

Cápsula 2: Escalas de D3

Hola, bienvenidxs a una cápsula del curso Visualización de Información. En esta hablaré sobre escalas de D3, funciones que nos apoyarán a hacer cálculo de propiedades geométricas al crear visualizaciones.

Para comenzar y motivar el uso de escalas, primero analizaremos el programa actual con el que quedamos en la cápsula pasada. Lo primero raro que podemos apreciar es que los tamaños completos de los elementos se manejan de forma extraña. El ancho del SVG principal depende del tamaño de los datos, mientras que la altura es fija.

En general, es mejor fijar un tamaño fijo para nuestras visualizaciones, y adaptar internamente los tamaños dentro de ella. Así, podemos asegurar visibilidad y legibilidad de ella a partir de cierto tamaño. Eventualmente podríamos adaptarlo a cambios de tamaño de ventana, pero no nos adelantemos.

Fijaremos dos variables constantes al comienzo, que definen el ancho y el alto del SVG completo y las usaremos en el resto del programa. Podemos mover de lugar esta otra declaración, ya que el SVG se altera una única vez ahora.

Ahora, debemos adaptar algunas características de las barras. Primero, el ancho y posición horizontal de estas pierde sentido ahora que fijamos el ancho completo como fijo. El espacio que ocupan depende del número de datos presentes y del espacio disponible.

Por ahora, calcularemos el ancho de una barra, como la mitad de dividir el espacio horizontal de forma equitativa entre todos los datos. Al usar la mitad, se podrá usar la otra mitad como espacio entre barras. Entonces, fijamos el ancho de las barras con este valor, y la posición horizontal como dos veces este ancho, de forma que deje espacios entre barras.

El otro aspecto extraño de las barras es su altura. Como el SVG tiene altura fija, las alturas de las barras podían pasarse de este alto y se cortaban. Para este caso, necesitamos escalar los largos de las barras al tamaño del SVG. Pero para hacer eso, necesitamos saber cual es el valor máximo del atributo que estamos representando, para saber cómo escalar los otros valores.

Para extraer el máximo de un atributo, podemos usar una de las funciones de utilidad de D3: "d3.max". Esta recibe en su primer argumento un arreglo de datos, y en su segundo argumento una función que extrae algún valor numérico de cada dato en el arreglo. Entonces usa función para extraer valores, y a partir de esos extrae el valor máximo y lo retorna.

Con eso, podemos extraer el valor máximo del atributo frecuencia. También existen funciones similares a esta, como “d3.min” para extraer el mínimo, y “d3.extent” que retorna el mínimo y máximo juntos en un arreglo.

Con el máximo del atributo frecuencia podemos escalar las alturas de las barras. El máximo ocupará la altura completa, y el resto es proporcional. Si probamos estos resultados en nuestro ejemplo, ¡vemos que funciona!

Ahora, otro tema extraño de este “gráfico de barras” es que está invertido. Esto es porque hasta el momento no nos hemos dado el trabajo de definir una posición vertical “y”, y se coloca como cero por defecto.

Si pensamos cómo definirla, recordando que la altura aumenta hacia abajo, esta es opuesta al alto de las barras. El valor máximo de atributo tendría posición 0, mientras que frecuencia igual 0 tendría posición en el fondo del SVG. Matemáticamente, podemos expresarlo así. Si lo probamos, ¡vemos que funciona!

Ahora, nota de este código que hemos necesitado hartas matemáticas geométricas para ordenar las cosas. ¡Y tiene sentido! Estamos haciendo figuras geométricas. Esto es bastante frecuente, por lo que D3 provee un tipo de objeto para manejar este tipo de situaciones: las **escalas**.

Las escalas **son funciones que permiten transformar valores que provienen de datos a cálculos de interés para una visualización**, como lo es posicionar elementos geométricos. Cómo funcionan es que definen un **dominio**, es decir un intervalo de datos que recibirán como función, y un **rango**, que es el intervalo de valores que van a retornar.

Si lo piensan, varios de los cálculos que hemos realizado son lineales: los valores aumentan o disminuyen proporcionalmente entre un máximo y un mínimo. Para esto existe un tipo de **escala lineal** en D3: “d3.scaleLinear”. Definiremos una para la altura de las barras y otra para la posición vertical de las barras.

Como mencionado, es necesario definir el dominio y rango de una escala. Para el caso de la altura de las barras, el dominio se mueve entre 0 y la frecuencia máxima que hay en los datos. Mientras que el rango, o los valores que retornará la escala, se mueven entre 0 y la altura máxima del SVG. En dominio existen los valores de datos, y en rango los valores en SVG. Si reemplazamos la definición anterior llamando a la escala recién usada. ¡Vemos que funciona igual!

Para el caso de la posición vertical es similar, pero justo al revés. El dominio es idéntico al anterior, pero el rango se invierte. Es por eso que si simplemente invertimos los valores entregados a “range”, hará los cálculos de forma invertida. Un 0 en datos sería altura máxima, y la frecuencia máxima sería 0 de altura. Reemplazamos con la escala.

Para el caso de la posición horizontal es un poco más elaborado pero similar. Lo complejo aquí es que en realidad queremos posiciones que se comportan de forma discreta más que real y en base a proporciones.

Las escalas como objetos son parte del subpaquete de D3: "[d3-scale](#)", y tiene muchos tipos de escalas para distintas situaciones. De hecho, tiene una perfecta para nuestra situación: "[scaleBand](#)". Esta permite definir bandas horizontales a lo largo de un rango de un largo uniforme.

Definiremos entonces para la posición horizontal una escala utilizando "scaleBand". Revisando su documentación, nos podemos encontrar que define sus atributos de otras formas a las anteriores. Por ejemplo, para el dominio permite ingresar todos los valores posibles del atributo, y en este caso nos permitirá posición por el atributo de categoría en vez del índice.

Si la reemplazamos y la ponemos en práctica, ¡vemos que funciona! E incluso nos posicionó de una forma mejor y espaciada las barras.*

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!

[*Faltó mencionar un detalle para el caso del eje X del gráfico. Ahora los anchos de las barras quedaban a disposición de la escala utilizada para posicionarlos, ya que esta también es capaz de calcular sus tamaños a partir de los datos entregados. La escala tiene el método "bandwidth" que entrega este valor, y la utilizamos en el código de ejemplo entregado para definir el ancho correcto.]