

---

# IIC2026

# Visualización de Información

— Hernán F. Valdivieso López —  
(2023 - 1 / Clase 15)

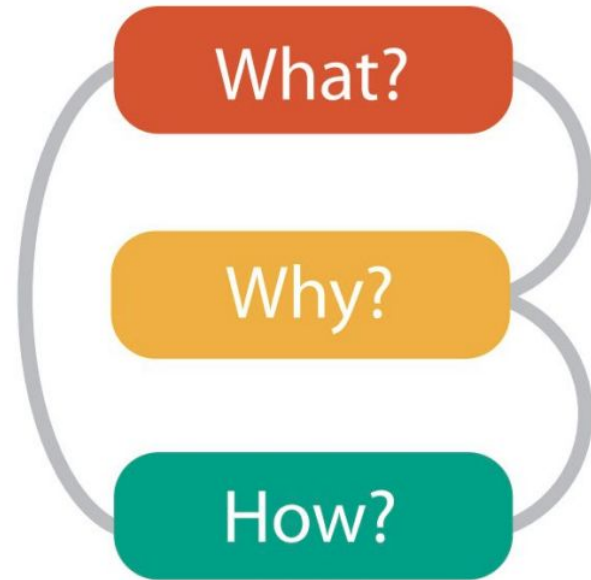
---

# Temas de la clase - Abstracción de datos

1. Semántica y tipos de datos singulares.
2. Tipos de *datasets* y su disponibilidad.
3. Tipos de atributos.

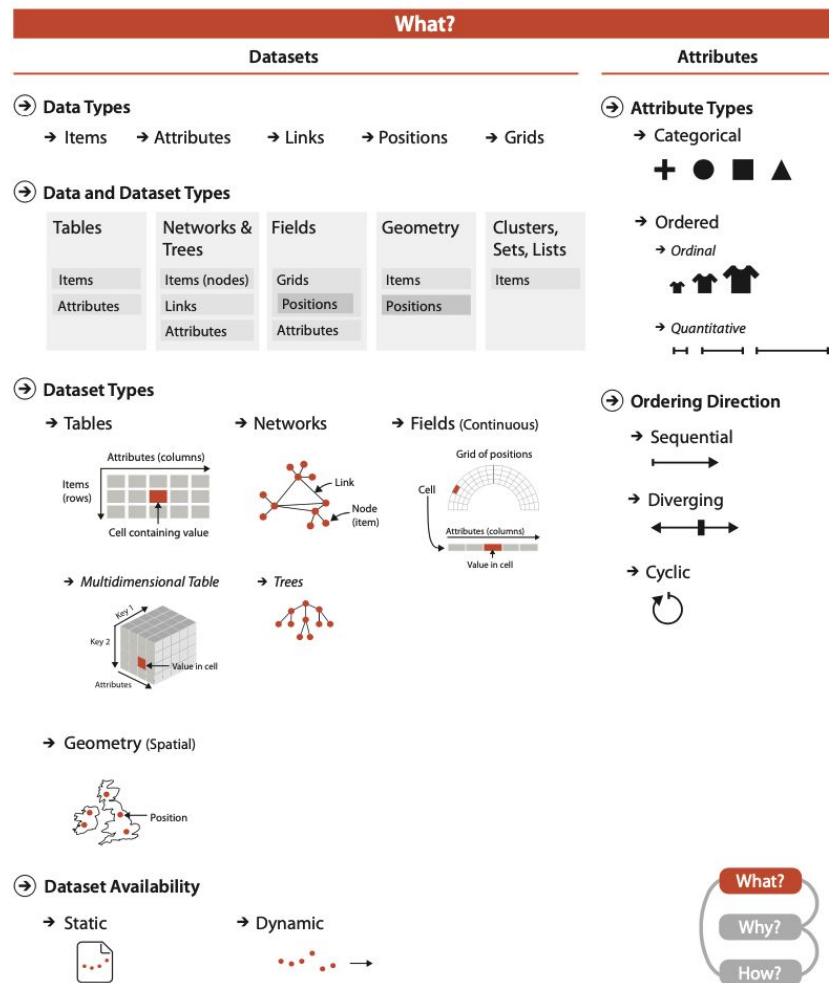
# Framework de Tamara Munzner

- El *framework* propone tres preguntas para analizar y crear una visualización:
- *What?* → qué información se está mostrando (data).
- *Why?* → por qué queremos visualizar esto (tareas visuales).
- *How?* → cómo se construye la visualización (*idioms*).



