

Visualización de información y analítica visual

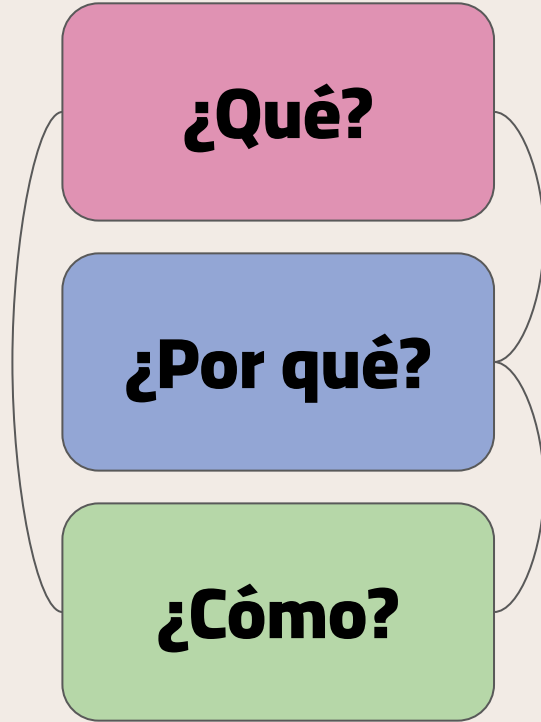
Clase 3: Abstracción de datos y tareas, modelo de validación



Outline de la clase de hoy

1. **Abstracción de datos**
2. **Abstracción de tareas**
3. **Cuatro niveles de validación**





¿Qué datos está viendo el usuario?

¿Por qué el usuario está usando esta visualización?

¿Cómo están contruidos los elementos y codificaciones visuales?

1. Abstracción de datos

Abstracción de datos

100, 150, 120, 340, 1, 1000



Abstracción de datos

100, 150, 120, 340, 1, 1000

3 puntos en el plano XY:

(100, 150), (120, 340), (1, 1000)

Abstracción de datos

100, 150, 120, 340, 1, 1000

**2 puntos en el plano XYZ:
(100, 150, 120), (340, 1, 1000)**

Abstracción de datos

100, 150, 120, 340, 1, 1000

Una variable en el tiempo

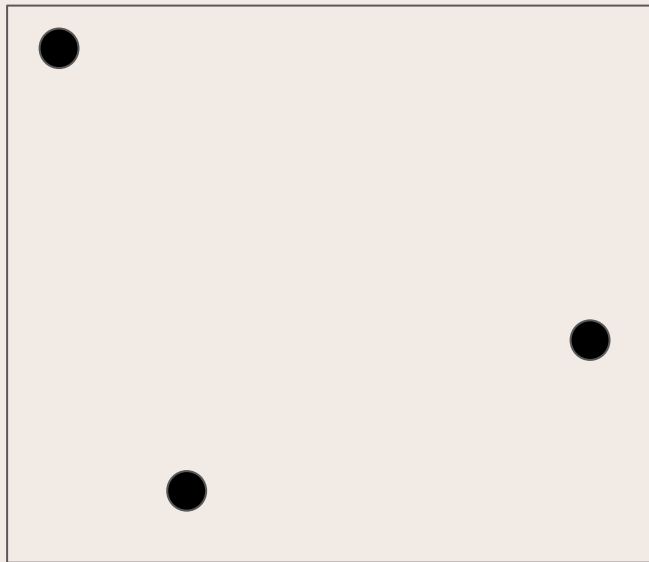
100, 150, 120, 340, 1, 1000

Abstracción de datos

Muchos aspectos del diseño de visualizaciones están dados por el tipo de datos con el que se trabaja

3 puntos en el plano XY:

