

Cápsula 1: Escalas para organización tabular

Hola, bienvenidxs a una cápsula del curso Visualización de Información. En esta hablaré sobre el uso de escalas de D3.js como método de organización tabular.

En el último grupo de cápsulas revisamos la decisión de organización espacial para datos tabulares, que implican la expresión cuantitativa, definición de regiones categóricas y orientación de ejes. Todas estas ideas son reproducibles utilizando D3.js, específicamente, una herramienta que mencionamos en una cápsula anterior nos ayuda con varias de ellas: **las escalas**.

Las escalas cuantitativas de D3 son aquellas que definen un dominio de valores cuantitativo y un rango de valores cuantitativo. Estas son perfectas para la sub decisión de **expresión cuantitativa**, ya que esta quiere codificar un atributo cuantitativo a una posición espacial en una imagen, que también se comporta como un valor cuantitativo.

La escala lineal era la escala cuantitativa más básica que revisamos, que crea una correspondencia lineal y proporcional entre los valores de dominio y rango. Por lo general nos referimos a una escala lineal cuando se genera un eje, como el que muestro en pantalla ahora que va desde uno a un millón. El eje resultante muestra a la mitad de la distancia el valor intermedio del dominio, el medio millón.

Ahora, hay situaciones donde es necesario ejes con una escala que aumenta de **forma no lineal**. Por ejemplo, cuando los órdenes de magnitud de nuestros datos son muy diferentes, el usar una escala lineal produciría que solo los datos más altos ocupen gran parte del espacio. En esos casos utilizar una **escala logarítmica** puede ser conveniente.

D3 provee tal escala, en "d3.scaleLog". En pantalla muestro la misma definición que hicimos con la escala lineal, pero cambio el tipo de escala. Si vemos el eje correspondiente, notamos que el aumento de valores es por orden magnitud, comenzando por el valor uno, y en cada paso aumenta en uno el exponente del valor codificado.

Si bien perceptualmente puede parecer raro, depende de la tarea a cumplir qué tipo de escala utilizar. Una escala lineal se puede necesitar si realmente se quiere apreciar diferencias de órdenes visualmente. En cualquier caso siempre hay que mostrar leyenda o marcar los ejes de forma que sea claro el tipo de escala utilizada, ya que la lineal es la que entenderemos por defecto.

Hay muchas otras escalas continuas que aumentan de distintas formas y también podemos personalizar cómo aumentan. Por ejemplo, la escala logarítmica por defecto aumenta con una base de 10, pero podemos alterar su base mediante el método "base". Podemos ver la diferencia de aumento en el ejemplo en pantalla, donde ahora se establece un dominio de valores con un máximo que es potencia de dos y la escala que está en base a dos.

Otras escalas cuantitativas son las de potencias, que aumentan según algún exponente. Pueden aumentar siguiendo la potencia de dos, o de tres, pero también potencias intermedias. Una que se suele ocupar bastante es la **raíz cuadrada**.

Aparece específicamente en situaciones donde se busca **codificar un valor cuantitativo mediante el tamaño de un área, y para dibujar tal área se necesita uno de sus tamaños unidimensionales**. Como la radio para el círculo. Si utilizamos una escala lineal para codificar el radio, nos encontramos que el **área aumenta de forma desproporcionada**, al cuadrado de hecho.

Pero si ocupamos una escala de potencia "scalePow" con valor de potencia un medio, el **radio aumenta como raíz cuadrada, y el área aumenta de forma proporcional al atributo a codificar**. Incluso hay una escala de raíz cuadrada ya nombrada en D3.

Por otro lado, está la sub decisión de definición de regiones categóricas. Con la idea de escalas lineales nos la podemos ingeniar para generar espacios, pero D3 nuevamente llega con herramientas ya solucionadas. Las escalas "d3.scaleBand" y "d3.scalePoint" nos ayudan a establecer y definir regiones categóricas.

