



---

# Visualización de Información y Analítica Visual

— Hernán Valdivieso López ([hfvaldivieso@uc.cl](mailto:hfvaldivieso@uc.cl)) —

---

# Resumen clase 1

# ¿Qué es la visualización de información?

Se identifican 3 temas claves:

- **Representación** de un concepto abstracto, como datos.
- Uso de imágenes **visuales** generadas por el computador.
- **Ayudar** a las personas: a entender, a convencer, a realizar acciones eficientemente, etc .

# ¿Para qué creamos visualizaciones?

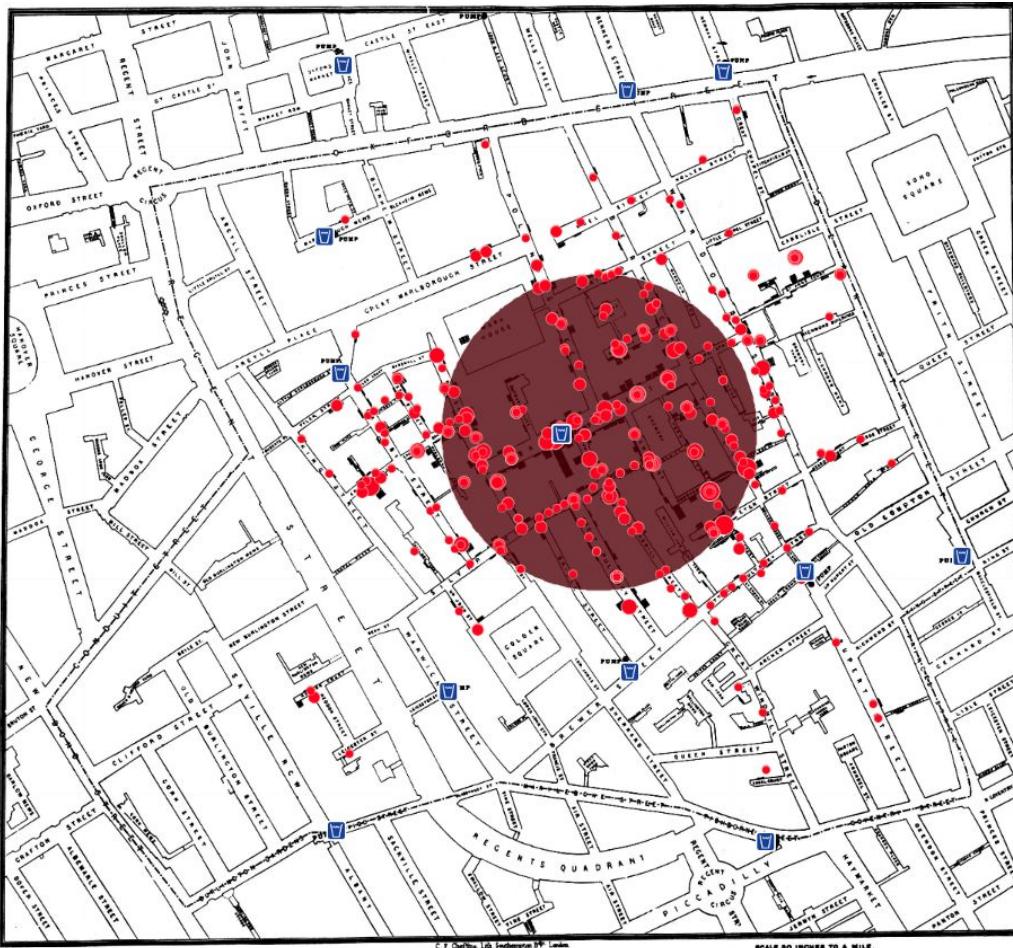
- Comprender las relaciones entre conjuntos de datos
- Entender algo sobre los datos
- Resaltar información importante
- Plantear un argumento convincente
- Podemos encontrar *outliers*
- Descubrir datos faltantes
- Reducir carga cognitiva para procesar información
- Comparar valores de forma efectiva
- Los gráficos son necesarios para explicar y verificar los datos

# ¿Para qué creamos visualizaciones?

- Comprender las **relaciones** entre conjuntos de datos
- Entender algo sobre los datos
- **Resaltar** información importante
- Plantear un argumento convincente
- Podemos encontrar **outliers**
- **Descubrir** datos faltantes
- Reducir carga cognitiva para procesar información
- **Comparar** valores de forma efectiva
- Los gráficos son necesarios para explicar y verificar los datos

# Un poco de historia

- Córera en Londres: John Snow (1854)
  - Durante una epidemia de Córera en Londres, el Dr. John Snow usa un análisis espacial para apoyar su hipótesis
  - Encontró que las muertes eran principalmente por una bomba de agua contaminada

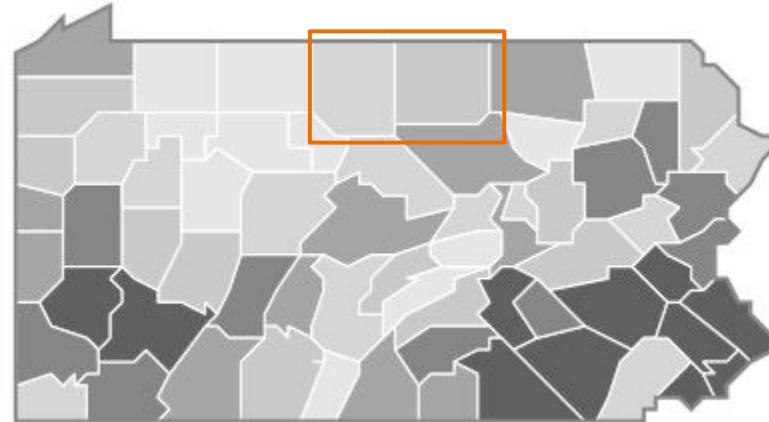


# Percepción

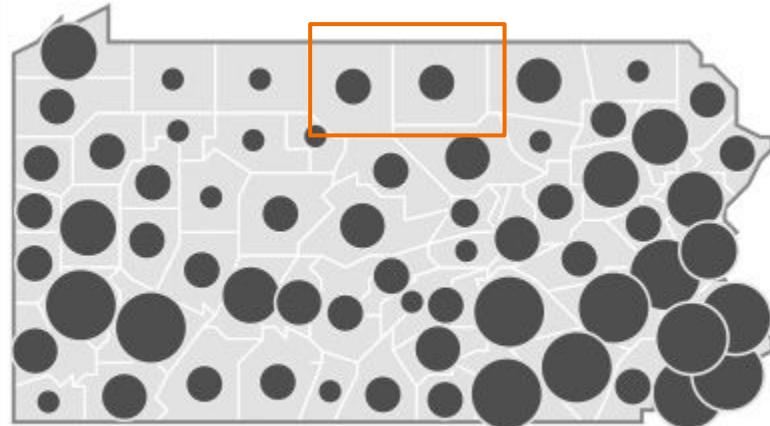


# Percepción

## Estimación de magnitud



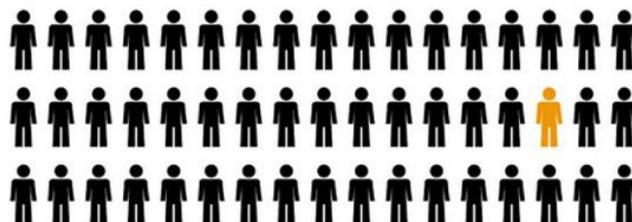
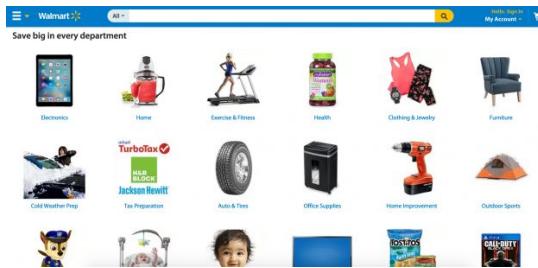
*dark means more, light less*



*large circles mean more, small less*

# Percepción

- Leyes de *Gestalt*
  - Estas leyes describen cómo percibimos a través de grupos
  - Nuestro cerebro aplica varios principios (proximidad, similaridad, cierre, etc.)



# Marcas y canales

Definición de **marca**:

- Es un elemento geométrico básico, que puede ser clasificado según el número de dimensiones espaciales que requiera.

→ Points



→ Lines



→ Areas



→ Containment



→ Connection

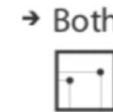
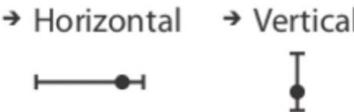


# Marcas y canales

Definición de **canal**:

- Un canal visual permite controlar **la apariencia de las marcas**, independientemente de las dimensiones que contengan.

④ Position



④ Color



④ Shape



④ Tilt



④ Size

→ Length



→ Area



→ Volume



# Marcas y canales

## Principio de expresividad

- Debe haber **coherencia** entre el tipo de canal (magnitud, identidad) con la semántica del atributo (cuantitativo, ordinal, categórico).
  - Los datos ordenados deben ser mostrados de tal forma que nuestro sistema perceptual los perciba como ordenados; inversamente, debe ocurrir lo mismo con los datos no ordenados.

## Principio de efectividad

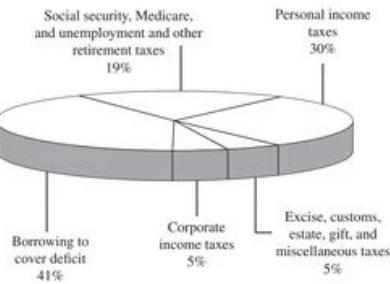
- Dicta que los **atributos más importantes** deben ser codificados con los **canales más efectivos**, para que sean más perceptibles

# Marcas y canales - Ejemplos

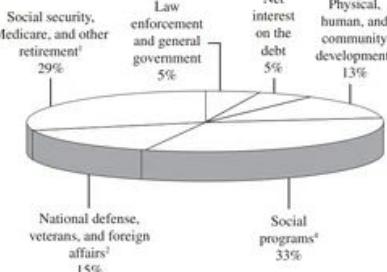
## Major Categories of Federal Income and Outlays for Fiscal Year 2021

Income and Outlays. These pie charts show the relative sizes of the major categories of federal income and outlays for fiscal year 2021.

### Income



### Outlays\*



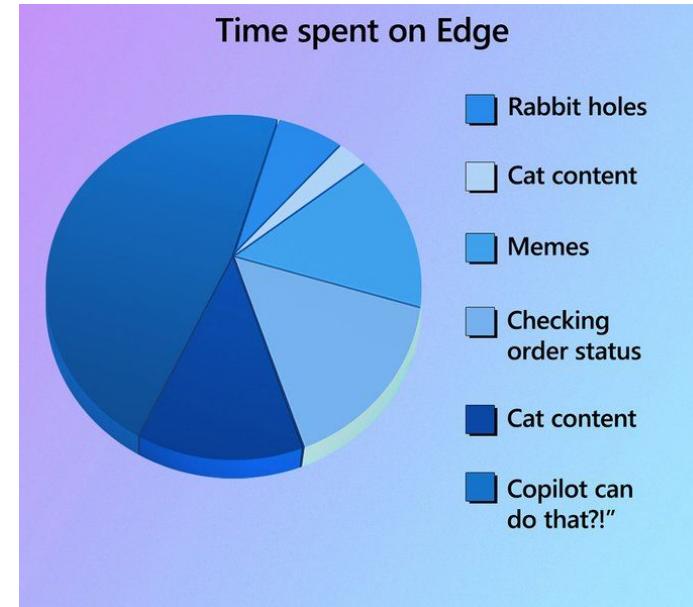
\* Numbers may not total to 100% due to rounding.

El uso de volumen + perceptiva no es el más efectivo para comparar los valores

No cumple el **principio de efectividad**

Se ocupa varias gamas de azules para datos categóricos que no tienen una relación de orden.

No cumple el **principio de expresividad**



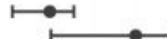
# Efectividad de canales

## ④ Magnitude Channels: Ordered Attributes

Position on common scale



Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



Color saturation



Curvature



Volume (3D size)



## ④ Identity Channels: Categorical Attributes

Spatial region



Color hue



Motion



Shape



▲  
Most

Effectiveness  
↓

Same  
Least

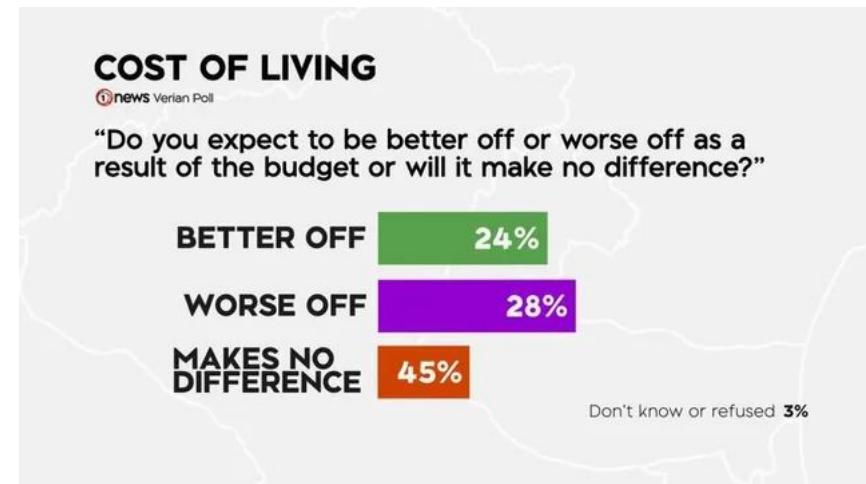
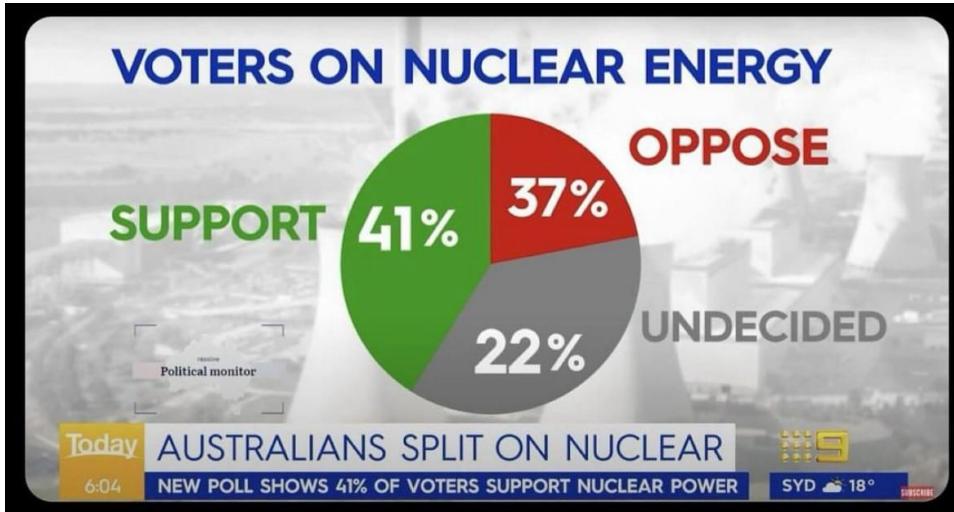
# Efectividad de canales

- 00 El ranking no es absoluto.