$(100, 150)$, $(120, 340)$,
 $(1, 1000)$

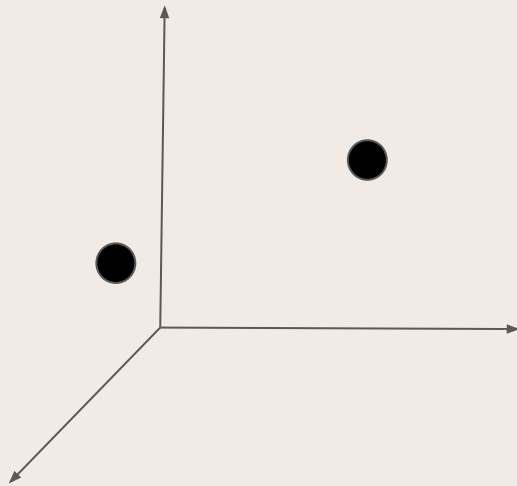


Abstracción de datos

Muchos aspectos del diseño de visualizaciones están dados por el tipo de datos con el que se trabaja

2 puntos en el plano XYZ:

$(100, 150, 120),$
 $(340, 1, 1000)$

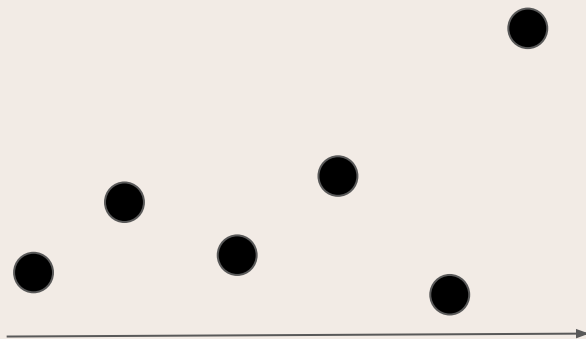


Abstracción de datos

Muchos aspectos del diseño de visualizaciones están dados por el tipo de datos con el que se trabaja

Una variable en el tiempo

100, 150, 120, 340, 1, 1000



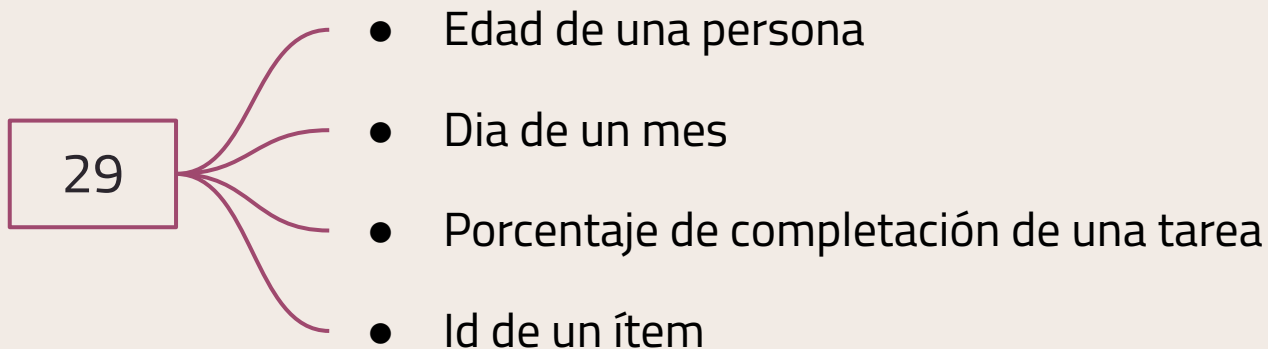
Abstracción de datos

Muchos aspectos del diseño de visualizaciones están dados por el tipo de datos con el que se trabaja, es decir,

La semántica de los datos

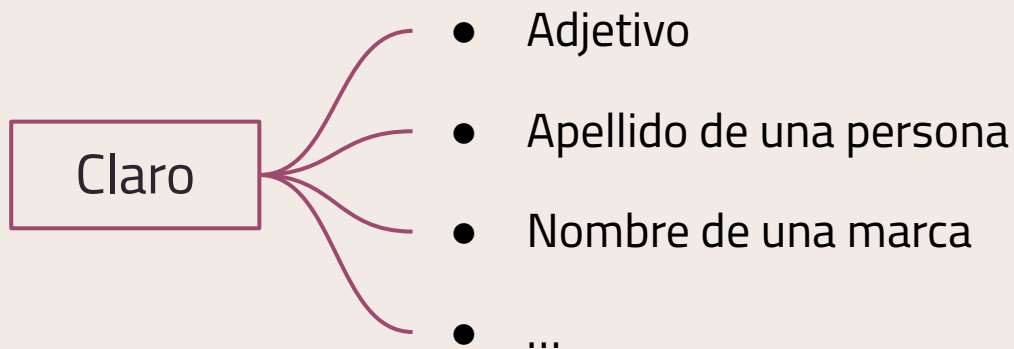
Datos: semántica y tipos

La **semántica** de los datos se refiere a su significado en el “mundo real”