# ¿What? Abstracción de datos

- ¿Qué semántica tienen los datos (significado en el mundo real)?
- ¿Qué tipo de datos/set de datos tenemos?
- ¿Qué tipo de disponibilidad tienen los datos?
- ¿Qué tipo de atributos tienen los datos?



# Semántica y tipos de datos singulares

---

# Semántica de los datos

- Muchos aspectos que guían el diseño de una visualización son **impulsados por el tipo de datos** que tenemos a nuestra disposición.
- Hay que preguntarse, entonces, **qué tipo** de datos tenemos, **qué información** podemos obtener directamente, y **qué sentido** tienen realmente.

# Semántica de los datos

1, Santiago, 20, 22.3, L, Los Simpsons

# Semántica de los datos

- Para salir de las adivinanzas, es necesario saber **dos tipos de información**: la semántica y el tipo de dato.
  - La **semántica** es su significado en el **mundo real** (¿qué es? ¿un nombre de una persona, una ciudad, una abreviación de un punto cardinal, una cantidad?).
  - El **tipo de dato** es **interpretación estructural** o matemática del dato (¿es un ítem, un enlace o un atributo? Si es atributo ¿es categórico u ordenado?).
- A veces, se necesita leer **información adicional** (conocida como metadata) para poder interpretar correctamente un dato.



# Semántica de los datos

1, Santiago, 20, 22.3, L, Los Simpsons

# Semántica de los datos

ID	Nombre	Edad	BMI*	Talla polera	Serie preferida
1	Santiago	20	22.3	L	Los Simpsons
2	Felipe	15	19.1	M	Game of Thrones
3	Francisca	25	15	M	HxH
...	...	...	...	...	...

\*BMI: Body Mass Index (Índice de Masa Corporal)

# *Data Types* (tipos de datos)

## ➔ Data Types

➔ Items   ➔ Attributes   ➔ Links   ➔ Positions   ➔ Grids

- Ítems
- Atributos
- Vínculos
- Posiciones
- Grilla

# Atributos

Es una **propiedad** específica que puede ser medida, observada o registrada. También se le conoce como variable o dimensión.

- Por ejemplo: temperatura, salario, precio, número de ventas, etcétera.

ID	Nombre	Edad	BMI*	Talla polera	Serie preferida
1	Santiago	20	22.3	L	Los Simpsons
2	Felipe	15	19.1	M	Game of Thrones
3	Francisca	25	15	M	HxH
...	...	...	...	...	...

# Ítems

Es una **entidad** discreta (e.g. fila en una tabla, nodo en un grafo).

- Por ejemplo: personas, ciudades, tiendas de computación.

ID	Nombre	Edad	BMI*	Talla polera	Serie preferida
1	Santiago	20	22.3	L	Los Simpsons
2	Felipe	15	19.1	M	Game of Thrones
3	Francisca	25	15	M	HxH
...	...	...	...	...	...

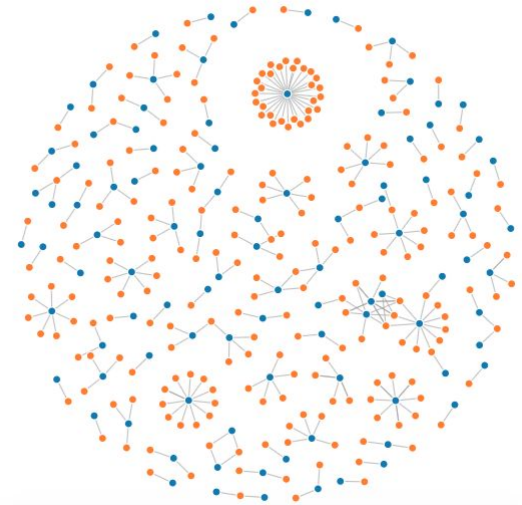
# Vínculos o enlace

Es una **relación** entre los ítems.