Estas dos escalas se consideran ordinales, porque el dominio y rango de valores que se entregan son discretos, no cuantitativos. El detalle es que los valores de rango son del tipo numérico y viven en un espacio cuantitativo. Esto las hace perfectas para la idea de definición de regiones: **áreas definidas con bordes y contiguas entre ellas**.

Revisaremos "scaleBand" ya que es más compleja. La idea detrás de ella es generar bandas de tamaño uniforme y orientadas como lista a lo largo de un espacio visual. La cantidad de bandas depende de la cantidad de valores en el dominio, ya que se reserva una banda por valor. La función resultante recibe un valor y retorna la posición del inicio de la banda correspondiente al valor entregado.

En pantalla hay código para generar regiones usando "scaleBand". En el dominio se especifican los valores de categorías a colocar, cuatro en este caso, y el rango es el intervalo de valores espaciales que se separa en bandas. En nuestro caso va desde cero al ancho completo disponible.

Bajo esa definición, agrego rectángulos, texto y líneas que usan esta escala para disponerse y las agregué para que veamos el resultado de separación de "scaleBand". Si lo vemos en ejecución, vemos que efectivamente se separa el espacio completo en cuatro espacios, uno por categoría. Con una línea segmentada se marca la mitad de cada región como referencia. **Recuerda que es el borde izquierdo el que retorna la escala, no el centro**.

Con esto vemos que por defecto se usa el espacio completo provisto para definir bandas, pero esto se puede alterar. Los métodos "paddingInner" y "paddingOuter" permiten definir cuánto espacio se dedica para espaciado entre bandas (*inner*) y espaciado a los bordes

externos (*outer*). Los valores son razones entre 0 y 1 de cuánto espacio se dedica a separación entre bandas y cuánto espacio se dedica efectivamente a bandas.

Si lo ejecutamos, vemos que efectivamente ahora se dedica espacio entre las bandas y alrededor de ellas. Hay otro método, "align", que permite alterar cómo se dedica el espacio exterior. **Juega por tu cuenta con los valores de este método de espaciado con este código y para ver que pasa.**

Con esto podemos simular perfectamente la idea de definición de regiones categóricas de lista para un atributo categórico y llave en un *dataset*. De la misma forma, podemos hacer uso de una segunda escala de banda para definir una escala orientada verticalmente, de forma que en conjunto con esta sea alineamiento matricial. Des-comentaré un código bastante similar al anterior, pero que orienta regiones en sentido vertical.

Si lo ejecutamos, vemos que efectivamente se genera una definición matricial de regiones. Puede parecer extenso el código, pero mucho de eso es simplemente para mostrar las regiones de forma visual, en la práctica se utilizarían directamente las escalas para posicionar elementos específicos.

Por ejemplo, al final agregué un arreglo de objetos que tienen dos atributos de categorías y un valor de color. A partir de él, agregué rectángulos que sigan las posiciones definidas por las escalas recién creadas, en base a los valores de categoría de cada objeto. Si lo ejecutamos, vemos que se genera lo esperado en la intersección de las regiones definidas.

El uso de "scaleBand" es conveniente gracias a que los valores que entrega son la posición izquierda de las regiones, de la misma forma en que funcionan muchos elementos SVG como "rect" que se mueven en base a la esquina superior izquierda. Pero alternativamente podríamos querer trasladar elementos alrededor de un eje central de la banda.

Podríamos usar el ingenio para generar eso utilizando "scaleBand, pero "scalePoint" tiene ese comportamiento directamente, que separa puntos y ejes espacialmente en un rango cuantitativo. En pantalla tengo código muy similar al anterior pero utilizando "scalePoint" y usaré estas escalas para posicionar círculos SVG, que se posicionan en relación a su centro.

Si lo ejecutamos, obtenemos el resultado en pantalla. **Te recomiendo revisar este código para notar las diferencias entre escalas y cómo distribuyen el espacio.**

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!