Lectura recomendada adicional:

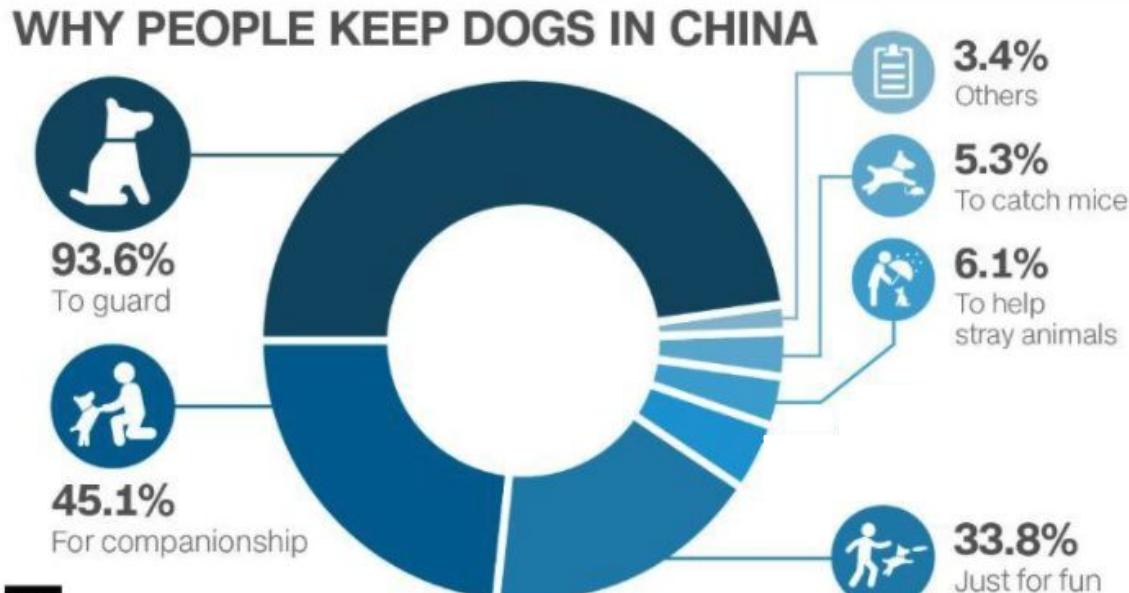
Bertini, Enrico, Michael Correll, and Steven Franconeri. "Why shouldn't all charts be scatter plots? Beyond precision-driven visualizations." 2020 IEEE Visualization Conference (VIS). IEEE, 2020.

# Efectividad de canales



Fuente: [Right-wing media forgets how pie charts work : r/dataisugly](#) y [In New Zealand 45% is half the size of 24% : r/dataisugly](#)

# Efectividad de canales



# Efectividad de canales



# Efectividad de canales

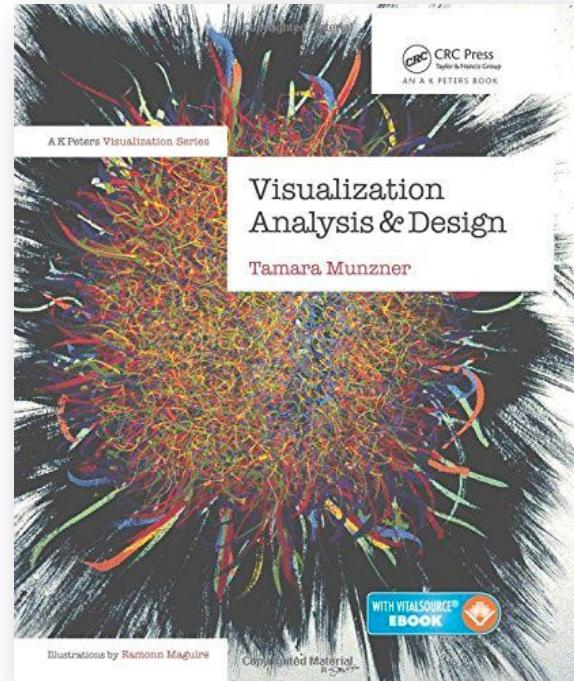


# Efectividad de canales



# Efectividad de canales

- Construir visualizaciones efectivas **no es trivial.**
- Existen *frameworks* (metodologías) que explican los pasos para obtener una visualización efectiva.
- Nos basaremos en el *Framework* de Tamara Munzner (2014).



## Clase 2: *Framework* anidado de Tamara Munzner

# Contenidos

- *Framework*
- Abstracción de datos (*What*)
- Abstracción de tareas visuales (*Why*)
- Desarrollo taller 1

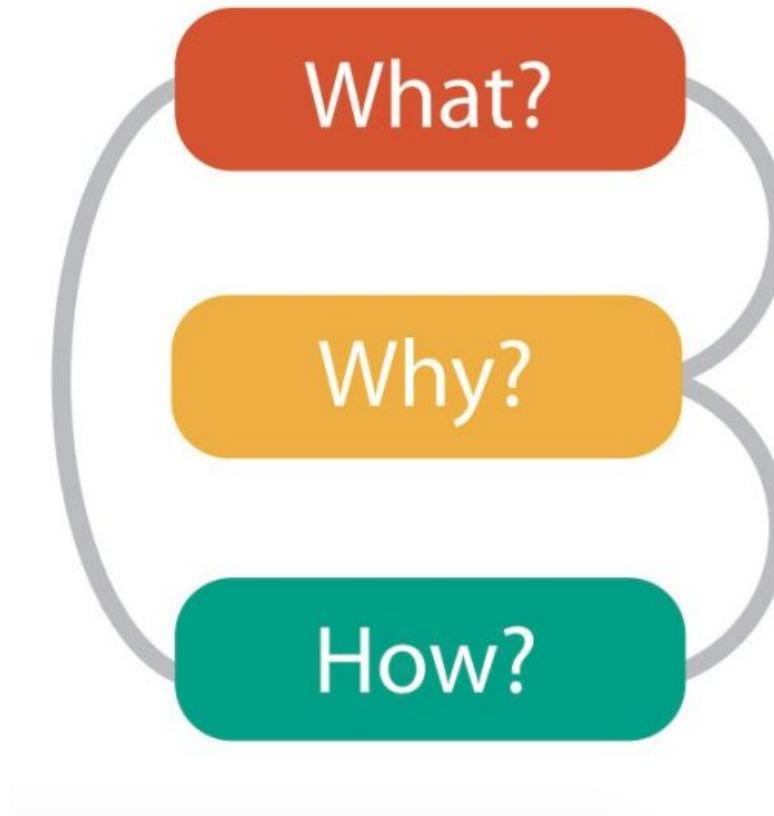
# **Framework para crear visualizaciones**

## **¿Por qué necesitamos un *framework*?**

- Cuando nos enfrentamos con un problema de visualización, debemos decidir una combinación de opciones dentro de un **enorme espacio posibilidades** que comunique mejor los datos.
- Para poder analizar estas opciones es necesario tener una **estructura** que nos permita entender las diferencias y ventajas de las diferentes combinaciones.

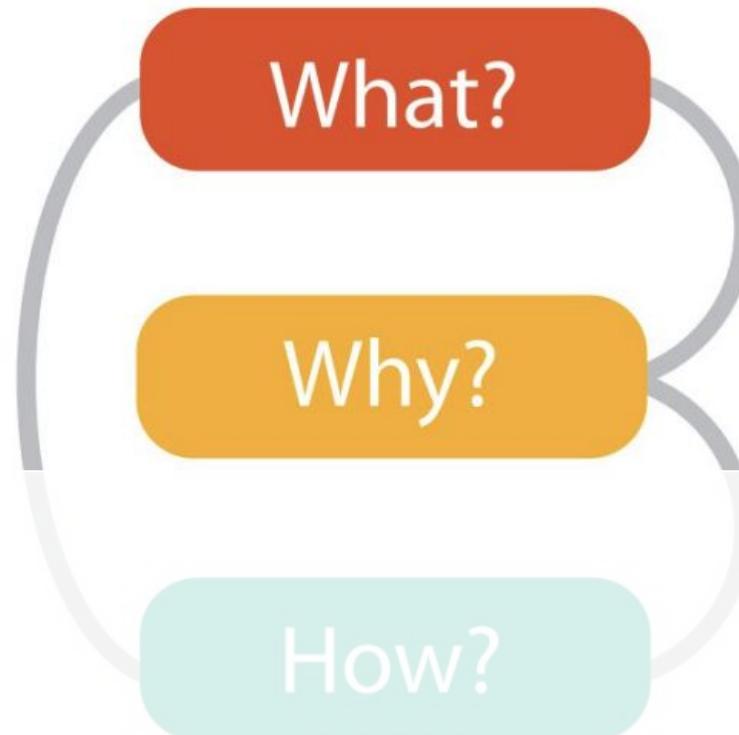
# Munzner's Framework

- **What** qué información se está mostrando (*data*).
- **Why** por qué queremos visualizar esto (*task*).
- **How** cómo se construye la visualización (*idiom*).



# Munzner's Framework

- **What** qué información se está mostrando (*data*).
- **Why** por qué queremos visualizar esto (*task*).
- **How** cómo se construye la visualización (*idiom*).



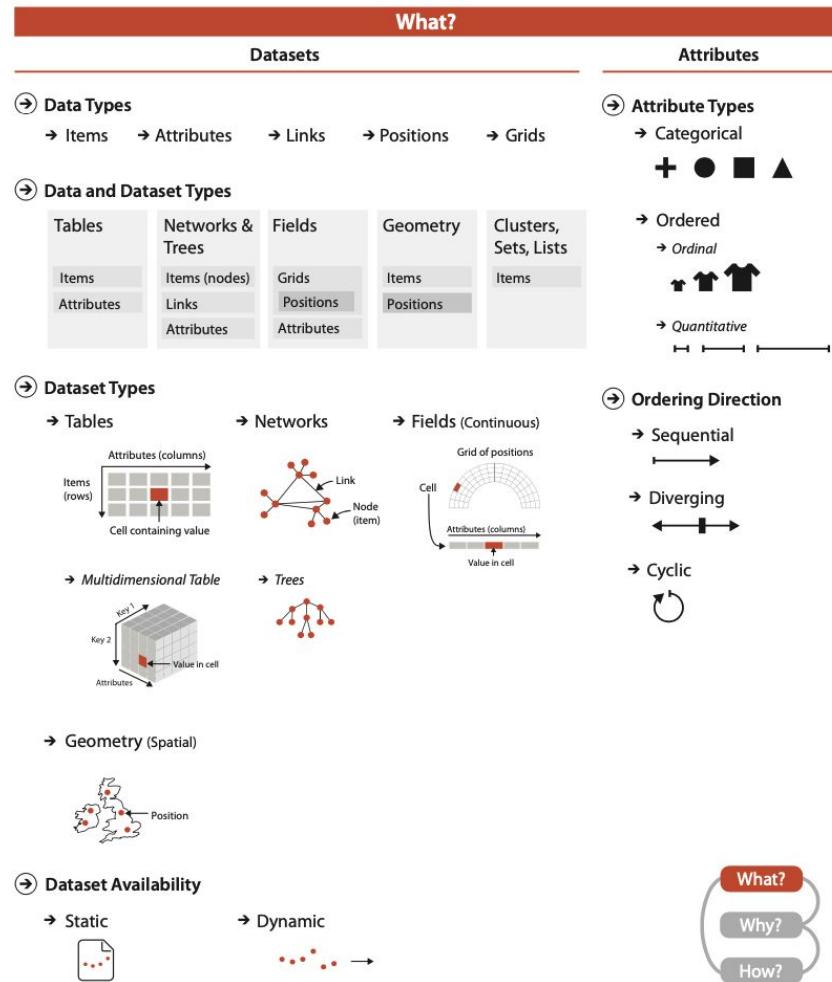
*What*

Abstracción de datos

---

# ¿What? Abstracción de datos

1. ¿Qué **semántica** tienen los datos (significado en el mundo real)?
2. ¿Qué **tipo de datos**/set de datos tenemos?
3. ¿Qué tipo de **disponibilidad** tienen los datos?
4. ¿Qué **tipo de atributos** tienen los datos?



# ¿Qué? Abstracción de datos

Semántica y tipos  
de **datos**

Tipos de  
***datasets***

**Disponibilidad**  
del *dataset*

Tipos de  
**atributos**

# ¿Qué? Abstracción de datos

Semántica y tipos  
de **datos**

Tipos de  
*datasets*

Disponibilidad  
del *dataset*

Tipos de  
atributos

# 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 cuestionar, entonces:
  - **¿Qué tipo** de datos tenemos?
  - **¿Qué información** podemos obtener directamente?
  - **¿Qué sentido** tienen realmente?