- Por ejemplo: amistades de una red social, viajes entre países/ciudades, árbol genealógico.

**Ojo:** Una relación también puede ser un ítem.

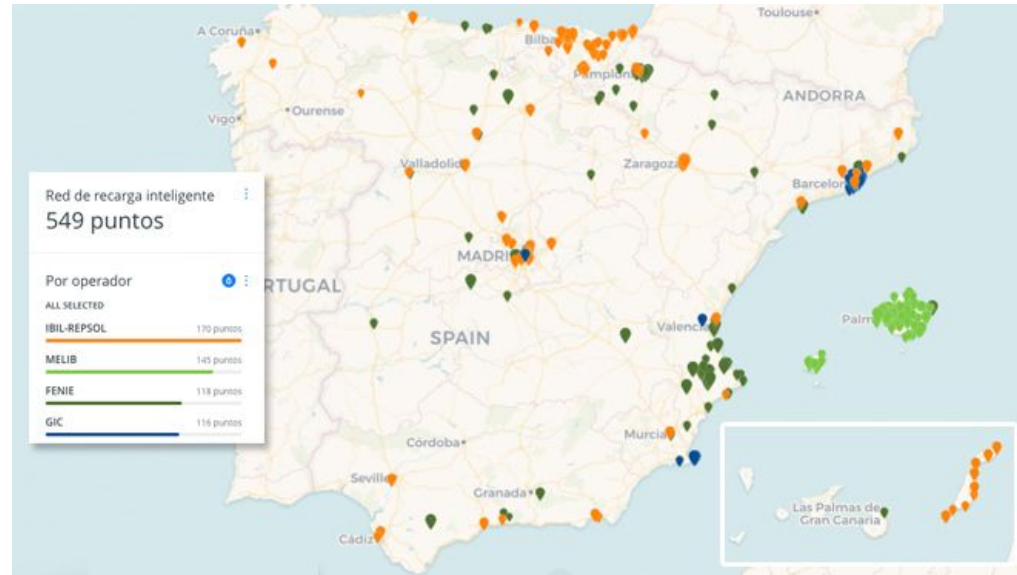
- Por ejemplo: **Comentario en youtube**. Un comentario permite relacionar el ítem “usuario” con el ítem “video”. Pero del mismo modo, el comentario por si solo es un ítem con atributos que tiene:
  - Fecha de creación.
  - Cantidad de *likes*.
  - Contenido.



# Posiciones

Es un dato **espacial**, que provee una **ubicación** en un espacio 2D o 3D.

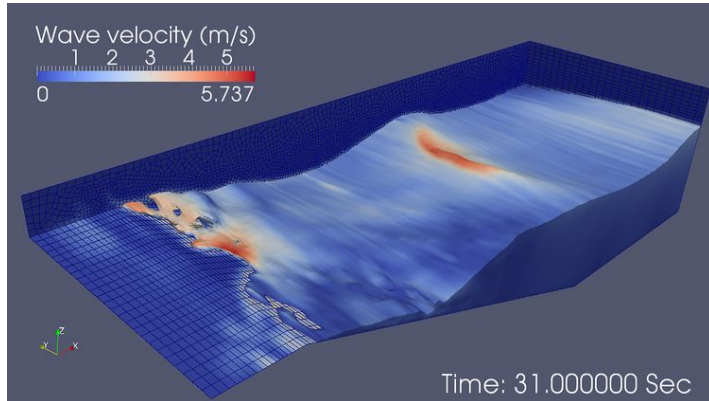
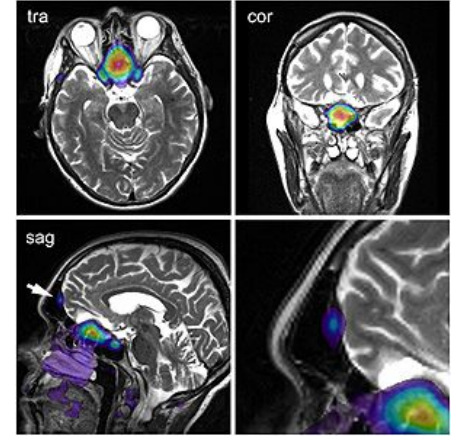
- Un par latitud-longitud mostrando una ubicación en la Tierra,
- La ubicación en la región de un escáner médico.
- La posición de las mesas para una boda.
- Posición de los íconos en una pantalla de computador.