Datos: semántica y tipos

La **semántica** de los datos se refiere a su significado en el “mundo real”



Datos: semántica y tipos

Tamara	E7	17	Pendiente
--------	----	----	-----------

Datos: semántica y tipos

Nombre	Asiento	Valor (usd)	Estatus pago
Tamara	E7	17	Pendiente

Datos: semántica y tipos

La **semántica** de los datos se refiere a su significado en el “mundo real”

El **tipo** de dato es su interpretación estructural o matemática

- **Numerico, Categorico**
- **Atributo**
- **Item**
- **Relación entre ítems**

Datos: semántica y tipos

La semántica y el tipo de datos no son independientes

18723600

- Tipo de dato: numérico
- Semántica del dato: RUT

Sumar números?

Sumar RUTs?

DV de un número?

DV de un RUT?

Tipos de datos singulares

- **Atributo**
- **Item**
- **Enlace**
- **Posición**
- **Grilla**

Tipos de datos singulares

- **Atributo**

Se refiere a una propiedad específica de algo que se puede medir, observar y registrar



- Altura
- Peso
- Color de ojos
- Color de pelo
- Edad
- ...

Tipos de datos singulares

- **Item**

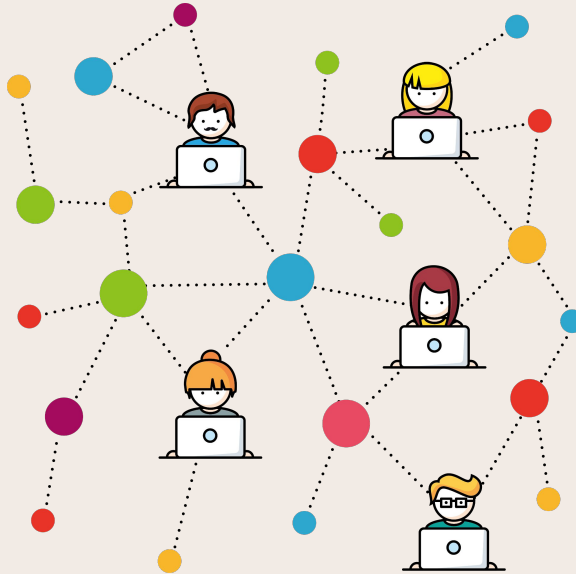
Corresponde a una entidad individual y puede estar descrita con uno o más atributos



Tipos de datos singulares

- **Enlace**

Relación entre dos o más ítems



Tipos de datos singulares

- **Posición**

Ubicación de algo en el espacio (1D, 2D o 3D)

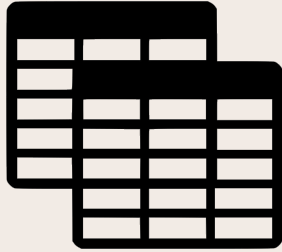


Conjuntos de datos: datasets

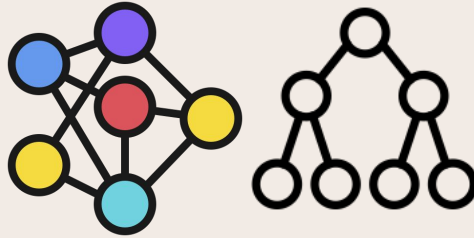
Un dataset es un **conjunto** o colección de **datos singulares**

Existen varios tipos de datasets, pero en este curso nos enfocaremos en tres: tabulares, redes/árboles y geométricos

