43] BE Rogowitz y AD Kalvin. El "proyecto de qué blair": un método visual rápido para evaluar mapas de color perceptivos. En *Visualización*, 2001. VIS'01. Actas, págs. 183–556. IEEE, 2001.
- [44] BE Rogowitz y LA Treinish. Uso de reglas perceptivas en visualización interactiva. En *IS&T/SPIE 1994 Simposio internacional sobre imágenes electrónicas: ciencia y tecnología*, págs. 287–295. Sociedad Internacional de Óptica y Fotónica, 1994.
- [45] BE Rogowitz y LA Treinish. Visualización de datos: el final del arcoíris. *Espectro IEEE*, 35(12):52–59, 1998.
- [46] BE Rogowitz, LA Treinish, S. Bryson, et al. Cómo no mentir con la visualización. *Computadoras en Física*, 10(3):268–273, 1996.
- [47] J. Ross y DC Burr. La visión detecta el número directamente. *Revista de visión*, 10(2):10–10, 2010.
- [48] KB Schloss y SE Palmer. Respuesta estética a las combinaciones de colores: preferencia, armonía y similitud. *Atención, Percepción y Psicofísica*, 73(2):551–571, 2011.
- [49] V. Setlur y MC Stone. Un enfoque lingüístico de la asignación de colores categóricos para la visualización de datos. *Transacciones IEEE sobre visualización y gráficos por computadora*, 22(1):698–707, 2015.
- [50] HM Shapiro. *Citometría de flujo práctica*. John Wiley & Sons, 2005.
- [51] S. Silva, BS Santos y J. Madeira. Uso del color en la visualización: una encuesta. *Computadoras y gráficos*, 35(2):320–333, 2011.
- [52] S. Smart y DA Szafir. Medir la separabilidad de forma, tamaño y color en diagramas de dispersión. En *Actas de la Conferencia de CHI de 2019 sobre factores humanos en sistemas informáticos*, págs. 1 a 14, 2019.
- [53] S. Smart, K. Wu y DA Szafir. Color crafting: Automatización de la construcción de rampas de color de calidad de diseñador. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1215–1225, 2019.

© 2020 IEEE. Esta es la versión del autor del artículo que se ha publicado en IEEE Transactions on Visualization and Computer Graphics. La versión final de este registro está disponible en: [xx.xxxx/TVCG.201x.xxxxxxx/](https://doi.org/10.1109/TVCG.2020.3000000)

- [54] DA Szafir. Modelado de la diferencia de color para el diseño de visualización. Transacciones IEEE sobre visualización y gráficos por computadora, 2017.
- [55] DA Szafir, S. Haroz, M. Gleicher y S. Franconeri. Cuatro tipos de codificación de conjunto en visualizaciones de datos. Revista de visión, 16(5):11–11, 2016.
- [56] C. Tominski, G. Fuchs y H. Schumann. Codificación de colores basada en tareas. En Visualización de Información, 2008. IV'08. XII Congreso Internacional, págs. 373–380. IEEE, 2008.
- [57] S. van der Walt y N. Smith. Mapas de colores de Matplotlib. <https://ofertas.github.io/colormap/>, 2015. [En línea; consultado el 20 de abril de 2020].
- [58] W3C. Módulo de unidades y valores CSS nivel 3. <http://www.w3.org/TR/css3-values/#absolute-lengths>, 2016. [En línea; consultado el 20 de abril de 2020].
- [59] C. Ware. Secuencias de colores para mapas univariados: Teoría, experimentos y principios. Aplicaciones y gráficos por computadora del IEEE, 8(5):41–49, 1988.
- [60] C. Artículos. Pensamiento visual: Para el diseño. Morgan Kaufmann, 2010.
- [61] C. Artículos. Visualización de Información: Percepción para el Diseño. Elsevier, 2012.
- [62] C. Ware, TL Turton, R. Bujack, F. Samsel, P. Shrivastava y DH Rogers. Medir y modelar las funciones de umbral de detección de características de los mapas de colores. Transacciones IEEE sobre visualización y gráficos por computadora, 2018.
- [63] H. Wickham, D. Cook, H. Hofmann y A. Buja. Inferencia gráfica para infovis. IEEE Transactions on Visualization and Computer Graphics, 16(6):973–979, 2010.
- [64] HM Widen, JB Elsner, S. Pau y CK Uejo. Inferencia gráfica en la investigación geográfica. Análisis geográfico, 48(2):115–131, 2016.
- [65] M. Wijffelaars, R. Vliegen, JJ Van Wijk y E.-J. Van Der Linden. Generación de paletas de colores mediante parámetros intuitivos. En Computer Graphics Forum, vol. 27, págs. 743–750. Biblioteca en línea de Wiley, 2008.
- [66] DN Williams. Herramientas de visualización y análisis de datos climáticos a ultraescala. Eos, Transacciones de la Unión Geofísica Americana, 95(42):377–378, 2014.
- [67] J. Winawer, N. Witthoft, MC Frank, L. Wu, AR Wade y L. Borodit sky. Los blues rusos revelan los efectos del lenguaje en la discriminación del color. Actas de la academia nacional de ciencias, 104(19):7780–7785, 2007.
- [68] L. Zhou y CD Hansen. Una encuesta de mapas de color en la visualización. IEEE Transactions on Visualization and Computer Graphics, 22(8):2051–2069, 2016.