# Grilla (no profundizaremos en el curso)

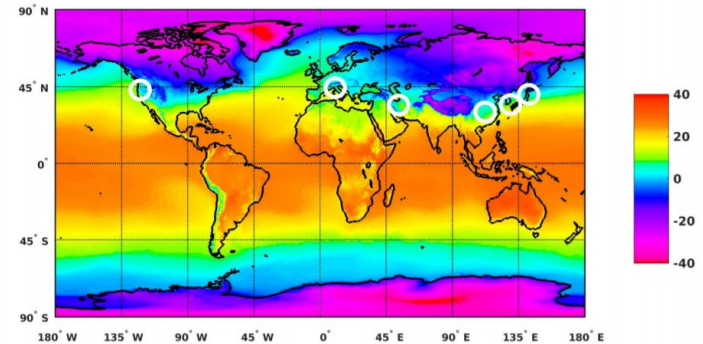
Ejemplos:

- En mediciones de escáneres médicos que actúan sobre cuerpos tridimensionales
- Simulación de fluidos.
- Mediciones meteorológicas a lo largo de la tierra.



16

Average 2-meter Temperature (Celsius) for Jan-Feb 2020 (ERA-5)





# Tipos de *datasets* y su disponibilidad

---

# Tipos de *dataset* (*dataset types*)

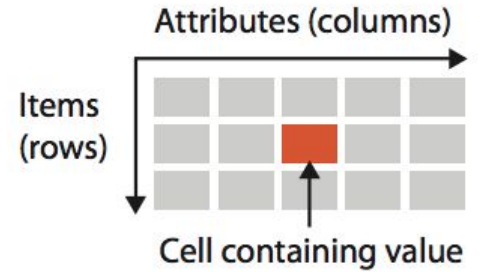
Principales tipos de *datasets* son:

- Tabulares.
- Redes y árboles.
- Geométricos.

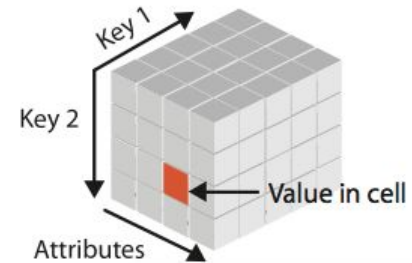
Cada uno de ellos, está compuesto por los tipos de datos recién vistos.

# Tabular

- Es el tipo de *dataset* más común.
- Viene en forma de filas y columnas (e.g. *spreadsheet*, excel) .
- Los tipos de datos son: **ítems y atributos**.
- Cada celda de la tabla es un valor para la combinación ítem-atributo.
- Además, existen las tablas multidimensionales, que tienen múltiples llaves.



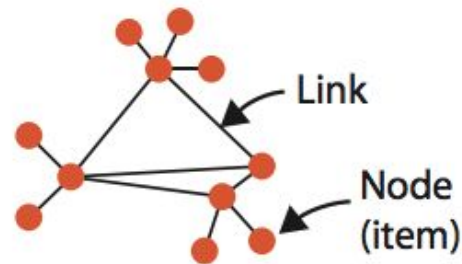
## → Multidimensional Table



# Redes y árboles

- Es apropiado para mostrar que existe algún tipo de relación entre dos o más ítems.
- Un **ítem** en una red es llamado nodo o vértice.
- Una **relación** entre dos o más nodos se llama enlace o vínculo.
- Por ejemplo: las personas pueden ser representadas como nodos y su relación de amistad entre ellas como vínculos.
- Adicionalmente, es posible asociar atributos a cada nodo y enlace.
- Más adelante abordaremos más sobre este *dataset*.