# Semántica de los datos

1, Santiago, 20, 22.3, L, Avatar

# Semántica de los datos

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

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

# *Data Types (tipos de datos singulares)*

- 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	Avatar
2	Felipe	15	19.1	M	Game of Thrones
3	Monica	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	Avatar
2	Felipe	15	19.1	M	Game of Thrones
3	Monica	25	15	M	HxH
...	...	...	...	...	...

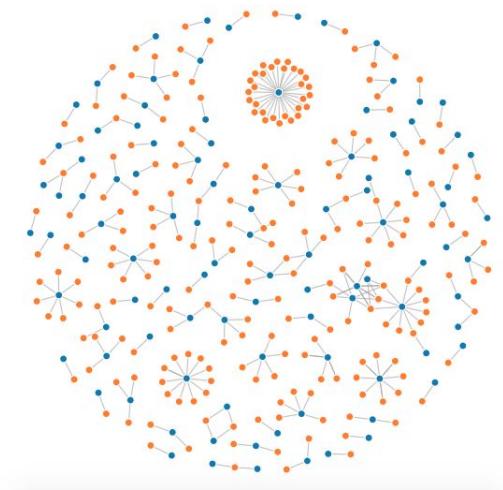
# Vínculos o enlace

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.

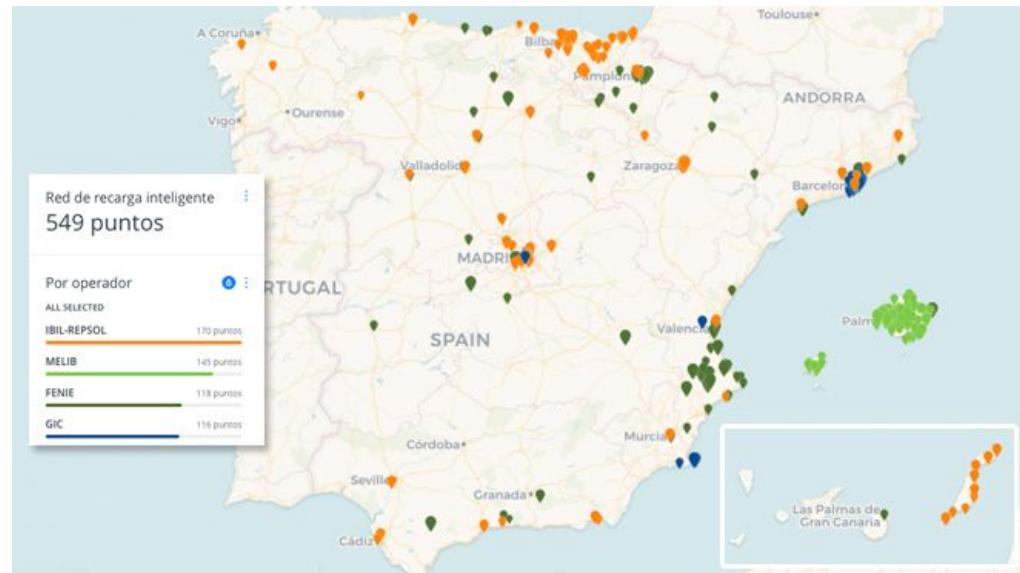
- Por ejemplo: **Comentario en youtube.** Un comentario permite relacionar el ítem “usuario” con el ítem “video”. Pero del mismo modo, el comentario por si solo es un ítem con atributos que tiene:
  - Fecha de creación.
  - Cantidad de *likes*.
  - Contenido.



# Posiciones

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

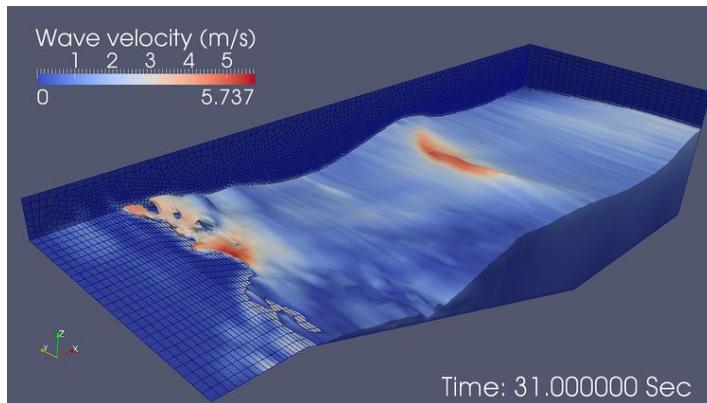
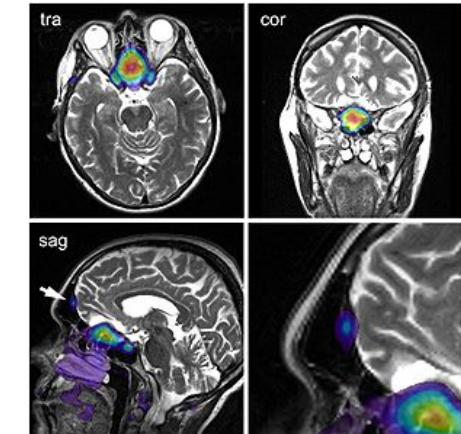
- Un par latitud-longitud mostrando una ubicación en la Tierra,
- La ubicación en la región de un escáner médico.
- La posición de las mesas para una boda.
- Posición de los íconos en una pantalla de computador.



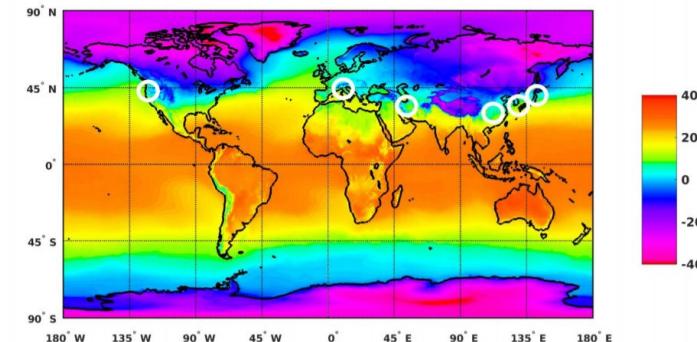
# Grilla (no profundizaremos en el curso)

Ejemplos:

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



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



Fuente: [Fluidos](#), [Scanner médico](#),

# Ejemplos de la importancia de la semántica

- (Santiago, Felipe)
  - Opción 1: **ítem** donde Santiago es nombre y Felipe es segundo nombre.
  - Opción 2: **enlace** donde Santiago es la ciudad y Felipe vive en esa ciudad.
  - Opción 3: **enlace** donde Santiago es una persona con una **amistad** a otra persona: Felipe
- (5, 4, 7)
  - Opción 1: **ítem** de 3 atributos: Persona de ID 5 compró 4 películas y 7 juegos.
  - Opción 2: **posición** en 3 dimensiones (X, Y, Z)
  - Opción 3: **enlace** que conecta Persona "5" es amiga de "4" y se hicieron amigos en la ciudad "7".

La **semántica** que le demos al dato, nos dará indicio de **qué tipo de dato es**.

# ¿Qué? Abstracción de datos

Semántica y tipos  
de **datos**

Disponibilidad  
del *dataset*

Tipos de  
***datasets***

Tipos de  
atributos

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

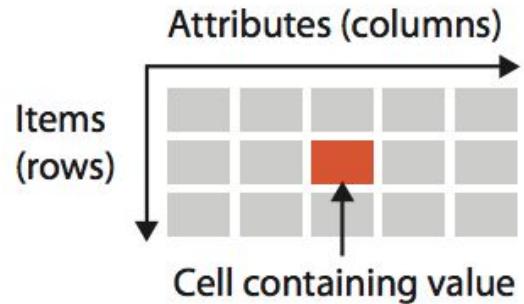
Principales tipos de *datasets* son:

- Tabulares.
- Redes y árboles.
- Geométricos.

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

# Tabular

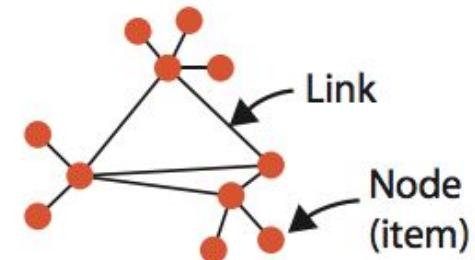
- Es el tipo de *dataset* más común.
- Viene en forma de filas y columnas (e.g. *spreadsheet*, *excel*).
- Los tipos de datos son: **ítems y atributos**.
- Cada celda de la tabla es un valor para la combinación ítem-atributo.



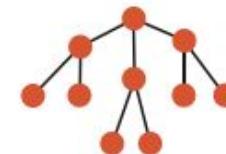
# Redes y árboles

→ Networks

- 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.



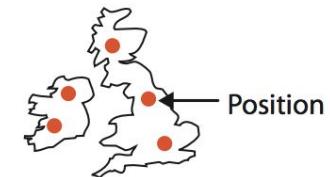
→ Trees



# Geométricos

- Los ítems pueden ser posiciones, curvas, superficies o volúmenes.
- **Ítems con posiciones** explícitas.
- Este tipo de *dataset* puede que no tenga atributos adicionales aparte de la posición.
- En caso de un *dataset* con información del terreno (geojson por ejemplo), es necesario saber con qué nivel de detalle se generan cada forma
  - A nivel de país, nivel de ciudad, nivel de continente, etc.

→ 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.) puede ser tabular y geométrico.

# 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.) puede ser tabular y geométrico.
- **Lo importante, es evaluar cuál será el enfoque principal que le daremos al *dataset*: tabular, geométrico y/o como red.**

# ¿Qué? Abstracción de datos

Semántica y tipos  
de **datos**

**Disponibilidad**  
del *dataset*

Tipos de  
*datasets*

Tipos de  
atributos

# 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**



# ¿Qué? Abstracción de datos

Semántica y tipos  
de **datos**

Tipos de  
***datasets***

Disponibilidad  
del ***dataset***

Tipos de  
**atributos**

# Atributos

- Tipos de atributos
- Dirección de ordenamiento

## ➔ Attribute Types

→ Categorical



→ Ordered

→ *Ordinal*



→ Quantitative



## ➔ Ordering Direction

→ Sequential



→ Diverging

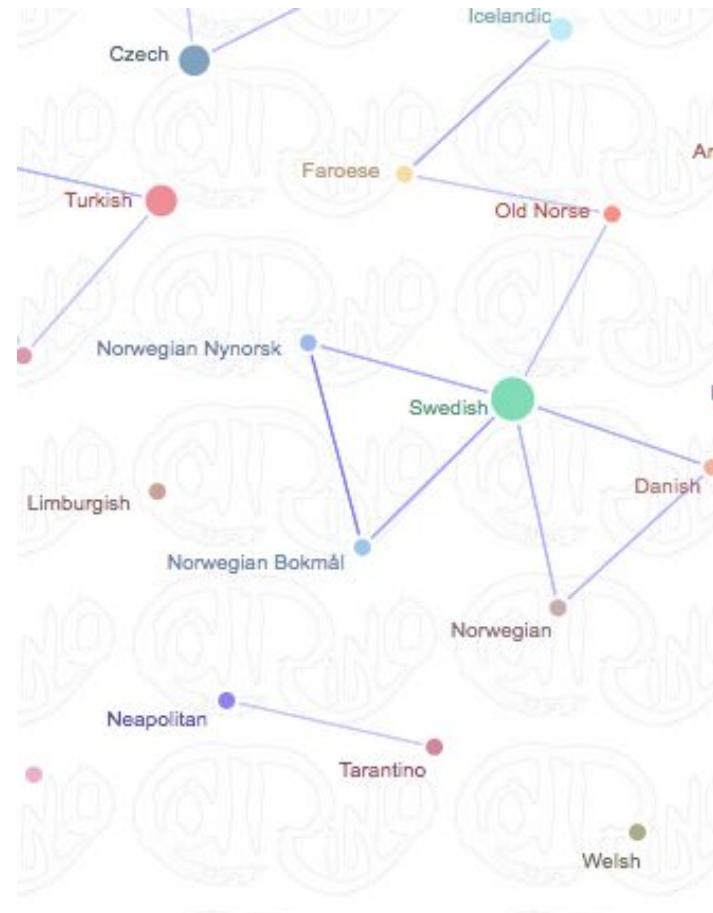


→ Cyclic



# 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



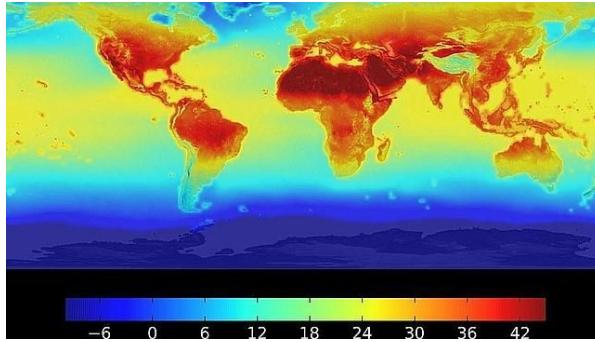
➔ Diverging



➔ Cyclic



# Dirección de ordenamiento

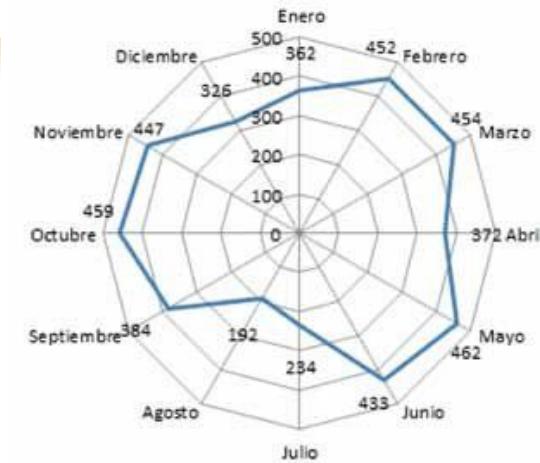


**Secuenciales**, dado un valor mínimo o máximo, el valor solo aumenta o disminuye. Hay una "única" dirección.

**Ejemplo:** la frecuencia de un dato, el peso de las personas, volumen de la música.

**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, altura a nivel del mar.



**Cíclicos**, en donde los valores vuelven hacia el punto inicial, en vez de crecer indefinidamente.

**Ejemplo:** los meses del año, latitud/longitud, los días de la semana.

# Tipos de atributos

- Atributos jerárquicos:
  - Existe una estructura jerárquica entre uno o múltiples atributos, es decir, es un atributo que permite agrupar los datos en un nivel superior conocido.
  - 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.
  - **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	Avatar
2	Felipe	15	19.1	M	Game of Thrones
3	Monica	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 jazz”.
- “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.