Tipos de conjuntos de datos



- contiene ítems y sus atributos
- normalmente se representa mediante tablas
- cada fila es un ítem
- cada columna es un atributo



- representa relaciones entre ítems
- compuestos por datos de tipo ítem (nodos)
- y datos de tipo enlace (aristas)

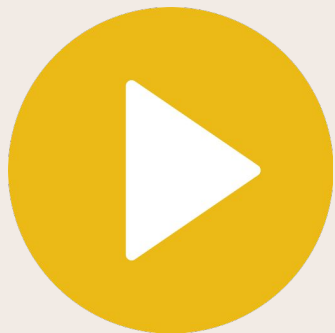


- especifican información sobre formas de los ítems
- y su ubicación (posición)

Tipos de conjuntos de datos



Datasets estáticos



Datasets dinámicos

Tipos de atributos

Atributo: propiedad específica de algo que se puede medir, observar y registrar

La clasificación más general de los atributos es: categóricos y ordenados

Atributos categóricos

Los atributos categóricos son aquellos que no poseen un orden intrínseco

Ejemplo: nombres

Tamara, Francisca, Katalina, Johnny

- Podemos ordenarlos (ej: por orden alfabético), pero ese orden es externo y no es inherente al atributo
- Podemos saber si son iguales o distintos

Atributos ordenados: ordinales y cuantitativos

Los atributos ordenados son aquellos que poseen un orden intrínseco

ORDINALES

- Tienen orden, pero no comparación aritmética exacta

Ej: pequeño, mediano, grande

CUANTITATIVOS

- Tienen orden y comparación aritmética exacta

Ej: altura, precios, etc.

Atributos ordenados: secuenciales y divergentes

Los atributos ordenados son aquellos que poseen un orden intrínseco

SECUENCIALES



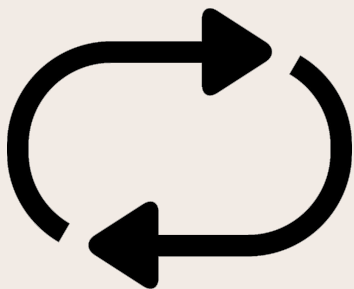
DIVERGENTES



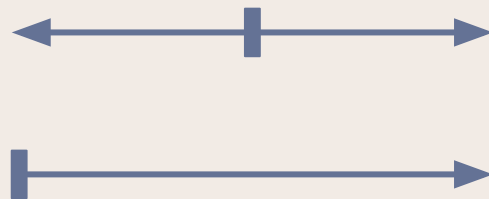
Atributos ordenados: cíclicos y no cíclicos

Los atributos ordenados son aquellos que poseen un orden intrínseco

CÍCLICOS



NO CÍCLICOS



Tipos de atributos: llave y valor

Como (probablemente) ya vieron en cursos de **bases de datos**, hay un tipo especial de atributos que son las **llaves**

Número de fila	Nombre	Edad	N de alumno	Promedio
1	Tamara	27	14638683	4.2
2	Francisca	24	17489364	5.0
3	Johnny	28	154789633	6.0

Tipos de atributos: llave y valor

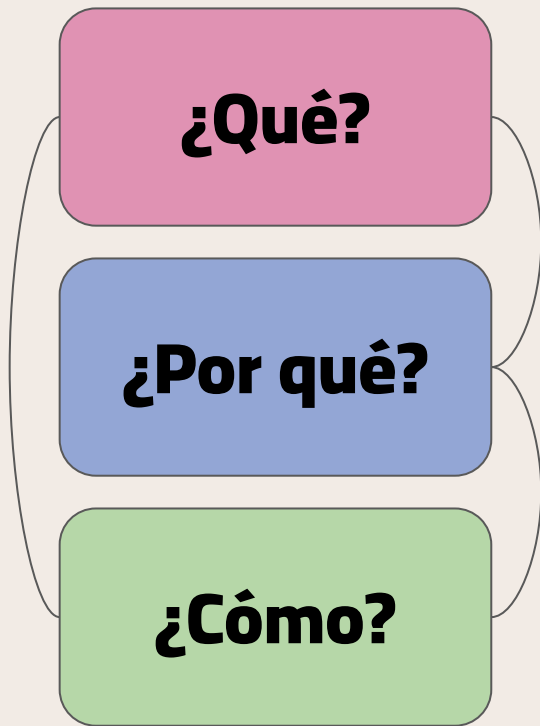
Como (probablemente) ya vieron en cursos de **bases de datos**, hay un tipo especial de atributos que son las **llaves**