## → Networks



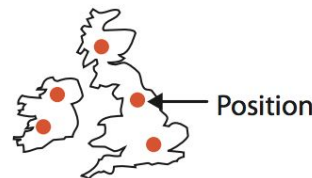
## → Trees



# Geométricos

- Los ítems pueden ser puntos, curvas, superficies o volúmenes.
- **Ítems con posiciones** explícitas.
- Este tipo de *dataset* puede que no tenga atributos adicionales aparte de la posición.
- También es necesario saber con qué nivel de detalle se generan las formas (*shapes*) desde datos geográficos crudos.
  - Por ejemplo, la frontera de un bosque, o de una ciudad, o también la curva de una carretera.

→ Geometry (Spatial)



# Tipos de *dataset* (*dataset types*)

- **No es necesario condicionar un *dataset* a un único tipo.**
- Un *dataset* más complejo puede estar compuesto por alguno de los datasets visto hasta ahora.
  - *Dataset* de Red Social con información de los usuarios y sus amistades será tabular y de red.
    - Si entre la información de los usuarios tenemos el país de origen o algún dato para posicionar al usuario, el *dataset* también será geométrico.
  - Un *dataset* sobre los países y estadísticas del país (índice per cápita, población, etc.) será tabular y geométrico.

# Disponibilidad del *dataset*

- Existen dos categorías: *datasets* estáticos y *datasets* dinámicos:
  - **Estático** (offline) es cuando el *dataset* está disponible all at once (i.e. todo en un instante)
  - **Dinámico** (online) es cuando nueva información llega a través del tiempo (streaming data)
- Cuando el *dataset* es **dinámico**, nuevos datos pueden ser agregados, otros eliminados o también actualizados .
  - Se agrega complejidad en varios aspectos al proceso de visualización comparado a un *dataset* estático.