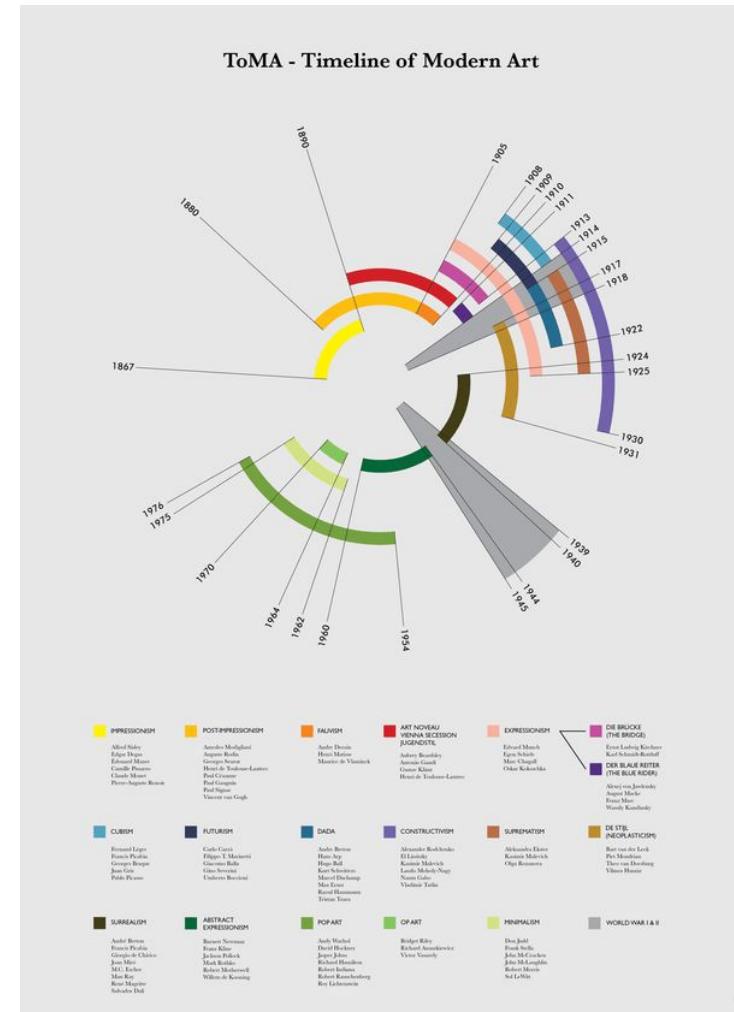
# ¿Cuál es la ventaja de abstraernos?

- **Caso 1:** Población por país en el tiempo después del año 0.
  - **Población:** cuantitativo - discreto - secuencial.
  - **País:** categórico.
  - **Tiempo:** cuantitativo - continuo - secuencial y jerárquico.
- **Caso 2:** Ganancias por diferentes productos en una empresa en el tiempo.
  - **Ganancia:** cuantitativo - discreto - secuencial.
  - **Producto:** categórico.
  - **Tiempo:** cuantitativo - continuo - secuencial y jerárquico.
- **Distintos datos, pero se abstraen de la misma forma.**

# Análisis de casos

La siguiente visualización ocupa un gráfico de barra radial para mostrar la línea de tiempo de diferentes expresiones del arte moderno. Las zonas en gris representan las guerras mundiales.

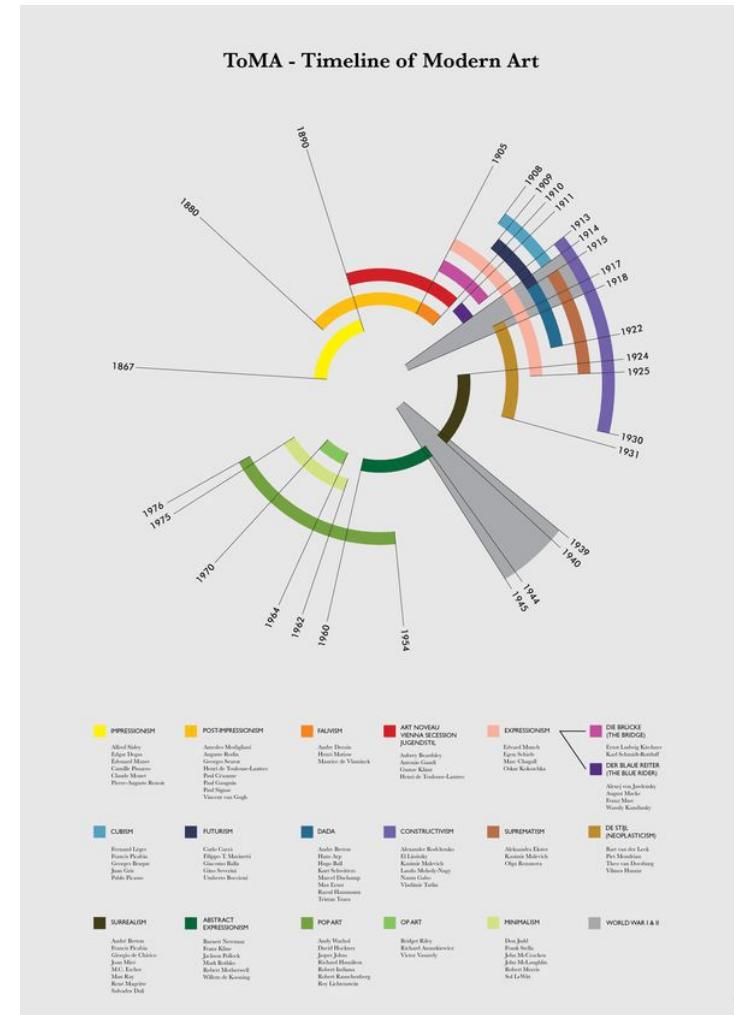
- ¿Qué datos tenemos?
  - Arte:
  - Año de inicio y término:
  - Artistas por arte:



# Análisis de casos

La siguiente visualización ocupa un gráfico de barra radial para mostrar la línea de tiempo de diferentes expresiones del arte moderno. Las zonas en gris representan las guerras mundiales.

- ¿Qué datos tenemos?
  - Arte: Categórico
  - Año de inicio y término: Cuantitativo con orden secuencial. Jerárquico.
  - Artistas por arte: Categórico.



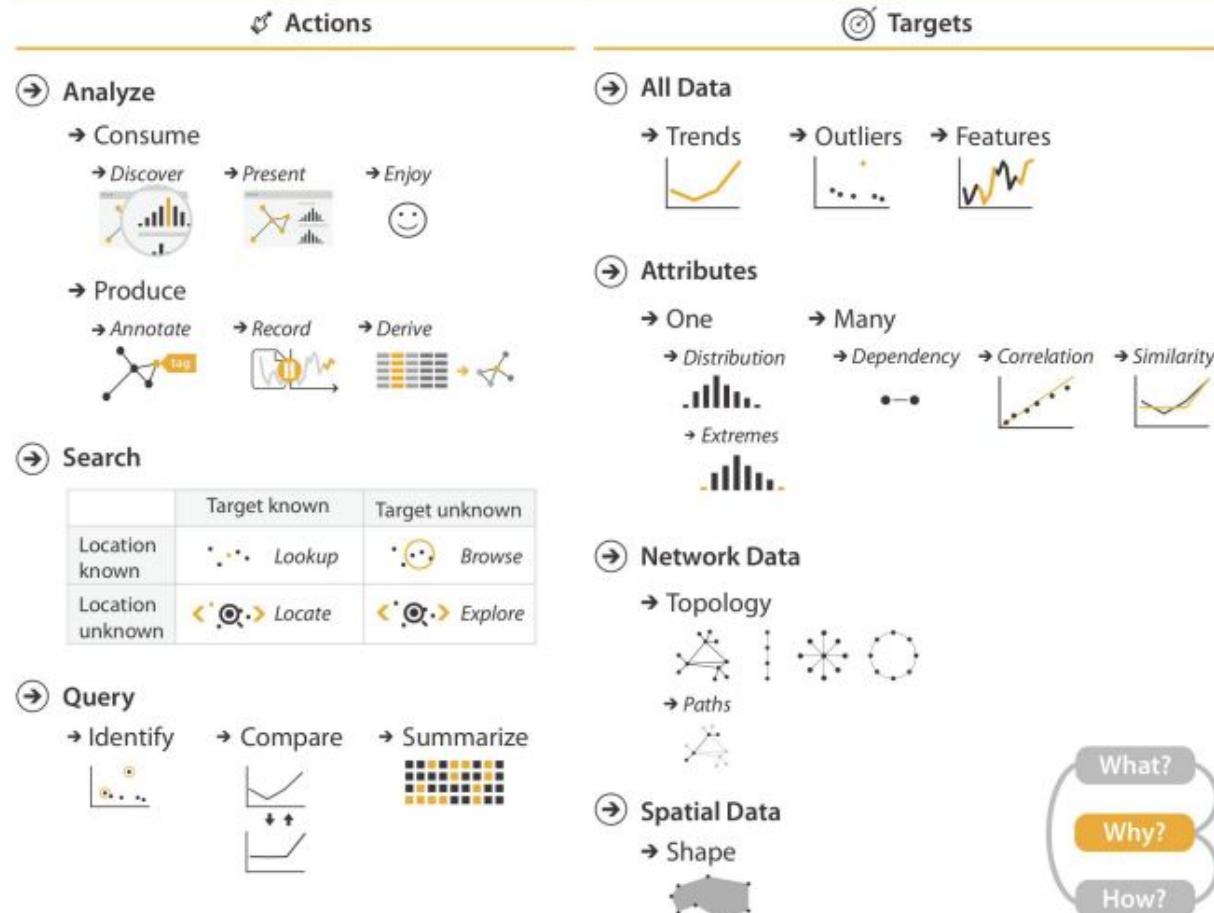
*Why*

Abstracción de tareas

---

# Why

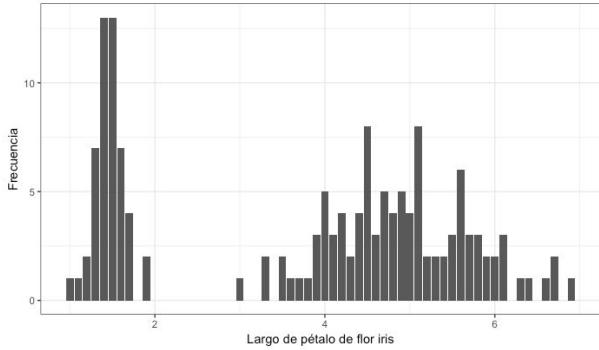
## Why?



# ¿Por qué necesitamos un “por qué”?

**Caso 1:** Si nos interesa identificar la distribución del tamaño de pétalos de una flor,

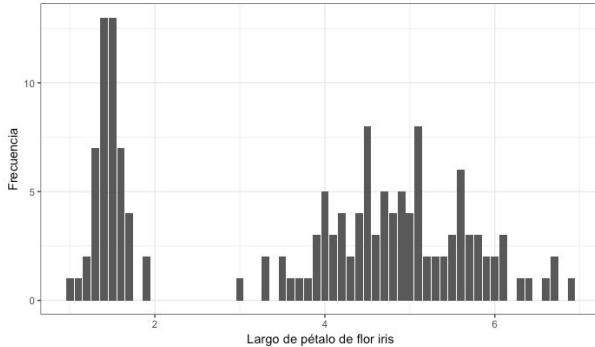
- Un histograma podría ser una buena opción.



# ¿Por qué necesitamos un “por qué”?

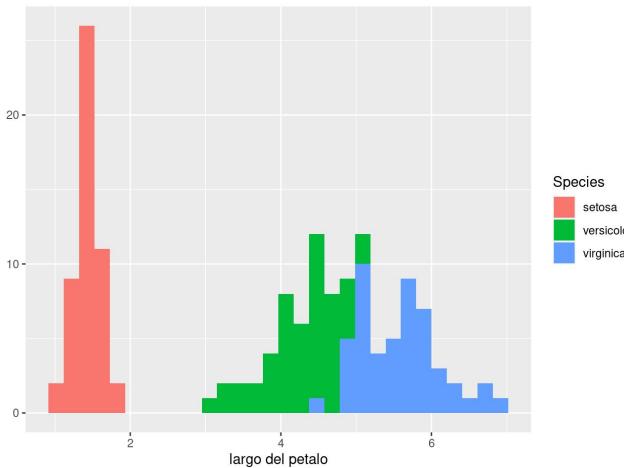
**Caso 1:** Si nos interesa identificar la distribución del tamaño de pétalos de una flor,

- Un histograma podría ser una buena opción.