Número de fila	Nombre	Edad	N de alumno	Promedio
1	Tamara	27	14638683	4.2
2	Francisca	24	17489364	5.0
3	Johnny	28	154789633	6.0

Ok, tenemos tipos de datos singulares y datasets.. entonces qué es la abstracción de datos?



Abstracción de datos



¿Qué datos está viendo el usuario?

¿Por qué el usuario está usando esta visualización?

¿Cómo están contruidos los elementos y codificaciones visuales?

Abstracción de datos

¿Qué?

¿Qué datos está viendo el usuario?

Abstraer los datos es llevarlos a sus clasificaciones abstractas y así entender mejor cuáles son las marcas y canales más apropiadas para representarlos visualmente

Ej: un dato ordenado divergente puede expresarse con un colormap divergente

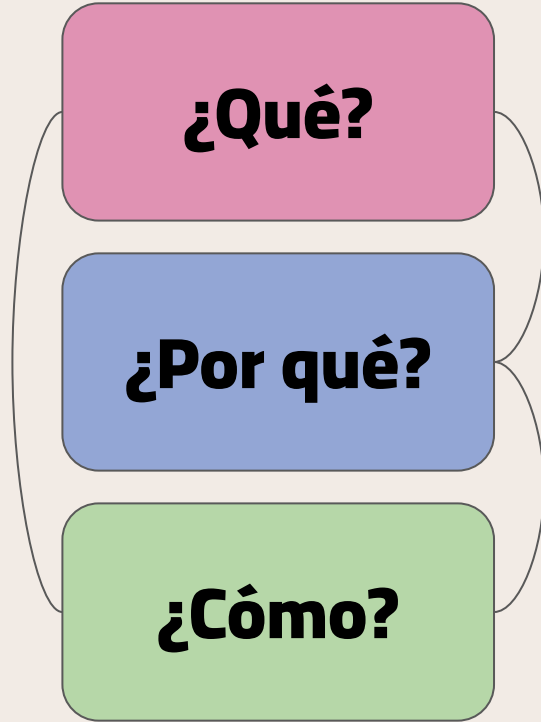
Derivación de datos

Si bien los datos nos van a venir dados de cierta manera, podemos **manipularlos/modificarlos.**

Ejemplo:

- Conversión de escalas de notas
- Temperatura en c° a adjetivos (frío, templado, caluroso, etc)
- Notas a aprobado/reprobado

2. Abstracción de tareas



¿Qué datos está viendo el usuario?

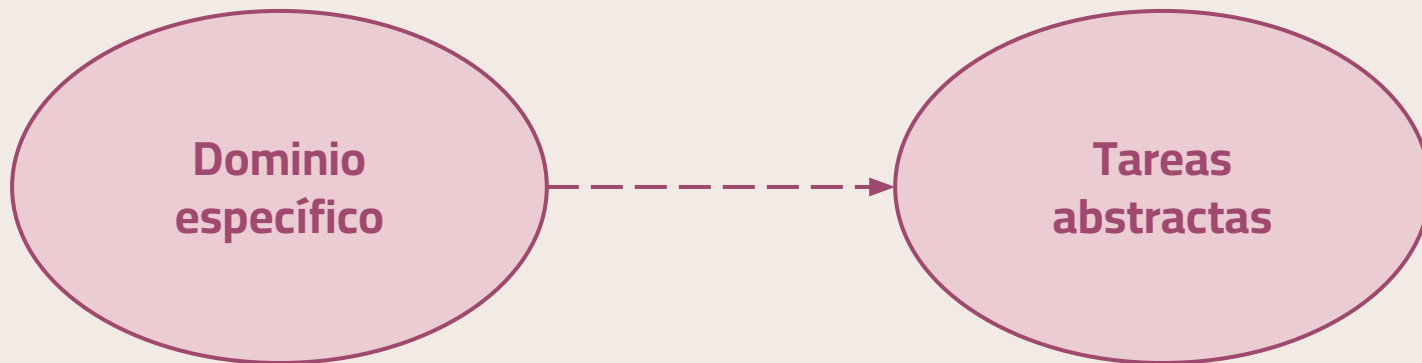
¿Por qué el usuario está usando esta visualización?

