

Cápsula 4: Abstracción de datos y derivaciones

Hola, bienvenidxs a una cápsula del curso Visualización de Información. En esta hablaré sobre la abstracción de datos y derivaciones de datos.

Como recordarán del modelo anidado que utilizamos, el segundo nivel busca abstraer la información recopilada del contexto de aplicación (el primer nivel) para traerse a un mismo campo de descripción. Una de esas abstracciones es la de datos, que responde a la primera pregunta de la triada: **qué** datos ve un usuario de una herramienta de visualización.

La gran cantidad de formas de categorizar los datos siguen esta idea. Que nos cuestionemos cómo se comportan los datos singulares, el detalle de los atributos, y el conjunto de datos completo. **¡Ya que estas descripciones nos indicarán mejores formas de representarlas visualmente!**

Algunos adelantos de ejemplo son: usar una gráfico de barras para codificar un atributo categórico no tiene sentido; usar tonalidad de colores no es tan efectivo al querer representar valores cuantitativos divergentes; un *dataset* de red se muestra de forma completamente distinta que un *dataset* geométrico; etc...

A su vez, es importante también considerar la posibilidad de transformar los datos entregados a otra forma, y realizar su abstracción una vez transformado. A la generación de nuevos datos o información desde un conjunto inicial se le llama **derivación de datos**. Y es un proceso importante a considerar a la hora de analizar el qué en el contexto de aplicación.

Por ejemplo, un conjunto de datos podría contener un atributo que describe la temperatura promedio en distintas localidades del planeta. Sus valores básicos y crudos reflejan a un atributo cuantitativo, al ser números reales en escala celsius o fahrenheit. Pero si consideramos más el contexto de uso de estos datos, puede que nos encontremos que los valores exactos no son importantes, y solo es necesario diferenciar entre frío, cálido y caluroso. Al transformar la temperatura a este nueva escala, nos encontramos con que es un atributo ordinal. Más extremo puede ser si en realidad se busca determinar si la temperatura es peligrosa para salir o no. En esa nueva escala se convierte en un atributo categórico.

El esfuerzo de traer el contexto a un nivel abstracto busca cuestionarnos su significado en la vida real y en su aplicación, y usar eso a la hora de armar una visualización de datos. **No es un paso fácil, ya que la vida real es complicada.** Los enlaces pueden tener atributos, los atributos pueden tener órdenes y múltiples categorías y los datos pueden transformarse de tipo.

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!