**Caso 2:** Si queremos comparar la distribución de tamaños de pétalos de 3 tipos de flores,

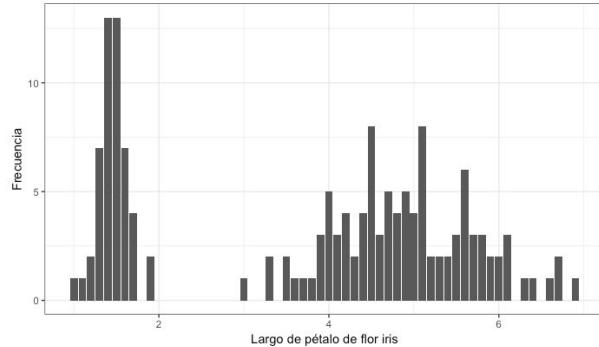
- Una visualización que agregue el canal de color y sobreponga 3 histogramas es una mejor opción.



# ¿Por qué necesitamos un “por qué”?

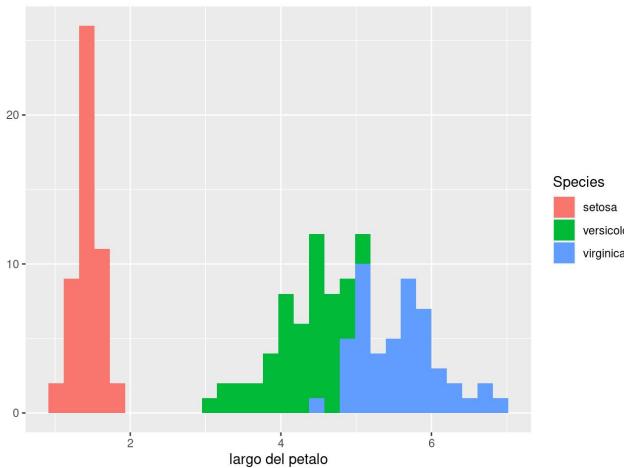
**Caso 1:** Si nos interesa identificar la distribución del tamaño de pétalos de una flor,

- Un histograma podría ser una buena opción.



**Caso 2:** Si queremos comparar la distribución de tamaños de pétalos de 3 tipos de flores,

- Una visualización que agregue el canal de color y sobreponga 3 histogramas es una mejor opción.



El por qué del gráfico, **modifica la decisión** del tipo de gráfico para los mismos datos

# ¿Por qué necesitamos un “por qué”?

**Caso 1:** Deseo contrastar la cantidad de pacientes por médico entre el año pasado y este año.

- Comparar valores entre 2 rangos de fechas

**Caso 2:** Revisar si las ventas por sucursal de este año fueron mejores que las del año pasado.

- Comparar valores entre 2 rangos de fechas.

**Ambos casos, aunque el contexto y los datos son distintos, presentan la misma tarea visual.**

# Abstracción de las tareas

- El *Framework* propone encontrar el:

**Par {acción, objetivo}**

- Por ejemplo:
  - **Identificar** la película con el **mayor costo (máximo)**
  - **Comparar tendencias** en el clima entre zonas cercanas al polo norte y al polo sur.
  - **Descubrir outliers** en las ventas del empresa
  - **Explorar la topología** en la red del metro.
  - **Resumir** las notas de los alumnos (**valores singulares**)

# Abstracción de las tareas

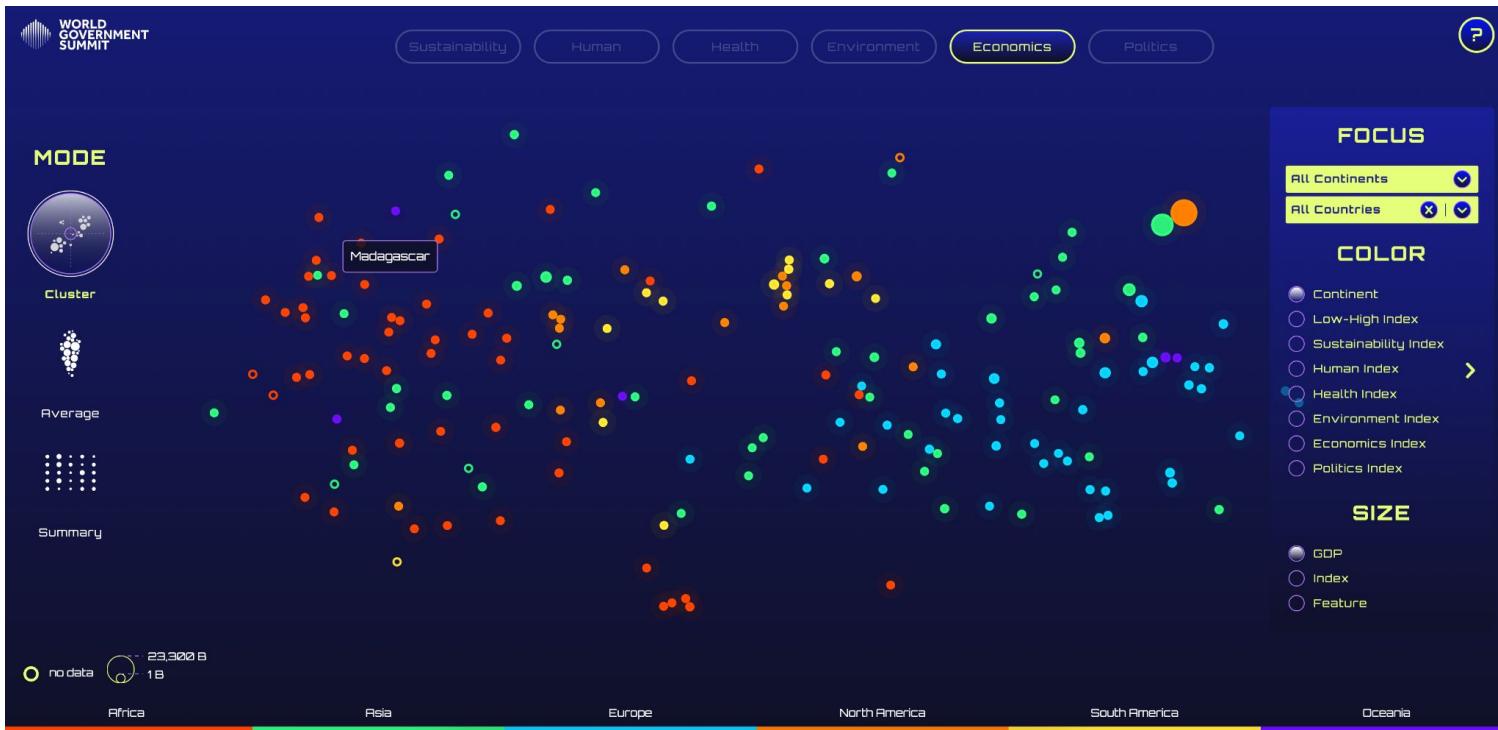
- Define tres niveles de acciones (del más alto al más bajo):
  - Analizar
  - Buscar
  - Consultar
- Estos tres niveles son independientes entre ellos. Incluso, es recomendable describir acciones en todos ellos.

# Abstracción de las tareas - Analizar

- Analizar:
  - Acción para **consumir** información.
  - Acción para **producir** información
- **Descubrir información**
  - Usar la visualización como herramienta para **encontrar conocimiento** no descubierto.
  - Vamos a "conocer" los datos.
- **Presentar información**
  - Usar la visualización como herramienta para **comunicar información**.
  - Vamos a "explicar" los datos.

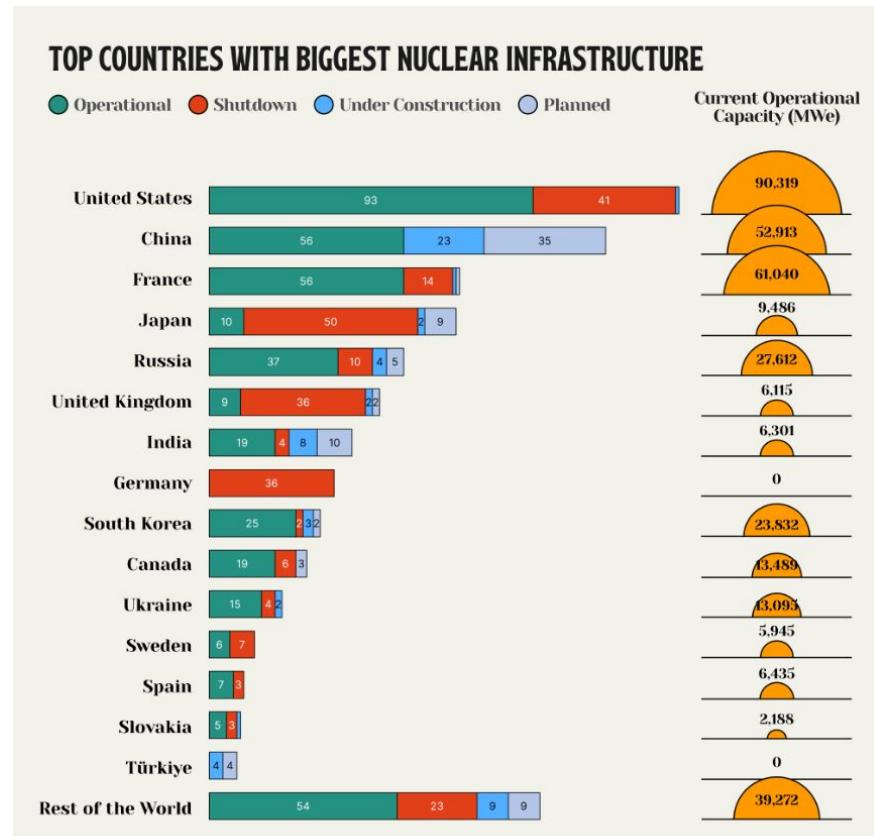
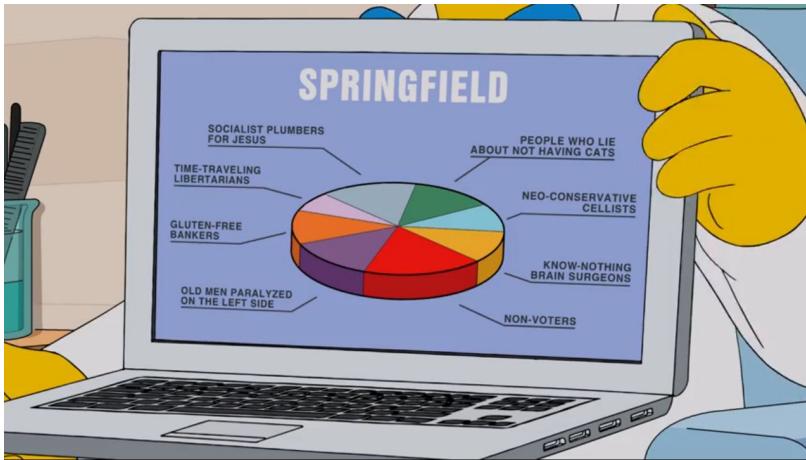
# Abstracción de las tareas - Analizar

- Analizar - Consumir - Descubrir



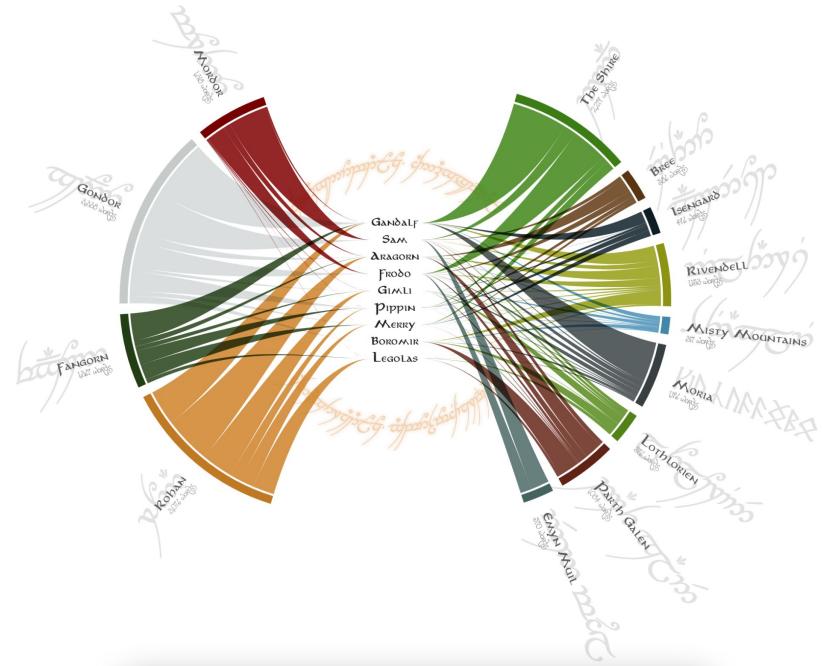
# Abstracción de las tareas - Analizar

- Analizar - Consumir - Presentar



# Abstracción de las tareas - Analizar

- Analizar:
  - Acción para **consumir** información.
  - Acción para **producir** información.
- **Disfrutar:** El usuario es atraído por curiosidad, y puede que no pertenezca a la audiencia principal

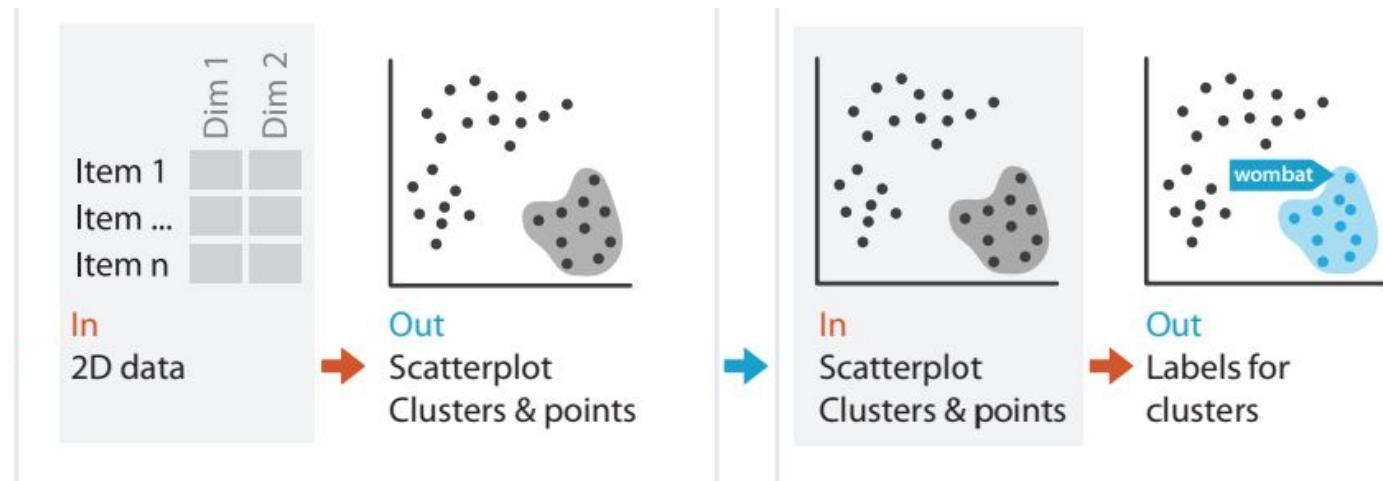


# Abstracción de las tareas - Analizar

- Analizar:
  - Acción para **consumir** información.
  - Acción para **producir** información.
- **Comentar/etiquetar:** Se agregan comentarios gráficos o textuales.
  - Por ejemplo, cuando aplicamos vemos un gráfico de puntos y le damos un "nombre" a un conjunto de puntos cercanos.

