

Cápsula 2: Definición de regiones categóricas

Hola, bienvenidxs a una cápsula del curso Visualización de Información. En esta hablaré sobre la sub decisión de definición de regiones categóricas.

Al concentrarnos en atributos categóricos y en su codificado espacial, se vuelve más complejo que en el caso de valores cuantitativos. **La posición en un eje tiene semántica ordenada y los atributos categóricos no tiene, según el principio de expresividad no van bien juntos.**

Pero si va bien junto a la idea de **regiones espaciales**: áreas distinguibles, con límites y contiguas entre ellas. Dibujar todos los ítems con el mismo valor para un atributo categórico en la misma región utiliza proximidad espacial para codificar su similaridad de forma que hace más sentido con el principio de expresividad.

De todas formas a estas regiones hay que darles posiciones para hacerlas visibles, por lo que inevitablemente se genera un orden. Es por eso que la idea de regiones también puede traducirse a atributos ordenados, no solo a categóricos. Es el uso de atributos de llave el que va mejor de la mano con la separación de regiones.

Además, solo definir regiones es una parte. Se puede hacer mucho más dentro de cada región. Podríamos pensar la idea de definición de regiones como tres operaciones distintas: **separar en regiones, alinear regiones y ordenar regiones**. La separación y orden deben darse de alguna forma, pero el alineamiento es opcional. Cómo se deciden esas operaciones puede ser en base a múltiples atributos, dependiendo de la intención.

Cuando se trata con **un** atributo llave, el separar las regiones utilizándolo entrega una región por ítem. En este caso generalmente se organizan como una **lista unidimensional** vertical u horizontal. En la dirección que avanza se dividen distintas regiones, y en su dirección perpendicular se expande la región de la llave.

Un ejemplo de organización unidimensional por atributo llave es el clásico **gráfico de barra**. Utiliza la marca de línea donde su largo codifica un atributo cuantitativo orientado en un eje. En dirección perpendicular se definen regiones según un atributo de llave.

Por lo general se alinean las líneas, de forma que el canal sea más efectivo. Tiene límites de escalabilidad, ya que es necesario espacio entre líneas para diferenciar las regiones. Hasta cientos de barras es posible diferenciar.

El orden de las regiones puede ser arbitrario o según algún atributo. Si la llave es ordenada, debiera seguirse su orden intrínseco. Para categóricos, es posible usar orden alfabético, lo cual tiene la ventaja de hacer más fácil la búsqueda de ítems, o podría ordenarse según el atributo cuantitativo codificado, a modo de comunicar la tendencia.

Los gráficos de barra apilados utilizan un glifo más complejo, de barras separadas en sub barras. El largo completo sigue codificando un valor, pero las sub barras permiten ver sub componentes de ese valor. **Para esta codificación son necesarios dos atributos que funcionan como llave en conjunto.** Las regiones en lista se definen según un atributo principal, y las sub barras se definen según el segundo atributo.

Por lo general se apoya de color para diferenciar las sub barras y así permitir tareas de ubicación según llaves. Al apilarse, es posible comparar de forma efectiva el largo total de cada barra y la de la barra inferior, gracias al alineamiento. Es más difícil para las otras barras, por lo que el orden en que se muestran las sub barras es clave para determinar esa ventaja.

Los gráficos de punto son similares al gráfico de barra, pero en vez de utilizar una línea de marca y usar su largo para codificar un atributo, utiliza una marca de punto y usa su posición para codificar el mismo atributo.

Los gráficos de línea extienden a los gráficos de punto al agregar conexiones mediante líneas entre puntos vecinos. El uso de líneas enfatiza la idea de tendencia y de comparación de un valor con sus vecinos.

Los gráficos de línea, punto y barra codifican un atributo de valor cuantitativo mediante posición o largo unidimensional y utilizan un atributo de llave para separar el espacio. Los tres se pueden extender con el uso de color para agregar otro atributo.

La diferencia primordial radica en que las líneas del gráfico de línea enfatizan conexión y orden entre los ítems, por lo que proveen mejor intuición de percepción de tendencias ordenadas. **Es por esto que los gráficos de línea deberían ocuparse con llaves ordenadas y no llaves categóricas**, de lo contrario, se violaría el principio de expresividad.

Podemos apreciar esta diferencia con el ejemplo en pantalla, donde se muestra uso de barras y líneas en dos casos distintos: uno con llave categórica (el género de personas), y uno con ordenado (edades de personas). En todos los casos se codifica verticalmente la altura promedio de un grupo de personas, atributo cuantitativo.

Perceptualmente el uso de gráfico de línea aplica de mucho mejor forma en la llave de edades, gracias al orden intrínseco, que en el caso del género. Esto último da a entender una tendencia al “pasar de femenino a masculino” en altura, que no es particularmente lo que se quiere transmitir.

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!