

Cápsula 1: Principios de expresividad

Hola, bienvenidxs a una cápsula del curso Visualización de Información. Esta cápsula corresponde a la primera parte de principios de diseño en visualización: principios de expresividad.

Los principios que comenzaremos a listar en esta cápsula y continuaremos en las siguientes, son **consejos o guías** de cómo estructurar y diseñar visualizaciones. Estos provienen del **conocimiento teórico y práctico colectivo en Visualización de Información**. No deben interpretarse como reglas fijadas en piedra, ya que **pueden existir excepciones** para cada principio, y no corresponde utilizarlas todas en todo momento.

Ya hemos mencionado varios principios, sobretodo de percepción, color y uso de canales. Uno de ellos era el principio de expresividad, que dicta que una visualización debe expresar todos los datos, y solamente, los datos. Con ella sacamos estas guías de emparejar canales y atributos en base a tipos.

En esta cápsula detallaremos otros dos principios que se extraen de la idea de expresividad: el *data-ink ratio*, y el *lie factor*. Ambos fueron propuestos por **Edward Tufte**, pionero de Visualización de Información.

El primero es la idea de estimar cuál es el ***data-ink ratio*** de una visualización. Este es una relación matemática, que ve la proporción entre **la tinta utilizada para codificar aspectos de los datos en la visualización**; y **la cantidad de tinta total** utilizada en dicha visualización. Por tinta, se refiere a las áreas y espacios visibles que la visualización ocupa. En nuestro caso, podemos pensarlo como la **cantidad de píxeles utilizados, en datos y visualización**.

Lo que dicta este principio es que **el *data-ink ratio* debe ser lo más cercano a 1 como podamos**. Eso se traduce en que todos los elementos gráficos se ocupan para representar datos y nada más. Es decir, busca que cada píxel utilizado, o al menos la mayoría, esté justificado y **tenga una razón** de estar ahí.

A modo de ejemplo, podemos analizar el ejemplo en pantalla. Es un gráfico de barras que por alguna razón se muestran como bloques tridimensionales. Como el hecho de que sean tridimensionales no aportan a la transmisión de los datos, el *data-ink ratio* se puede considerar bajo. Si removemos ese efecto, las barras ocupan menos espacio para transmitir lo mismo, por lo que aumenta el *data-ink ratio*.

De modo más extremo, esta idea nos deja un modelo crítico que nos permite **cuestionar elementalmente todos los aspectos visuales de una visualización**. Por ejemplo, podríamos partir con la visualización en pantalla e ir removiendo elementos visuales con el fin de aumentar el *data-ink ratio*.

Primero, el color de fondo no cumple ninguna función aparente. Perfectamente podemos removerlo y dejarlo como un fondo blanco. Segundo, la grilla que encierra todo es innecesaria. Las líneas verticales son redundantes porque la posición de las etiquetas ayudan a identificar correspondencia, así que también podemos removerlas.

Las líneas horizontales son más discutibles, ya que permiten identificar los pasos de alcance de cada barra en relación al eje Y. Una opción es usar líneas punteadas y colocarlas detrás, a modo de sacarles énfasis sobre las barras. Pero incluso podríamos darlo vuelta y solo realizar cortes en las barras que efectivamente tienen intersección.

El segundo principio mencionado, es el de **lie factor**. De forma similar al anterior, esta es una relación matemática, pero ahora es la relación entre **el efecto o diferencias de magnitud en los datos y los efectos de magnitud en la visualización**.

Siempre se debe apuntar a un lie factor de 1, es decir, que los efectos de magnitud en la visualización sean **fieles** a los efectos de magnitud en los datos.

Por ejemplo, en pantalla se codifican ciertos valores cuantitativos mediante área. Si revisamos la relación de magnitudes de los datos mostrados y las áreas respectivas, podemos apreciar que no hay una correspondencia 1 a 1. El problema específico aquí es que se usó el radio de las áreas para codificar el atributo, no el área directamente. Se modifica para que la correspondencia sea correcta, obtenemos un *lie factor* de 1.

Bajo este principio cae **el buen uso de ejes**. Lamentablemente es muy común encontrarnos con visualizaciones que usan de forma engañosa los ejes, lo cual **afecta al lie factor y lo aleja de 1**.

El caso más común es con gráficos de barra. Si se ocupa una barra para codificar un atributo, entonces perceptualmente utilizamos el largo de cada barra para hacer las relaciones de magnitudes. **Si el eje no está escogido desde cero, entonces las magnitudes de barras visibles no corresponden a los atributos que representan**.

¿Qué otras situaciones te has encontrado donde el lie factor es muy distinto de 1? Déjalos en los comentarios del video.

Con eso termina el contenido de esta cápsula. Recuerda que si tienes preguntas, puedes dejarlas en los comentarios del video para responderlas en la sesión en vivo de esta temática. ¡Chao!