## ➔ Dataset Availability

### ➔ Static



### ➔ Dynamic



# Tipos de atributos

---



# Atributos

- Tipos de atributos
- Dirección de ordenamiento

## ➔ Attribute Types

➔ Categorical



➔ Ordered

➔ *Ordinal*



➔ *Quantitative*



## ➔ Ordering Direction

➔ Sequential



➔ Diverging

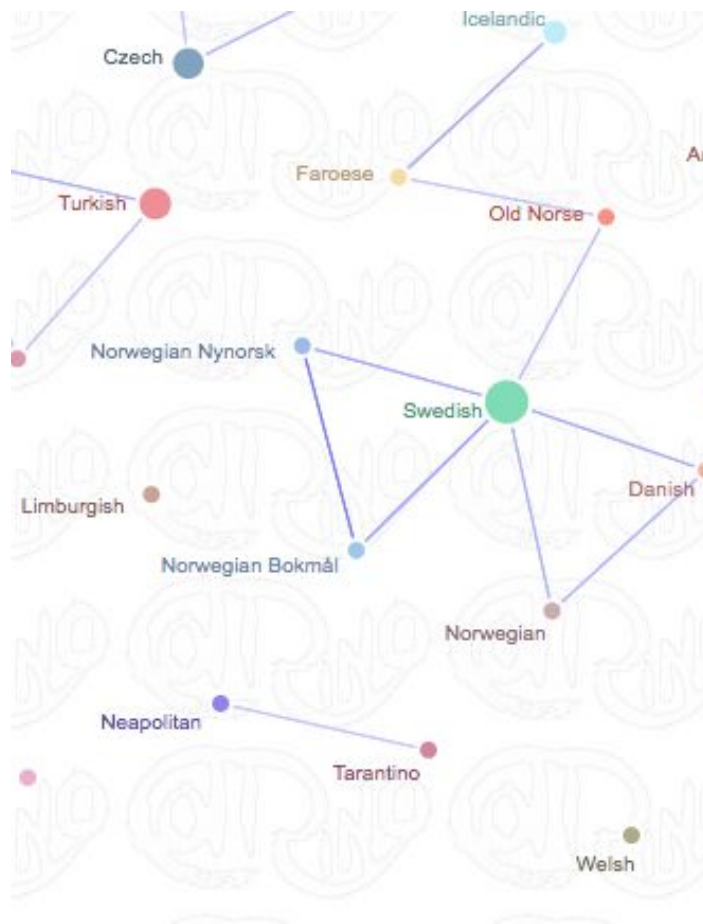


➔ Cyclic



# Tipos de atributos: categóricos

- También conocidos como **nominales**.
- No tienen un orden explícito.
- Podrían, eso sí, ser ordenados de forma arbitraria por datos externos
- Por ejemplo, los países:
  - Por si solos no podemos decir que **Japón > Chile**. Pero podemos usar información externa como la cantidad de habitantes, índice per cápita o nombre alfabético para ordenarlos.



# Tipos de atributos: ordenados

Esto puede ser subdividido en: datos **ordinales** y datos **cuantitativos**.

- En los **datos ordinales**, no existe una aritmética bien definida entre sus componentes, pero sí es posible ofrecer un orden (e.g. tallas de poleras).
- En los **cuantitativos**, existe una magnitud que sí permite una comparación aritmética. Ejemplos: altura, peso, temperatura, fechas, etcétera.

→ Ordered

→ *Ordinal*



→ *Quantitative*



# Dirección de ordenamiento

## ➞ Ordering Direction

➞ Sequential



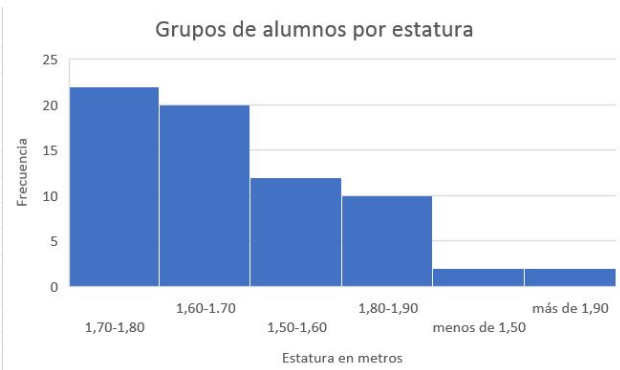
➞ Diverging



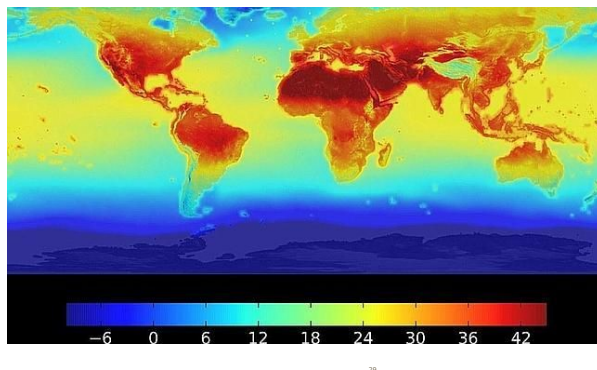
➞ Cyclic



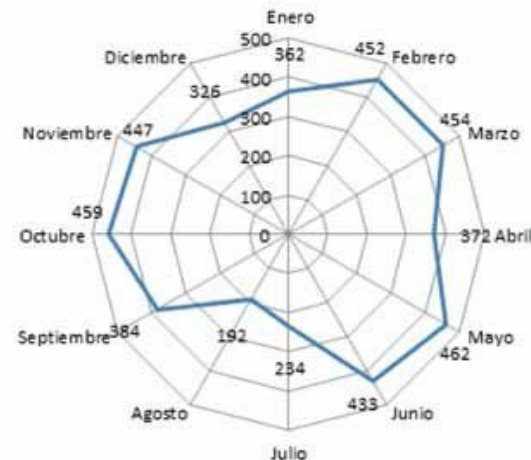
# Dirección de ordenamiento



**Secuenciales**, en donde existe un rango homogéneo desde un valor mínimo hasta uno máximo. Ejemplo: la frecuencia de un dato.