¿Cómo están contruidos los elementos y codificaciones visuales?

Abstracción de tareas

¿Por qué?

¿Por qué el usuario está usando esta visualización?



Abstracción de tareas

Un analista de datos del retail quiere revisar si las ventas del CyberDay del año pasado fueron superiores a las de este año

=

Un epidemiólogo quiere contrastar la incidencia de una enfermedad en una ciudad urbana versus una rural

Comparar valores entre dos grupos de datos

Tareas en visualización

Una tarea en visualización tiene la siguiente estructura:

Tarea = acción (verbo) + objetivo (sustantivo) + ...

***Presentar** la **tendencia** de precipitaciones en Santiago a lo largo de un año.*

Acciones

¿Por qué?

¿Por qué el usuario está usando esta visualización?

- Analizar
- Búsqueda
- Consulta

Puede que el usuario quiera hacer más de una acción...

Acción: analizar

-Consumir-

- Descubrir
- Presentar
- Disfrutar

-Producir-

- Anotar
- Registrar
- Derivar

Acción: analizar

-Consumir-

- **Descubrir**
Aprender nueva información
- Presentar
- Disfrutar

-Producir-

- Anotar
- Registrar
- Derivar

Acción: analizar

-Consumir-

- Descubrir
Aprender nueva información
- **Presentar**
Usar la visualización como
método de comunicación
- Disfrutar

-Producir-

- Anotar
- Registrar
- Derivar

Acción: analizar

-Consumir-

- Descubrir
Aprender nueva información
- Presentar
Usar la visualización como
método de comunicación
- **Disfrutar**
Alimentar la curiosidad del
usuario

-Producir-

- Anotar
- Registrar
- Derivar

Acción: analizar

-Consumir-

- Descubrir
Aprender nueva información
- Presentar
Usar la visualización como
método de comunicación
- Disfrutar
Alimentar la curiosidad del
usuario

-Producir-

- **Anotar**
El usuario adiciona texto o
elementos gráficos
- Registrar
- Derivar

Acción: analizar

-Consumir-

- Descubrir
Aprender nueva información
- Presentar
Usar la visualización como
método de comunicación
- Disfrutar
Alimentar la curiosidad del
usuario

-Producir-

- Anotar
El usuario adiciona texto o
elementos gráficos
- **Registrar**
Guardar elementos de las
visualizaciones
- Derivar

Acción: analizar

-Consumir-

- Descubrir
Aprender nueva información
- Presentar
Usar la visualización como
método de comunicación
- Disfrutar
Alimentar la curiosidad del
usuario

-Producir-

- Anotar
El usuario adiciona texto o
elementos gráficos
- Registrar
Guardar elementos de las
visualizaciones
- **Derivar**
Crear nuevos elementos a
partir de la visualización

Acción: búsqueda

El usuario quiere encontrar datos o elementos de interés en la herramienta

		Conozco el elemento	
Conozco la ubicación		Si	No
	Si	<i>Lookup</i>	<i>Browse</i>
	No	<i>Locate</i>	<i>Explore</i>

