

## Cápsula 2: Tipos de *datasets*

Hola, bienvenidxs a una cápsula del curso Visualización de Información. En esta les hablaré sobre los tipos de conjuntos de datos o *datasets*.

Un **conjunto de datos o *dataset***, es cualquier colección de datos. Aquí describiré tres tipos de *datasets* básicos: las tablas, las redes y *datasets* geométricos. Hay muchas otras formas de agrupar datos, como los campos, los conjuntos, clusters y listas, pero por ahora nos enfocaremos solo en los primeros tres.

También es importante notar que si bien describiremos tipos distintivos, es completamente posible que un *dataset* pertenezca a múltiples tipos al mismo tiempo, y en la vida real es posible que nos encontremos con combinaciones de tipos de *datasets*.

El tipo más común por su formato son las **tablas o datos tabulares**, que es cualquier colección de datos que contiene entidades y descripciones de dichas entidades. Les llamamos tabulares ya que es muy frecuente encontrar este tipo de datos como tablas, que organizan la información en filas y columnas.

Una tabla de datos generalmente especifica **atributos en sus columnas**, y cada fila presenta un **ítem distinto** a ser descrito. Las celdas especificadas por la combinación de fila y columna, muestra el **valor** asociado a dicho par. Al ser una forma común de describir entidades, es normal ver *datasets* que combinan este tipo con otro, como red o geométrico.

Nota que en la práctica no es necesario que los datos se almacenen como tabla para ser considerados tabulares. Basta que lo que almacene es información de entidades y descripciones de dichas entidades. Es de particular interés aquellos *datasets* tabulares cuya información no tiene una forma de representarse visualmente predeterminada, y que nos deja espacio a quienes diseñamos visualizaciones el trabajo de idear cómo codificarlas.

Los ***datasets* de redes** surgen al describir situaciones donde existen relaciones entre dos o más ítems en el conjunto de datos. El tipo de dato singular base son los enlaces que vinculan dos ítems, que en el contexto de redes generalmente se llaman nodos.

Por ejemplo, en un *dataset* sobre una red social, los usuarios son nodos de la red, y la relación de seguimiento entre usuarios son enlaces. En una red de interacción entre genes, los genes son nodos y un enlace entre nodos representa si se conoce que dichos genes interactúan entre ellos. No solo los nodos pueden tener atributos, si no que los enlaces también son ítems con la capacidad de ser descritos mediante atributos.

¿Se te ocurre algún dato de tipo enlace que tenga atributos distintivos? ¡Déjalo en los comentarios! Los *datasets* de redes tienen muchas particularidades, y pueden representar subtipos de *datasets* como los árboles o jerarquías.

**Los conjuntos de datos geométricos** son aquellos con datos intrínsecamente espaciales. Especifican información sobre la forma de ítems con posiciones espaciales explícitas. Estos ítems pueden tratarse de puntos, líneas o curvas unidimensionales, o incluso regiones bidimensionales o incluso volúmenes tridimensionales. Su uso aparece en situaciones donde es necesario entender formas o posiciones.

Este tipo no necesariamente tiene atributos asociados a sus ítems, y de ser el caso, el desafío se reduce a simplemente mostrar visualmente sus formas. Datos puramente geométricos son solo interesantes en el contexto de visualización de información cuando es transformado o derivado de una forma que necesitan aplicarse decisiones de diseño. Cómo producir formas o elementos gráficos que repliquen la forma de información espacial.

Finalmente, otra forma de describir conjuntos de datos es según su **dinamismo**, es decir, si son **estáticos o dinámicos**. Un *dataset* estático es aquel que no cambia y por lo tanto se cuenta con su estado completo desde el comienzo del proceso de diseño, y algunas consideraciones del proceso se simplifican. Los dinámicos son aquellos que durante el uso de la herramienta los datos que se observan cambian.

Pero también podemos encontrarnos con *datasets* dinámicos que cambian gracias a una corriente de información entrante que ocurre incluso mientras se usa una herramienta de visualización. Potenciales consideraciones a tener en ese caso es la actualización de datos considerados en la herramienta, o agregar nuevas entidades que vayan apareciendo.

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!