# Abstracción de las tareas - Analizar

- Analizar - Producir - **Comentar/Etiquetar** (cuando le damos un nombre a un *cluster de puntos*)

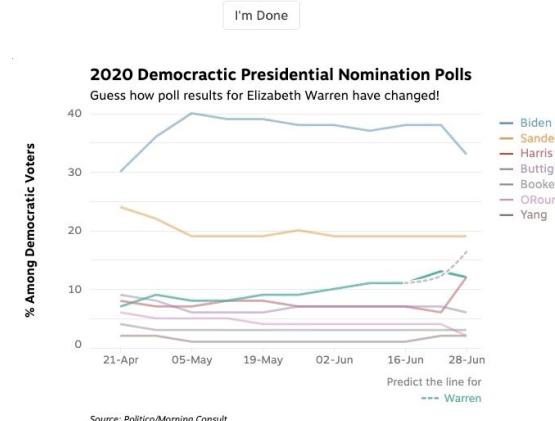
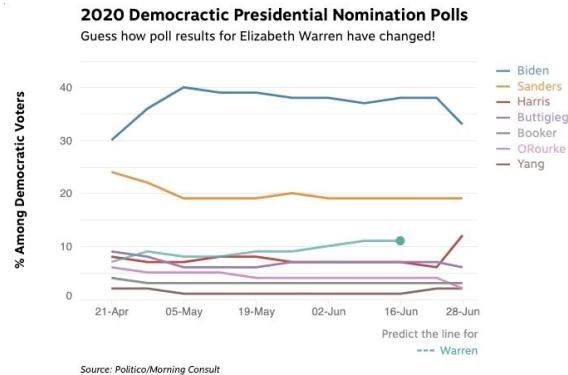


Identificamos clusters

Definir un nombre

# Abstracción de las tareas - Analizar

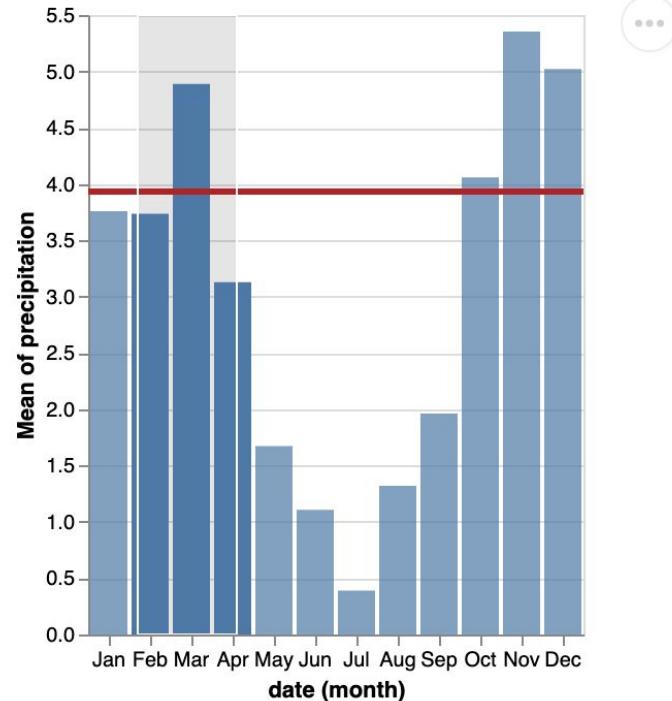
- Analizar:
  - Acción para **consumir** información.
  - Acción para **producir** información.
- Al producir información, uno puede:
  - **Grabar:** Capturar un artefacto persistente a partir de la visualización (por ejemplo, *screenshot*, grabación de pantalla, acciones del usuario)



Show Others' Predictions

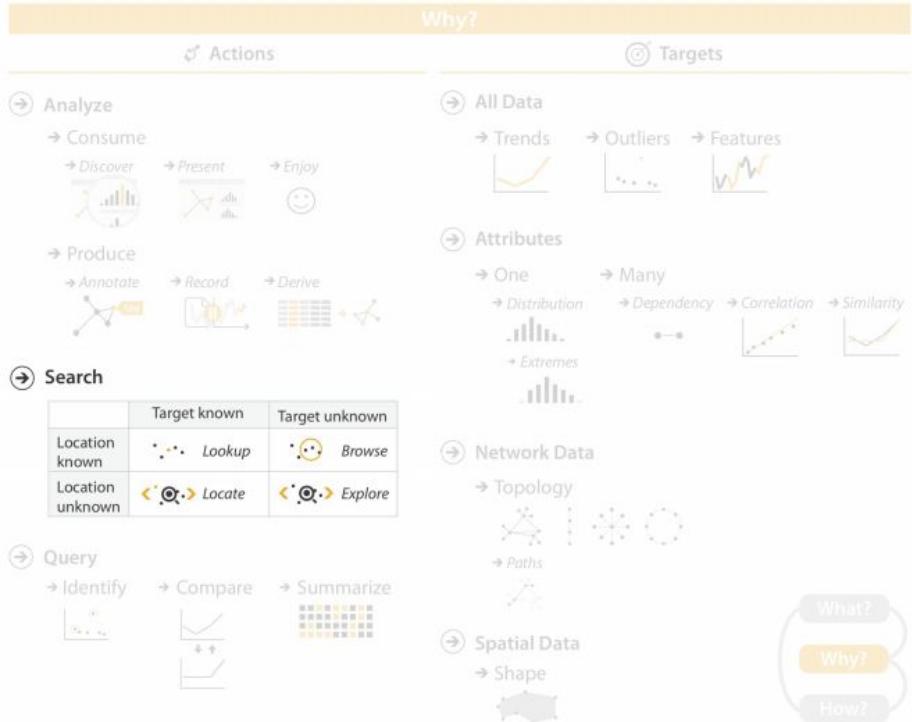
# Abstracción de las tareas - Analizar

- Analizar:
  - Acción para **consumir** información.
  - Acción para **producir** información.
- Al producir información, uno puede:
  - **Transformar**: Producir nueva información a partir de datos existentes.



# Abstracción de las tareas - Buscar

- Segundo nivel de tareas. El nivel de búsqueda que realizará el usuario en la visualización
- Una instancia para empezar a tener conciencia con mayor profundidad el tipo de usuario que esperamos que interactúe con la visualización.

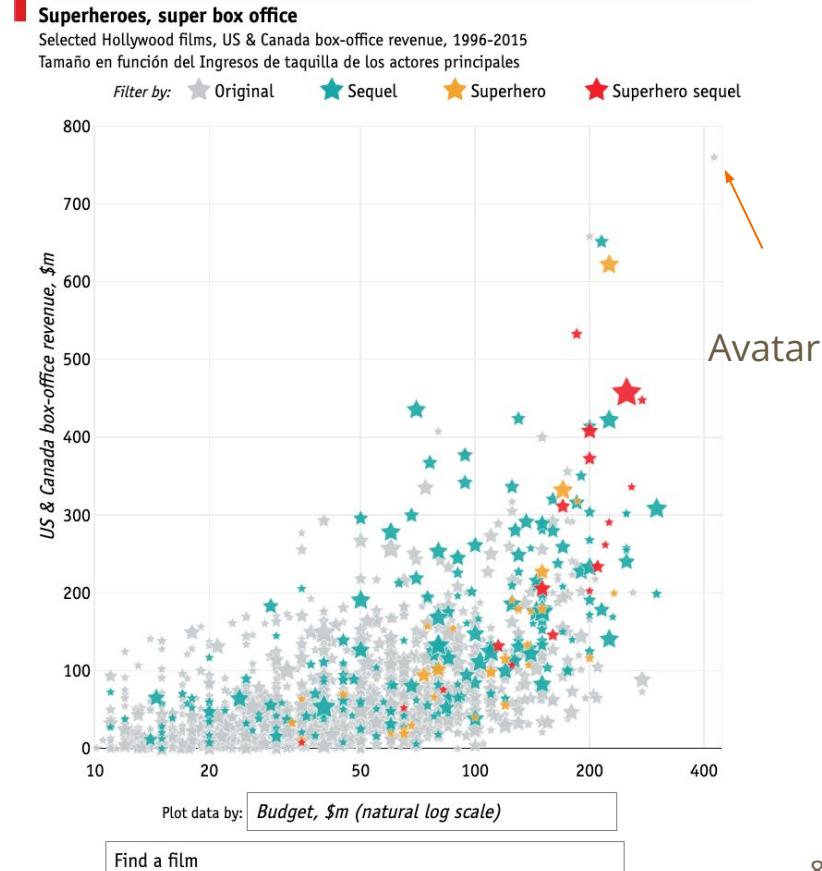


# Abstracción de las tareas - Buscar

- Buscar
  - ¿Sé exactamente el dato que busco?
  - ¿Tengo información que me ayude a saber por donde buscar el dato?

# Abstracción de las tareas - Buscar

- Buscar
  - ¿Sé exactamente el dato que busco?
  - ¿Tengo información que me ayude a saber por donde buscar el dato?
- Navegar
  - No sé con qué me encontraré pero tengo información al respecto.
  - ¿Cuál fue la película con más ingresos?



# Abstracción de las tareas - Buscar

- Buscar
  - ¿Sé exactamente el dato que busco?
  - ¿Tengo información que me ayude a saber por donde buscar el dato?
- Explorar
  - No sé lo que encontraré y tampoco sé por donde buscar.
  - ¿cómo distribuye la cantidad de mujeres en el gobierno entre todos los países?

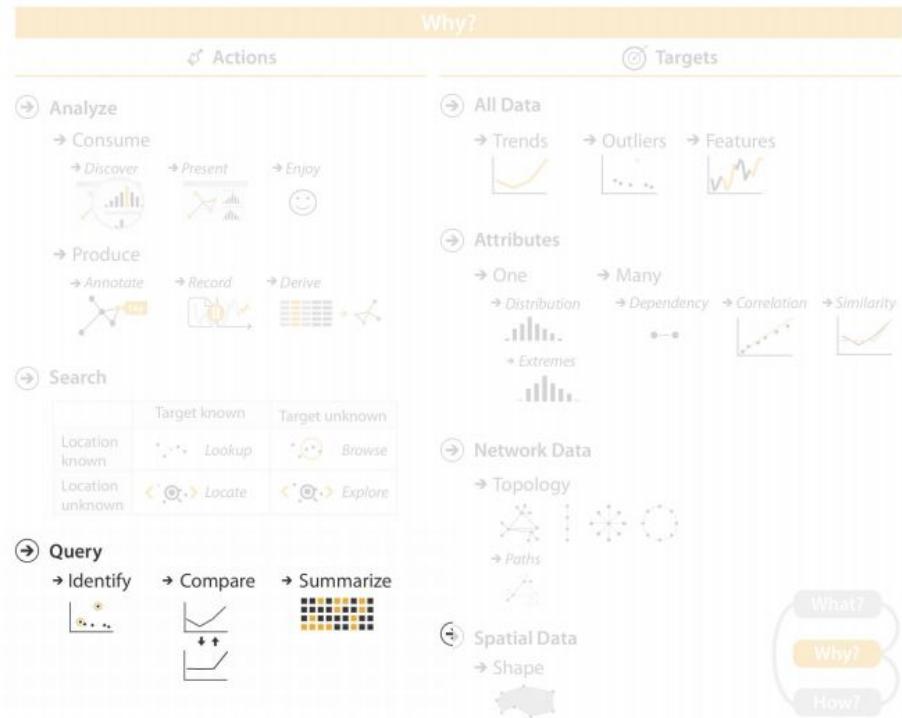


# Abstracción de las tareas - Buscar

	Sabemos exactamente <b>qué</b> queremos buscar	No sabemos lo que buscamos o solo sabemos <b>una propiedad</b>
	Objetivo conocido	Objetivo desconocido
Localidad conocida	 <b>Ubicar</b> <i>lookup</i>	 <b>Navegar</b> <i>Browse</i>
Localidad desconocida	 <b>Localizar</b> <i>locate</i>	 <b>Explorar</b> <i>Explore</i>

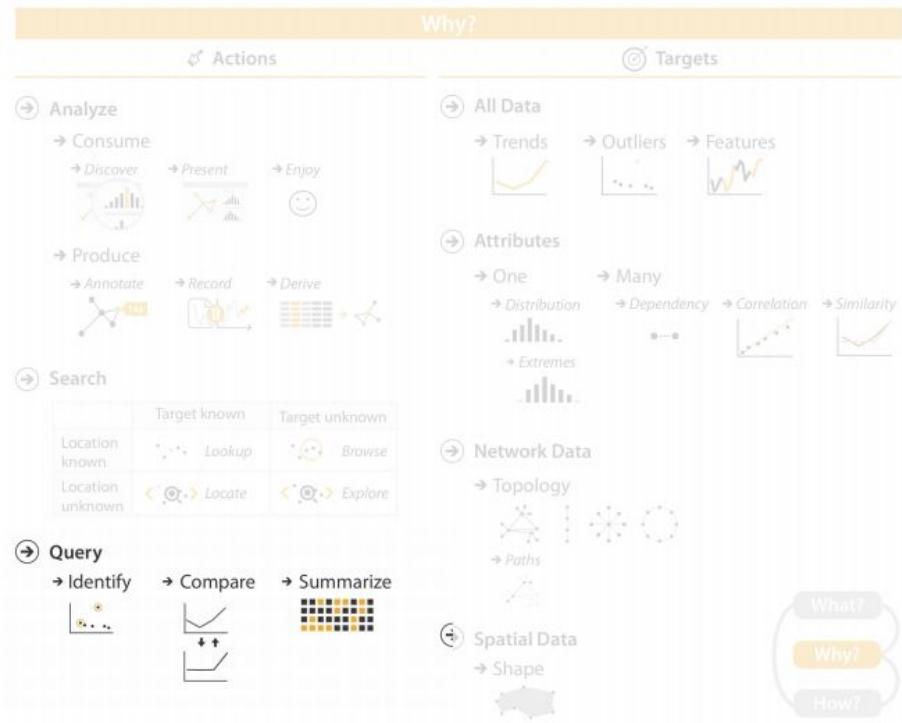
# Abstracción de las tareas - Consultar

- Tercer nivel de tareas. El nivel de consulta a realizar en la visualización.
  - Identificar.
  - Comparar.
  - Resumir.
- Ellos se diferencian en la cantidad de información considerada.