**Divergentes**, que puede ser descompuesto en dos secuencias que van en direcciones opuestas. Ambas se encuentran en un punto en común. Ejemplo: la temperatura en celsius.



**Cíclicos**, en donde los valores vuelven hacia el punto inicial, en vez de crecer indefinidamente. Ejemplo: los meses del año.

# Tipos de atributos

- Atributos jerárquicos:
  - Existe una estructura jerárquica entre uno o múltiples atributos.
  - Por ejemplo: Los precios de acciones recolectadas a lo largo de una década, donde **el tiempo** representa un **atributo jerárquico**.
    - La fecha la podríamos ver en días, meses, trimestre, cuatrimestre, semestre, años, décadas, siglos, etc.

# Tipos de atributos

- 🤔 ¿Llave o Valor?
  - **Llave**: atributo utilizado como índice para encontrar los atributos de valor. Se pueden ver como datos categóricos únicos por ítem.
  - **Valor**: atributos que fue medido y puede estar repetido en diferentes ítems.

ID	Nombre	Edad	BMI*	Talla polera	Serie preferida
1	Santiago	20	22.3	L	Los Simpsons
2	Felipe	15	19.1	M	Game of Thrones
3	Francisca	25	15	M	HxH
...	...	...	...	...	...

# ¿Se puede abstraer más los datos?

Según el *framework* de Tamara Munzner ya es suficiente abstracción. Pero otros expertos ofrecen más alternativas.

Qué pasa si tenemos estos datos:

- “Me gusta el anime”.
- “Salgo de casa todos los días”.
- “No me gusta el curso de visualización”.

Tenemos otro tipo de atributo: **el textual**

Con alta probabilidad, tendremos que aplicar una transformación a estos datos para visualizarlos.



# ¿Se puede abstraer más los datos? (II)

¿Hay diferencia entre estos 2 atributos?

Cantidad de alumnos                      VS                      Peso de los alumnos

- Ambos son atributos cuantitativos con orden secuencial.
- Pero...

# ¿Se puede abstraer más los datos? (II)

¿Hay diferencia entre estos 2 atributos?

Cantidad de alumnos                      VS                      Peso de los alumnos

- Ambos son atributos cuantitativos con orden secuencial.
- Pero... El primero es **discreto**, mientras que el segundo es **continuo**.
  - **Discreto**: No hay valores entremedio de los atributos discretos... No tendremos 1.5 alumnos.
  - **Continuo**: Hay valores intermedios de los atributos continuos... Podemos tener datos como 40.1 kilos, 40.05 kilos, 40.005 kilos, etc.

# ¿Cuál es la ventaja de abstraernos?

- Caso 1: Valor de distintas monedas en el tiempo. Ejemplo: dólar, yen, etc.
  - Valor: cuantitativo - discreto - secuencial.
  - Moneda: categórico.
  - Tiempo: cuantitativo - continuo - secuencial y jerárquico.
- Caso 2: Ganancias por diferentes productos en una empresa en el tiempo. Ejemplo: ganancias por venta de televisores, ventas de celulares, etc.
  - Ganancia: cuantitativo - discreto - secuencial.
  - Producto: categórico.
  - Tiempo: cuantitativo - continuo - secuencial y jerárquico.
- **Distintos datos, pero se abstraen de la misma forma.**

# Derivación de datos

Proceso de transformación de datos iniciales a otra forma, y realizar su abstracción una vez transformado.