Accion: consulta

- **Identificar**

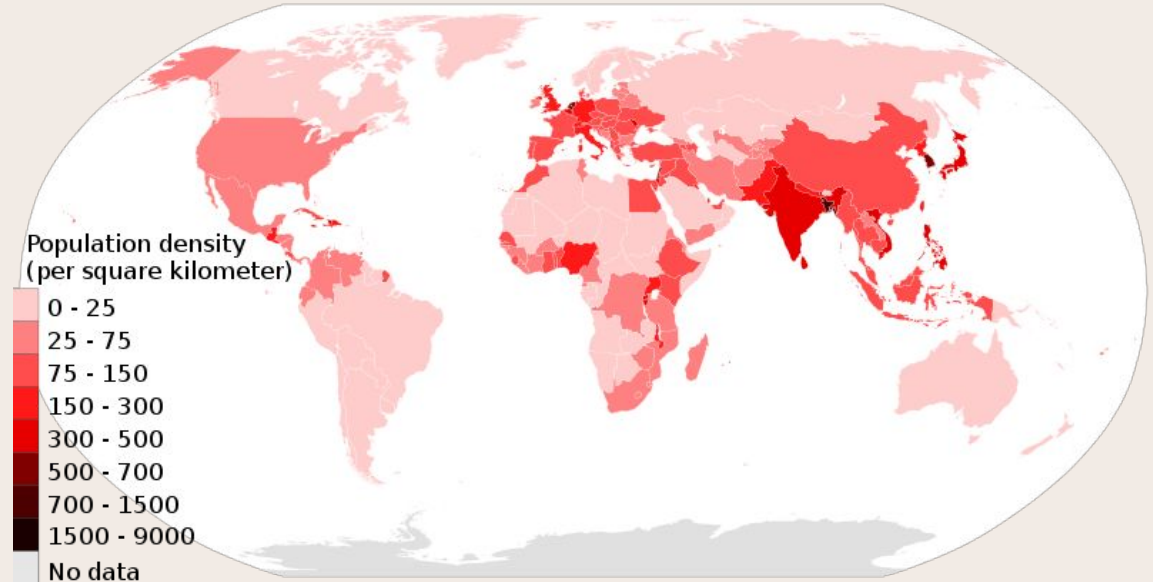
Un objetivo

- **Comparar**

Múltiples objetivos

- **Resumir**

Conjunto completo de objetivos



Tareas en visualización

Una tarea en visualización tiene la siguiente estructura:

Tarea = acción (verbo) + objetivo (sustantivo) + ...

