

## Resumen de materia abarcada hasta antes de la ayudantía N°2

“Los sistemas computacionales de visualización proveen representaciones visuales de conjuntos de datos, diseñadas para ayudar a personas a realizar tareas más eficientemente.”

- Tamara Munzner

Es una interdisciplina que incluye conceptos de la computación, diseño gráfico y psicología perceptiva/cognitiva.

El Modelo Anidado o Framework Anidado fue propuesto por Tamara Munzner en Visualization Analysis and Design. Tiene por objetivo garantizar y estructurar el proceso de diseño en visualización de la información. Así sistematizar la experiencia de creación y separarlo en etapas. Se busca guiar el proceso, no atarlo.

- Qué -> Datos; ¿qué datos vamos a usar? ¿cómo se caracteriza el dataset?
- Porqué -> Acciones y Tareas; ¿por qué se hará la visualización? ¿Cuáles acciones o tareas se necesitan para el usuario?
- Cómo -> Codificaciones o idioms; ¿cómo será nuestra visualización? ¿qué codificaciones o idioms utilizaremos?

Donde se utilizan triplas (dato, tarea e idiom) para describir y justificar decisiones de diseño en una herramienta de visualización. Esta tripleta puede ser simple o complejizarse según los tipos de datos o la tarea.

Modelo de cuatro etapas:

- 1) **Caracterización de dominio:** Entender el campo de conocimiento y contexto de aplicación. Entender datos, usuarios y tareas. Entrevistas observaciones o investigación mediante literatura.
- 2) **Abstracción de datos y tareas:** Traer las particularidades del contexto al mismo lenguaje y abstraerlos. Definir el **qué** y el **porqué** de la herramienta. Seleccionar los aspectos a consideración. Ayuda a simplificar y aterrizar decisiones de codificación posteriores.
- 3) **Codificación visual y de interacción:** Definir el **cómo**. Elección de idioms visuales y de interacción de una visualización.
- 4) **Implementación algorítmica:** Nivel de implementación técnica, mediante plataformas o programación. Se revisa eficiencia, escalabilidad y dificultad de programación.

Se dice que es anidado (o en cascada), porque resultados obtenidos en etapas previas repercuten en las posteriores, pudiendo arrastrar errores para más adelante. Es importante

la noción de separar dudas y procesos de diseño en subprocesos a resolver. Sin embargo, no implica completar un nivel totalmente antes de pasar al siguiente, sino que es algo iterativo.

Un **idiom** es un método o enfoque distintivo de codificar información o manipularla. Dicho de otra forma son las decisiones tomadas para transformar los datos en una visualización. Pueden ser de dos tipos: **idioms visuales** e **idioms de interacción**. El primero se refiere a cómo se muestra algo de manera estática, y el segundo a la posibilidad de albergar interactividad o manipulación de parte del usuario sobre la herramienta.

Las **marcas** son elementos gráficos básicos (o elementales) en una imagen. Son elementos geométricos simples, clasificable según el número de dimensiones que tenga. Suelen ser utilizadas para representar entidades.

Marca cero dimensional es un punto, unidimensional son líneas, bidimensionales son áreas y tridimensionales son volúmenes.

Los **canales** son formas de controlar la apariencia de las marcas. Por ejemplo, posición, forma, curvatura, rotación, color, tamaño. A su vez pueden tener **subcanales** como, por ejemplo, el color que tiene matiz, saturación y luminosidad. Los canales se encuentran sujetos al tipo de marca en particular.

Hablar de marcas y canales permite desagregar parte por parte las decisiones a tomar, y cómo codificar visualmente la información.

Los **glifos** son objetos gráficos complejos conformados por múltiples marcas primitivas.

Los tipos de canales se separan en **canales de identidad** y **canales de magnitud**. El primero ayuda a entender qué es algo, o donde está algo. El segundo ayuda a entender de cuánto hay algo.

No todos los canales son iguales o equivalentes. Codificar un mismo dato con canales distintos es percibido de manera distinta. Para una correcta elección de marcas y canales utilizamos los siguientes principios:

- 1) **Principio de expresividad:** Debe haber coherencia entre el tipo de canal (magnitud, identidad) con la semántica del atributo (cuantitativo, ordinal, categórico). Planteado de otra forma, una codificación visual debe expresar toda y solamente, la información del atributo objetivo.
- 2) **Principio de efectividad:** Los atributos más importantes deben ser codificados con los canales más efectivos, perceptiblemente hablando.

La semántica de los datos es el significado que estos tienen en el mundo real. Que está muy vinculado al tipo de dato, el cual es una interpretación estructural o matemática. A veces es necesario algo de contexto o información adicional, conocido como metadata.

Tipos de datos singulares:

- **Atributo:** propiedad específica que se puede medir, observar o registrar. También, conocido como variable o dimensión.
- **Ítem:** entidad individual y discreta que puede ser simple o compleja
- **Enlace:** relación entre ítems
- **Posición:** ubicación en espacio unidimensional, bidimensional o tridimensional. Dato espacial.
- **Grilla:** estrategia para obtener una muestra de datos continuos, lo que determina relaciones geométricas y topológicas entre celdas

Tipos de datasets:

- **Tabulares:** es el más común, se conforma por filas y columnas. Sus tipos de datos son ítems y atributos. Donde se hace una descripción de los atributos para cada ítem.
- **Redes y árboles:** representan las relaciones entre dos o más ítems. Donde el ítem recibe el nombre de nodo y la relación el nombre de enlace o vínculo. Pudiendo cada uno ser asociado con atributos adicionales. Un árbol es un caso particular de red en el que no existen ciclos y suele haber jerarquía.
- **Geométrico:** son aquellos que especifican información sobre la forma de ítems con posiciones espaciales explícitas.

Los datasets pueden ser clasificados en base a la disponibilidad que tengan. Por un lado, se tienen los datasets **estáticos** (offline) que es aquel que no cambia y se encuentra disponible en su totalidad desde el comienzo. Por otro lado, están los datasets **dinámicos** (online) en donde la información llega o se ve modificada a través del tiempo o de la interacción.

Tipos de atributos:

- **Categoricos (nominales):** no tienen un orden intrínseco. La comparación únicamente ocurre desde la igualdad o diferencia.
- **Ordenados:** poseen un orden intrínseco. Se separan en:
  - **Ordinales:** no tienen comparación aritmética exacta entre valores, pero poseen un orden.
  - **Cuantitativos:** sí permite la comparación aritmética exacta entre sus valores.
- **Jerárquicos:** Existe una estructura jerárquica entre uno o múltiples atributos.
- **Llave:** atributo utilizado como índice para identificar otros atributos. No posee duplicados.
- **Valor:** atributo medido y puede tener repeticiones.

Tipos de atributos (según dirección de ordenamiento):

- ***Secuencial:*** valores distribuidos en un rango homogéneo con un valor mínimo y máximo.
- ***Divergente:*** puede ser descompuesto en dos secuencias con direcciones opuestas que poseen un punto común.
- ***Cíclicos:*** valores siguen un orden que los devuelve al valor mínimo
- ***No cíclicos:*** aumenta indefinidamente o llegan a un máximo.