- Temperatura 🌡️

25°C	-14°C	27°C	11°C	40°C	(Cuantitativo continuo)
Cálido	Muy Frío	Caluroso	Frío	Caluroso	(Ordinal)
Seguro	Peligroso	Seguro	Seguro	Peligroso	(Categórico)



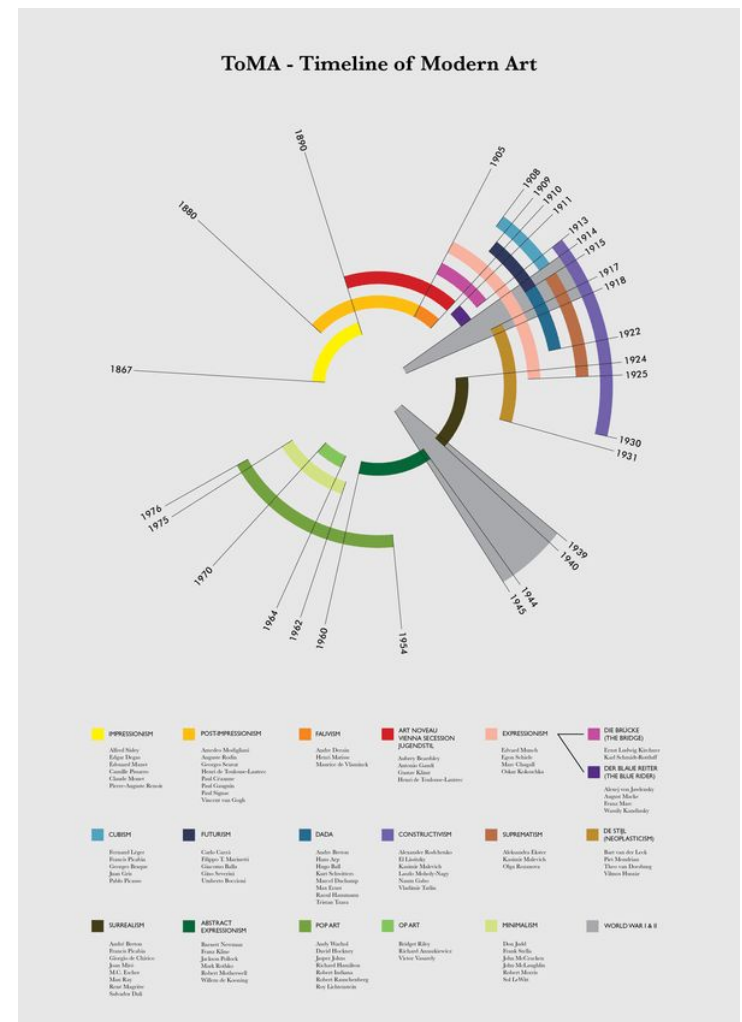
# Análisis de casos

## ¿Marcas y canales?

- **Marca:** Arcos, líneas, segmento de circunferencia.
- **Canales:** Colores, posición, largo

## Para reflexionar... ¿Qué ocurre con los principios de eficiencia y expresividad?

- ¿Los canales y marcas representan adecuadamente la semántica de los datos?
- ¿Los canales utilizados son los más efectivos para los atributos identificados?



# Antes de salir... Revisión de contenidos (RC)

Se acaba de publicar un mini control de alternativas en Canvas sobre lo que **vimos en la clase de hoy**.

- **Duración:** 2 semanas (sin considerar semana de receso) para realizarlo a partir de hoy. Una vez terminado el plazo, tendrán retroalimentación por cada pregunta.
- **Intentos para responder:** ilimitados.
- **Extensión:** 7 preguntas de 1 punto c/u.
- **Condición para obtener el punto RC:** Al menos 6 puntos de 7.
- Cada vez que respondan, verán el puntaje total logrado, pero no cuáles preguntas están correctas e incorrectas.

# Próximos eventos

## Próxima clase (de vuelta de la semana de receso)

- Percepción, Visión y eficiencia de canales

## Próxima ayudantía (de vuelta de la semana de receso)

- Analizar de visualización respecto a marcas, canales y datos utilizados.

## Tarea 4

- Se publica el **miércoles 10 de mayo** a las 20:00.
- Consiste en analizar visualizaciones.
- Será con corrección automatizada.



---

# IIC2026

## Visualización de Información

— Hernán F. Valdivieso López —  
(2023 - 1 / Clase 15)

---