***Presentar** la **tendencia** de precipitaciones en Santiago a lo largo de un año.*

Objetivos de tareas

Objetivos aplicables a todo tipo de datos

- Tendencias
- Outliers (partes aisladas)
- Características

Objetivos de tareas

Objetivos aplicables a atributos

Un atributo

- Valor individual
- Distribuciones
- Extremos

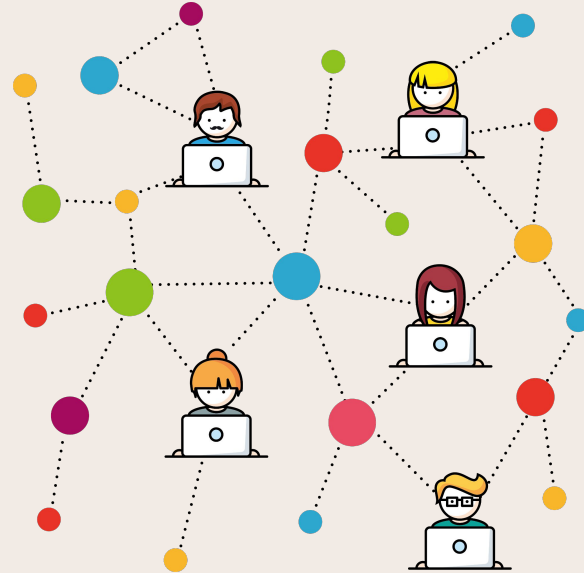
Grupo de atributos

- Dependencia
- Correlación
- Similitud

Objetivos de tareas

Objetivos aplicables a redes

- Topología
- Caminos



Objetivos de tareas

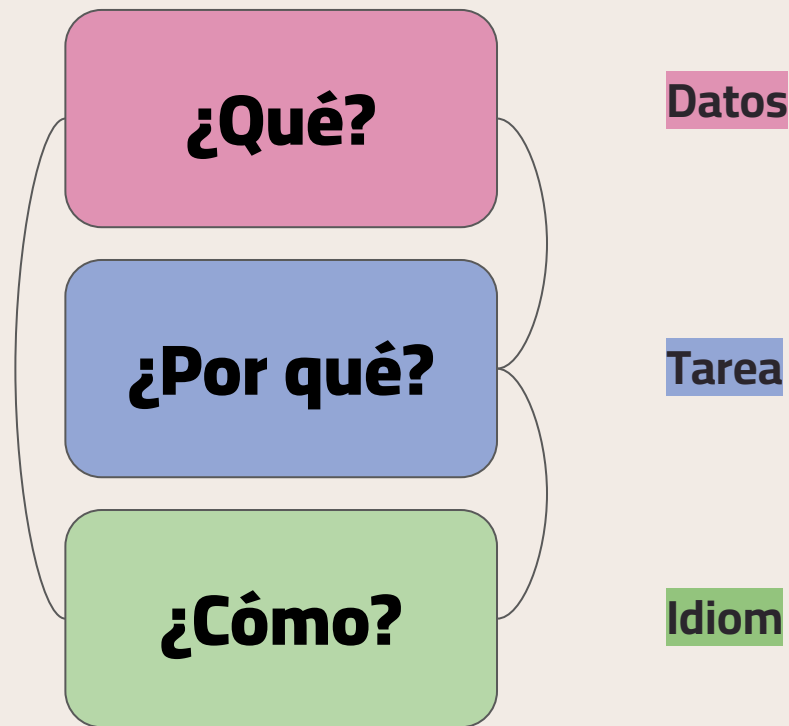
Objetivos aplicables a datasets espaciales o geométricos

- Forma



3. Cuatro niveles de validación

Framework anidado de Tamara Munzner



Podemos describir una
visualización a través de
uno o más triples
(dato, tarea, *idiom*)

Framework anidado de Tamara Munzner

Caracterización del dominio

Abstracción de datos y tareas

Codificación visual y de interacción

Implementación algorítmica

Busca sistematizar el proceso de creación de visualizaciones

Caracterización del dominio

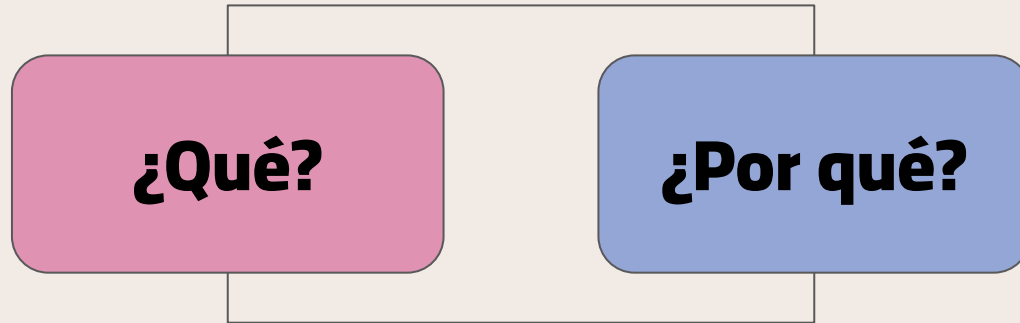
- Entender el campo de conocimiento y contexto
- Entender los datos, el usuario y las tareas específicas



Podemos lograr esto haciendo entrevistas, observaciones e investigación

Abstracción de datos y tareas

- Traducir las tareas y datos específicos a un lenguaje común
- Definir el que y el por que de la herramienta, en este lenguaje común
- Seleccionar los aspectos relevantes a tener en cuenta



Codificación visual y de interacción

- Definir la manera en que vamos a representar los datos en la visualización (definir el *idiom*)
- Decisiones visuales (marcas y canales) y de interacción (Clase 7)

¿Cómo?

Idiom

Implementación algorítmica

- La implementación misma de la visualización (el código)
- Elección de herramientas, librerías, lenguajes de programación
- Tener en cuenta eficiencia, escalabilidad y dificultad



*Existen muchas
herramientas para hacer
visualizaciones...*

Tutorial 2:

Introducción a Streamlit

Visualización de información y analítica visual

Clase 3: Modelo anidado, abstracción de datos y tareas

