

Cápsula 1: Reducción de datos

Hola, bienvenidxs a una cápsula del curso Visualización de Información. En esta hablaré sobre reducción de datos.

La idea de reducción es otra estrategia para manejar la complejidad de *datasets* en visualizaciones. La generalidad de estas opciones de decisiones de diseño es reducir la cantidad de información visible en una visualización y reducir la complejidad. Esta idea puede aplicarse sobre ítems de datos, o sobre atributos generales.

En situaciones estáticas se suele solamente reducir lo que se muestra, pero en situaciones dinámicas la opción de incrementar también es posible. Muchas de las ideas que revisamos en este grupo de cápsulas son bidireccionales, servirán para reducir o incrementar los datos visibles.

Hay dos opciones genéricas en este grupo: filtración y agregación. Aplicar filtros esconden elementos, mientras que la agregación crea un nuevo elemento que representan y reemplazan un grupo de varios otros elementos. Cada opción tiene sus ventajas y desventajas.

Filtración, por ejemplo, es simple de entender para un usuario, pero no es raro que se genere un sesgo por olvido de la existencia de los elementos que quedan fuera de la vista. Agregación por otro lado no oculta nada, solo reemplaza muchos elementos con un representante, pero este por definición muestra menos información.

Comenzaremos detallando la filtración que consiste en reducir el número de elementos visibles en pantalla escondiendo un grupo de ellos. Esta idea puede aplicarse tanto sobre ítems completos, como sobre atributos a través de todos los ítems.

La idea de filtración es bastante simple, el reto aparece al diseñar un sistema de visualización donde la filtración permite explorar el *dataset* de forma efectiva. Una forma directa es estableciendo qué ítems se mostrarán según sus valores individuales de atributos categóricos tal vez. O estableciendo un rango para un atributo cuantitativo, y a partir de ese rango se determinan qué ítems mostrar en la visualización y cuáles no.

Desde el punto de vista de implementación eso es bastante simple de hacer, pero desde el punto de vista de usuario no necesariamente. En situaciones donde las tareas a cubrir son descubrir o explorar aspectos del conjunto de datos, un usuario no tiene conocimiento previo de posibles valores para rangos. Este tipo de experiencia puede mejorarse con interacción y actualización en vivo en la herramienta, pero son detalles de interacción que hay que ir considerando caso a caso.

Una opción es filtrar por ítems, que corresponde a eliminar ítems en pantalla en base a los valores de atributos específicos. Menos ítems son visibles, pero el número de atributos mostrados no cambia.

En pantalla se muestra una imagen de la herramienta MoodPlay. Esta expone en un plano distintos artistas musicales, distribuidos según familiaridad en base al estado anímico que provocan.

Entre otras formas de navegación que permite, provee filtración mediante botones de categorías de estados anímicos generales. De esta forma provee inicialmente una panorámica del *dataset*, y luego permite revisar detalles según demanda del usuario.

La filtración de atributos es la otra forma u opción mencionada. Esta de forma inversa reduce la cantidad de atributos visibles en una herramienta, pero mantiene los ítems en vista. De todas formas, ambas opciones se pueden combinar, reduciendo tanto atributos como ítems visibles.

En el ejemplo en pantalla vemos un *idiom* de sigla DOSFA, que muestra a la izquierda múltiples codificaciones de estrella que traducen 215 atributos por cada ítem-estrella. A la derecha se muestra la versión filtrada de atributos, donde patrones se logran efectivamente apreciar.

Este tipo de filtración de atributos dinámico, también conocido como filtración de dimensionalidad, puede realizarse en base a datos derivados de atributos. Una forma es calculando algún tipo de medida de similaridad de un atributo, que crea un atributo ordenado en base a los valores que presenta el atributo.

Una opción puede ser calcular la varianza de atributos, por ejemplo. Esto permite comparar según varianza distintos atributos y permitir en base a esa medida filtrarlos de la visualización. O simplemente dejarle esa decisión al usuario, que escoja qué atributos mostrar.

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!