# Abstracción de las tareas - Consultar

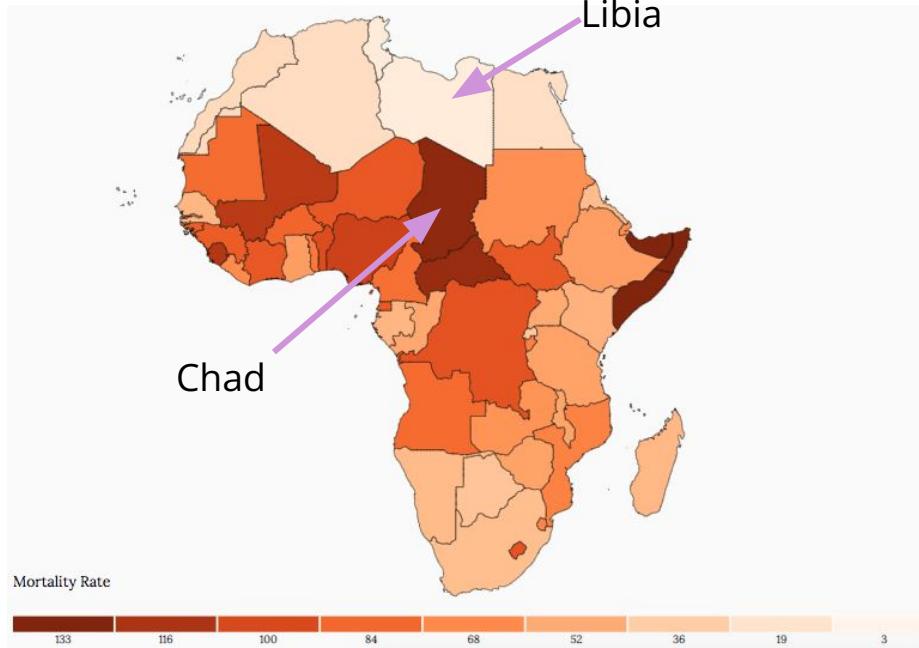
- Tercer nivel de tareas. El nivel de consulta a realizar en la visualización.
  - Identificar.
  - Comparar.
  - Resumir.
- Ellos se diferencian en la cantidad de información considerada.
  - **Identificar:** 1 objetivo.
  - **Comparar:** múltiples objetivos.
  - **Resumir:** conjunto completo de datos.



# Abstracción de las tareas - Consultar

Para exemplificar, tenemos el caso de mortalidad infantil en África.

- **Identificar:** la mortalidad específica de Chad.
- **Comparar:** la mortalidad de Chad y Libia.
- **Resumir:** la mortalidad general de África.

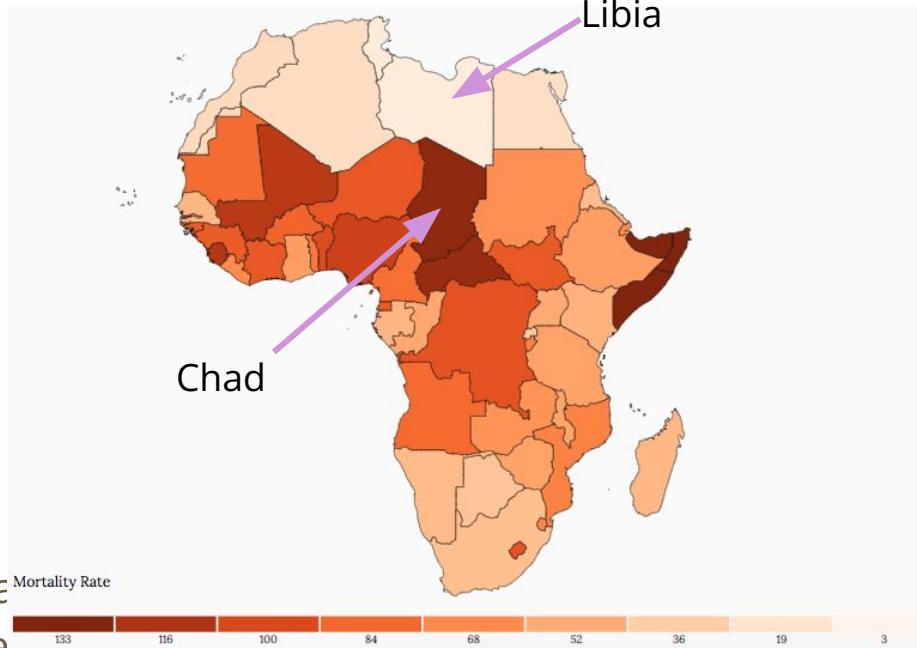


# Abstracción de las tareas - Consultar

Para ejemplificar, tenemos el caso de mortalidad infantil en África.

- **Identificar**: la mortalidad específica de Chad.
  - **Comparar** la mortalidad de Chad y Libia.
  - **Resumir**: la mortalidad general de África.

¿Cuál tarea visual es óptima para este gráfico? Asumamos que apenas conocemos los países de África.

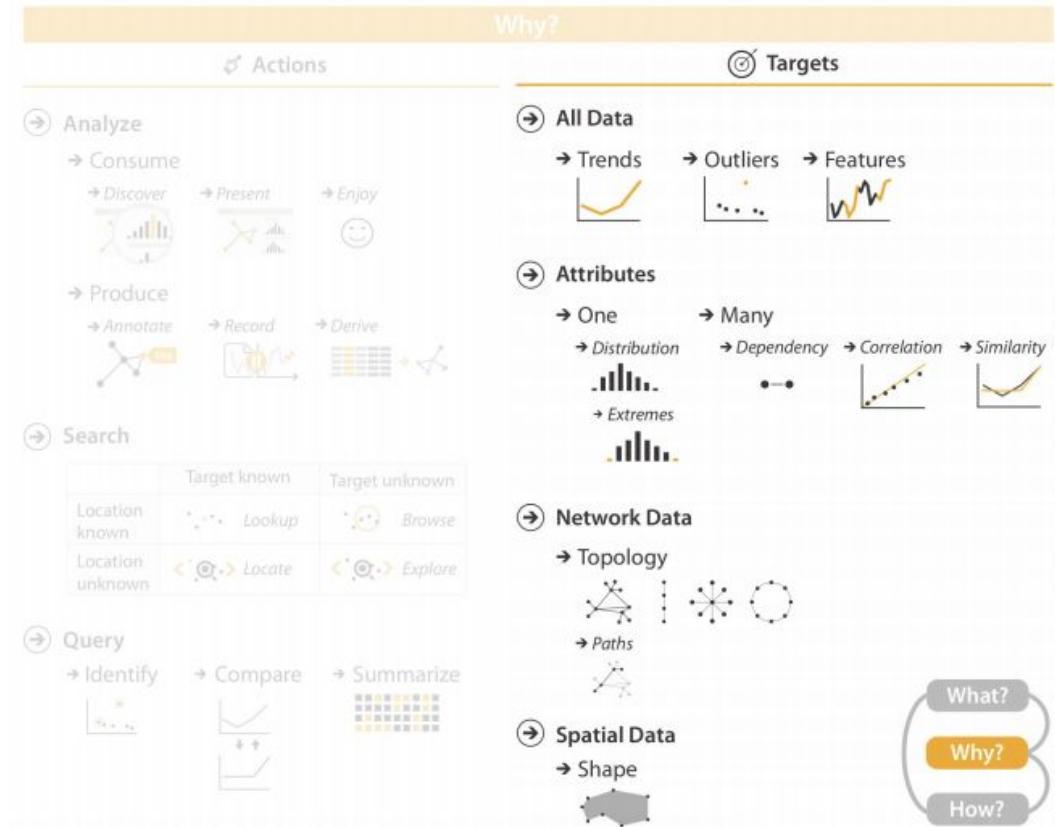


# Objetivo

—

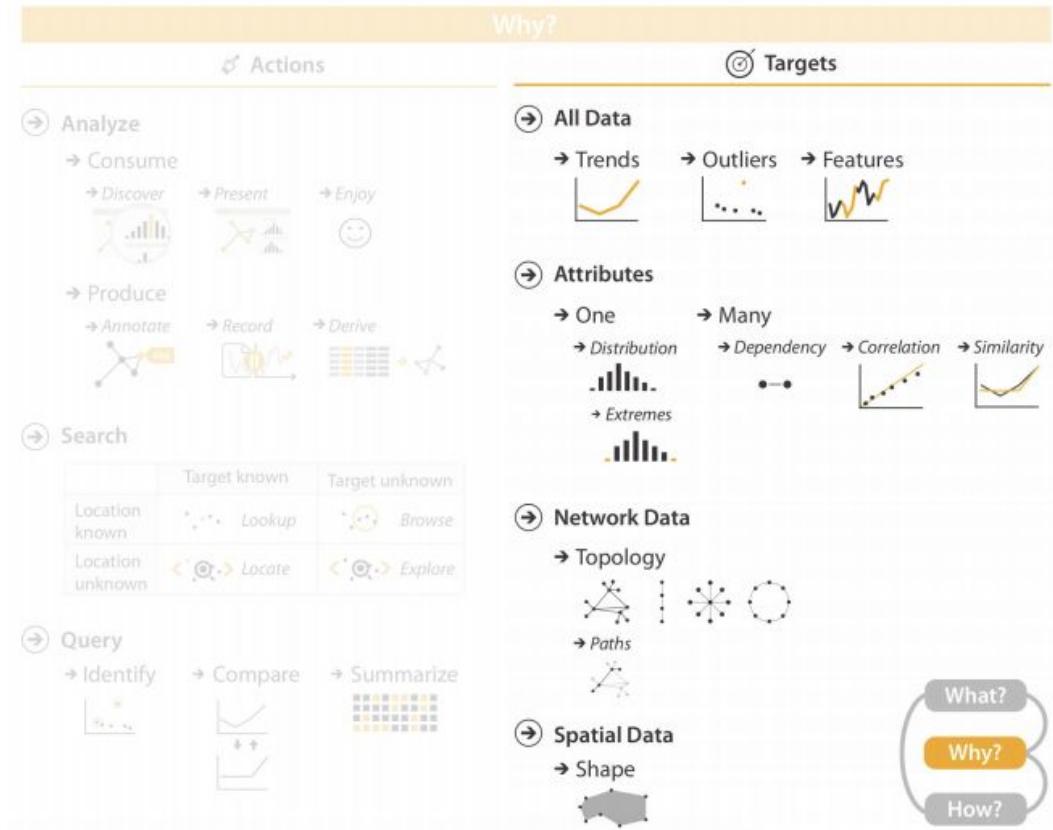
# Abstracción de las tareas - Objetivos

- Los *targets* se refieren a algún **aspecto del dato** que queremos visualizar. Como puede ser la tendencia de los datos, la distribución de un atributo, etc.
- Corresponde al sustantivo de la tarea.



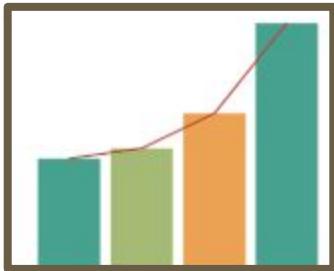
# Abstracción de las tareas - Objetivos

- Los *targets* se refieren a algún **aspecto del dato** que queremos visualizar. Como puede ser la tendencia de los datos, la distribución de un atributo, etc.
- Corresponde al sustantivo de la tarea.
- **Importante:** el objetivo **NO es la meta de la visualización**, sino que es una propiedad del dato que esperamos observar gracias al gráfico.



