

## Cápsula 3: Partición en vistas

Hola, bienvenidxs a una cápsula del curso Visualización de Información. En esta hablaré sobre partición en vistas.

Esta idea de partición es distinta a la mencionada de particiones de ítems de un *dataset*, donde simplemente se dividen ítems en grupos disjuntos. Ahora, nos referimos a cómo organizar un *dataset* de múltiples atributos en grupos significativos en base a sus valores y cómo organizar dichos grupos espacialmente en vistas.

Se relaciona con la idea de definición de regiones categóricas mencionada en visualización de datos tabulares, pero la generaliza. La idea es pensar en cómo organizar datos relacionados considerando atributos que podrían no ser llave, y como usar el espacio en ese caso.

Por ejemplo, una partición usual es agrupar según los valores de un atributo categórico. Ese caso generalmente se realiza cuando tienen pocos niveles, o valores. Se pueden agrupar personas por su identidad de género, por ejemplo. La misma idea se podría aplicar en atributos ordinales, pero también es posible mediante atributos cuantitativos si se genera un dato derivado que agrupa valores en rangos.

Cuando se definen regiones categorías en base a atributos llave, eso lleva a agrupamientos de un solo ítem, pero al considerar particiones con atributos de valor, se generaron grupos de múltiples ítems.

Al mismo tiempo, se puede decidir cuántas particiones sucesivas realizar: inicialmente se puede particionar por un atributo categórico, y luego dentro de cada grupo resultante se particiona por un atributo ordinal. También se puede alterar ese orden de partición, y realizándolo de forma inversa.

En esta situación de múltiples particiones, la organización espacial se vuelve más complicada: hay más formas de subdividir el espacio. Podemos subdividir una vista en regiones para una primera partición, y luego cada región en glifos según una segunda partición.

Aquí hay un ejemplo donde la partición de datos es la misma, pero la partición en vistas varía. En pantalla vemos a la izquierda un gráfico de áreas apiladas, y a la derecha una alternativa de múltiples pequeños que utiliza gráficos de barra individuales.

En ambos casos, los datos están particionados por una llave primaria, que es el atributo que define las regiones horizontales que contienen las agrupaciones de barras; y un atributo de secundario llave para definir las barras individuales. A la izquierda se subdividen en el

sentido horizontal dentro de las regiones intermedias, y a la derecha se distribuye en el sentido perpendicular, con su propia alineación por llave.

Para cada partición en vistas, se facilita una tarea diferente. El gráfico de barras agrupado facilita la comparación de atributos por grupo primario, mientras que el múltiples pequeños facilita la comparación de atributos individuales a través de grupos.

La idea de partición puede usarse como método de exploración, donde la forma de realizar la partición puede ser seleccionada mediante interacción de usuario también.

Por ejemplo, en pantalla vemos la vista inicial del sistema HiVE, que muestra una partición inicial de cuatro regiones basadas en un atributo categórico del *dataset*. Cada región luego divide nuevamente el espacio en pequeñas regiones cuadradas, según otro atributo categórico. Finalmente, en lo más bajo, hay pequeños cuadrados dispuestos matricialmente donde el color codifica un atributo cuantitativo.

Esta vista enfatiza los patrones según la partición superior o inicial, se puede apreciar que las cuatro categorías son bien distintas y presentan perfiles únicos. De cierta forma se realiza una partición recursiva entre regiones, y usa la contención entre áreas para mostrar el orden de la partición.

Alternativamente, se puede alterar la distribución, intercambiando el orden de dos particiones antes presentadas. Nuevamente pone énfasis en la partición superior, donde es posible apreciar regiones que destacan inmediatamente.

La idea de particionar en vistas da para mucho ¡Muchas posibilidades! Es un aspecto que afecta directamente cómo se lee la información de una visualización, y afecta directamente a las tareas que intenta cubrir. Lo podemos considerar sobretodo en situaciones de *dataset* complejos o de múltiples atributos.

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!