

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 (del 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 en una herramienta de visualización.

La gran cantidad de formas de categorizar los datos siguen esta idea. Que nos cuestionemos cuál es la semántica de los datos singulares, cómo entendemos el detalle de los atributos individuales, y cómo se comporta el conjunto de datos completo. **Estas descripciones nos indicarán mejores formas de representarlas visualmente.**

Algunos ejemplos de eso son ideas como las siguientes: usar una gráfico de barras para codificar un atributo categórico no tiene mucho sentido: el largo de barras tienen una connotación ordenada y los atributos categóricos no; usar matices de colores no es tan efectivo al querer representar valores cuantitativos divergentes, ya que estos tienen un orden muy específico implícito; un *dataset* de red se muestra de forma completamente distinta que un *dataset* geométrico; y así podríamos seguir.

Entonces, una actividad que le agrega mucho valor al proceso de diseño de una visualización es catalogar y abstraer las distintas partes de los datos que se trabajan. Después de conocer el contexto en el cual se está trabajando, esta actividad pediría enlistar las distintas partes de los datos a trabajar, y clasificarlos según las ideas que revisamos en las cápsulas anteriores.

Primero puede ser verificar que tipos de datos singulares hay presentes: ítems y atributos son los más usuales pero; ¿hay enlaces o posiciones presentes? Luego, al identificar los atributos, describir si son categóricos u ordenados, de llave o valor, si son ordenados: son cuantitativos u ordinales, son cíclicos o no, son secuenciales o divergentes. Finalmente puedes determinar el tipo de *dataset* en general, e identificar si es estático o dinámico.

A su vez, es importante también considerar la posibilidad de transformar los datos con los que contamos 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 o al abstraer los datos.

Por ejemplo, un conjunto de datos podría contener un atributo que describe la temperatura promedio en distintas localidades del planeta. Los valores con los que contamos 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 realmente importantes. Tal vez el contexto determina que se quiere transmitir la sensación general de la temperatura, y solo es necesario diferenciar entre frío, cálido y caluroso.

Al transformar la temperatura a este nueva apreciación, nos encontramos con que es un atributo ordinal ahora en vez de cuantitativo. Más extremo puede ser el caso si en realidad se busca determinar si la temperatura es peligrosa para salir de la casa o no. En esa nueva versión 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!