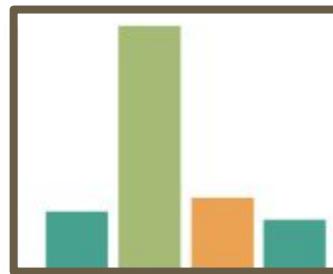
# Abstracción de las tareas - Objetivos

A nivel de todos los datos



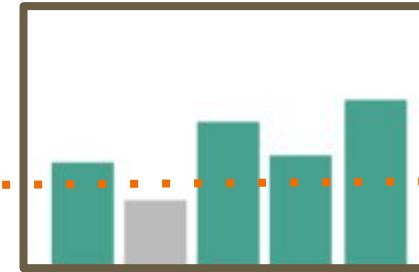
Tendencias

Descripción general  
de un patrón en los  
datos



Outliers

Datos atípicos, que  
no siguen un patrón



Features  
(Extractos)

Conjunto de datos con  
ciertas características  
en común

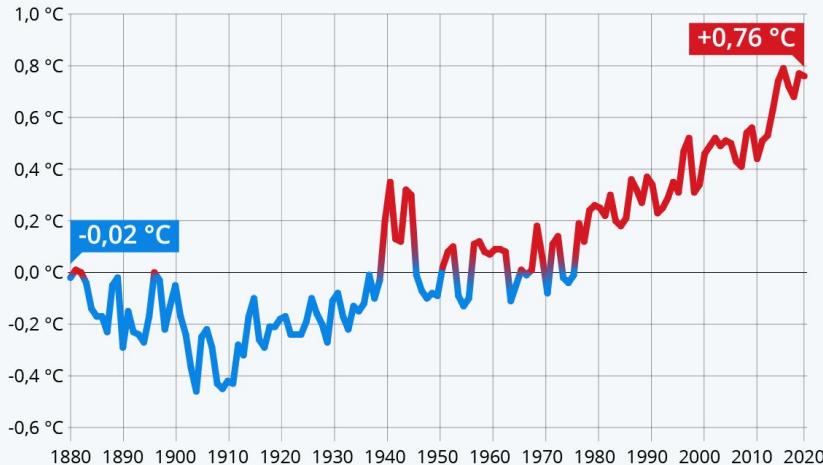
# Abstracción de las tareas - Objetivos

## A nivel de todos los datos - Ejemplo

- Identificamos la **tendencia** de la temperatura en el tiempo (va en aumento).
- Presentamos **features** en relación a los años en donde la temperatura aumentó o disminuyó.

### Los océanos, cada vez más cálidos

Anomalías de la temperatura superficial global de los océanos con respecto a la media del siglo XX



Fuente: Centros Nacionales de Información Ambiental (NCEI) de la NOAA



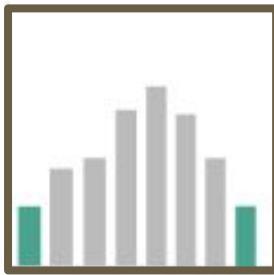
# Abstracción de las tareas - Objetivos

## A nivel de atributos



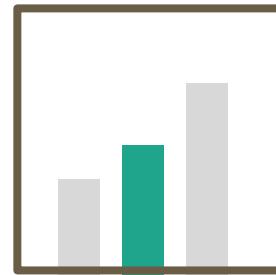
**Distribución**

Frecuencia de los valores presentes en un atributo



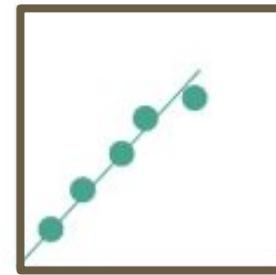
**Extremos**

Rango numérico en el que viven los atributos.



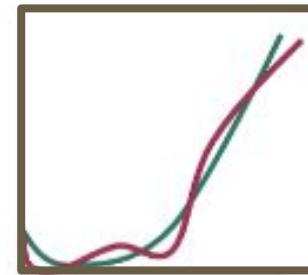
**Valor singular**

El valor específico de algún dato.



**Correlación**

Si la variación de un atributo depende de otro atributo.



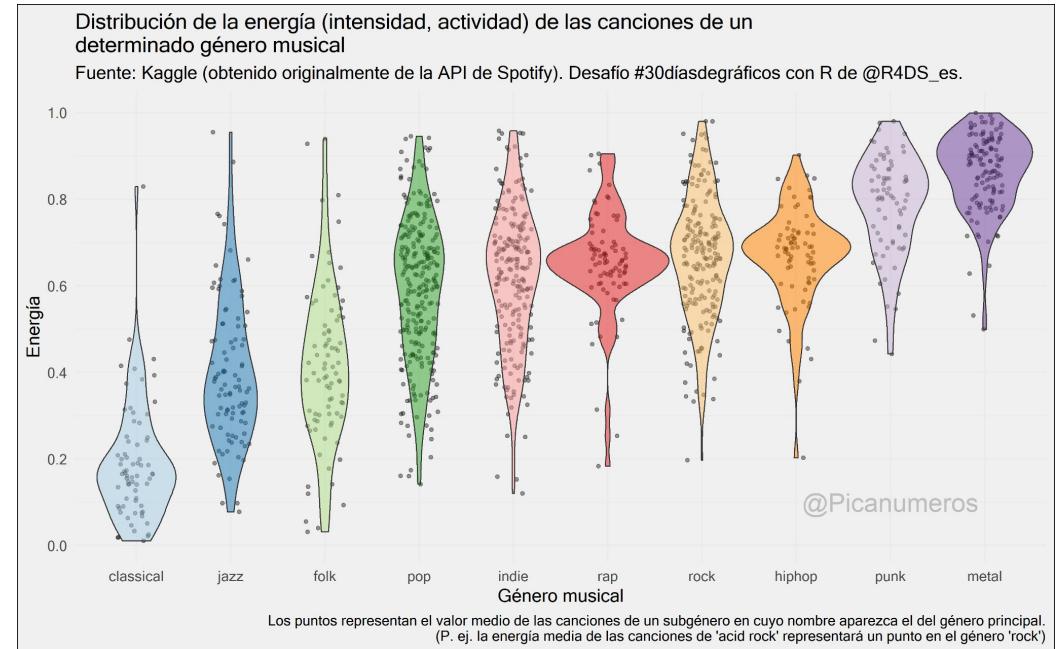
**Similitud**

Qué tan semejantes son los valores de 2 atributos.

# Abstracción de las tareas - Objetivos

## A nivel de atributos - Ejemplo

- Comparamos la **distribución** de la energía según género musical.
- Identificamos los **extremos** de energía por género musical.
- Descubrimos valores **outliers** en cada género.



# Objetivos (*targets*)

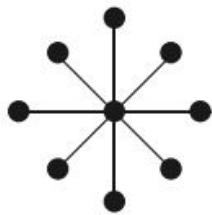
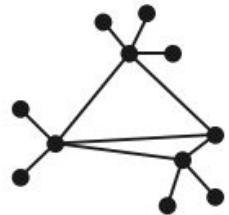
## A nivel de atributos - Ejemplo

- Comparar la **distribución** del valor de las acciones de Nintendo.
- Presentar **features** (acción subió o bajó).
- Identificar **valores singulares y extremos** de las acciones.
- Descubrir **tendencias** en las acciones.



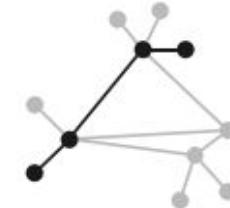
# Abstracción de las tareas - Objetivos

## Otros tipos de datos - Redes y geométricos



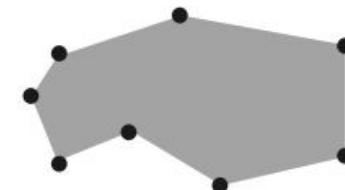
Topología

Estructura general de cómo se relacionan las diferentes entidades



Caminos

Secuencia de enlaces entre nodos conectados



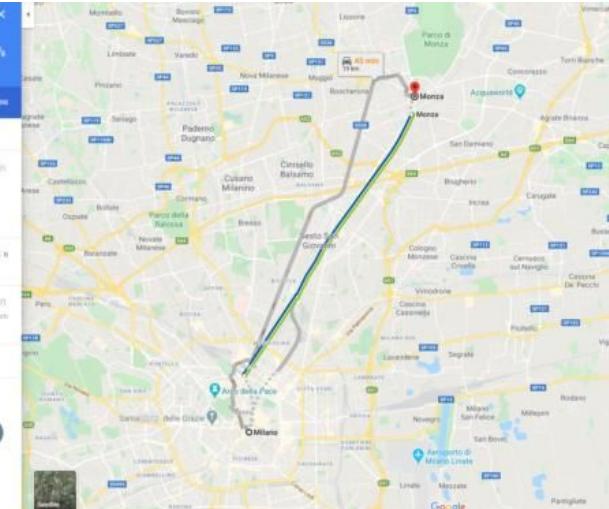
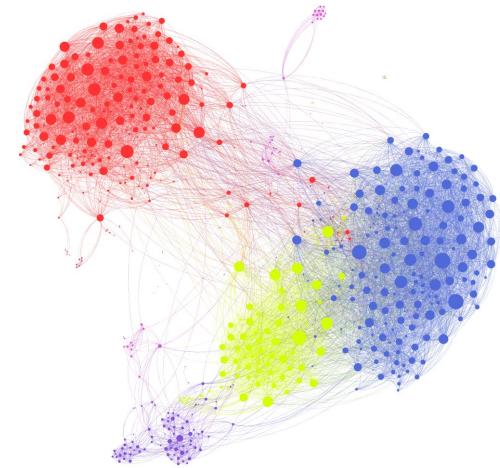
Forma

La forma de entidades o fenómenos geométricos.

# Abstracción de las tareas - Objetivos

## Otros tipos de datos - Ejemplo

- Podemos observar la topología o estructura general del grafo.
- Se distinguen 4 comunidades o clusters de nodos, donde hay uno (el rojo) que se separa más del resto.
- En el mapa, podemos observar diferentes caminos para llegar a un destino y se destaca un camino en particular.



# Resumen

Lo importante es quedarse con qué queremos que **haga el usuario (acción)** y a qué **propiedad del dato** nos interesa aplicar la acción (**objetivo**)

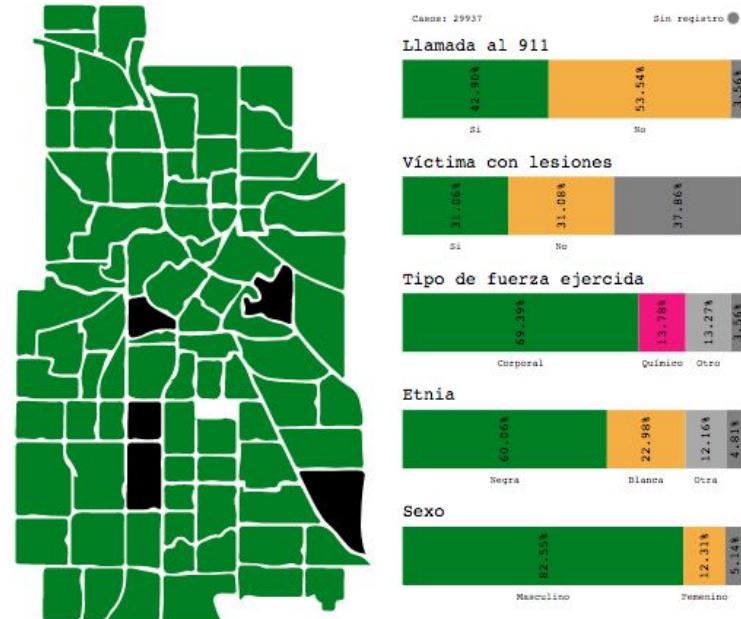
- Comparar - valores
- Identificar - mínimo
- Presentar - correlaciones
- Descubrir - datos atípicos
- Explorar - tendencias
- Localizar - máximo

# Análisis de casos

Visualización sobre la Violencia Policial en Minneapolis (ciudad de Minnesota, USA).

- **Descubrir distribuciones.** En particular, cómo distribuye los casos de violencia según atributos como el género o la raza en diferentes barrios de la ciudad.
- **Explorar features.** En particular, los **diferentes barrios** de Minneapolis y conocer sus características particulares dentro del dataset.

## Minneapolis: Violencia Policial y Racismo

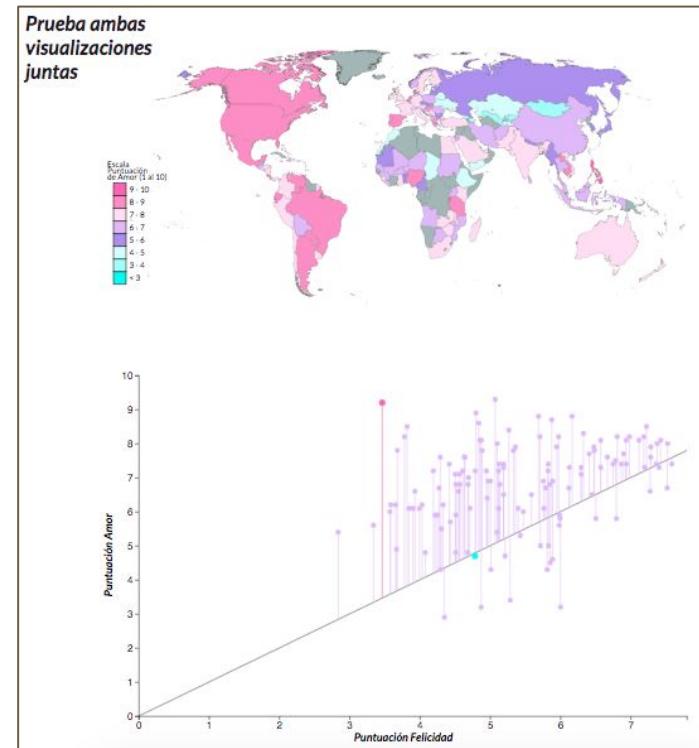


# Análisis de casos - Para la casa

Visualización sobre el amor y la felicidad en el mundo. Se recomienda hacer *zoom out* a la pantalla para ver bien la visualización

## ¿Qué tareas permite enfrentar su visualización?

- ¿Qué se intenta descubrir?
- Un usuario puede poner el cursor (*mouse*) sobre un punto del gráfico de abajo y el mapa enfatiza dicho país en el mundo. ¿Qué tipo de búsqueda será?
- El gráfico de punto destaca el punto más alejado y más cercano a la diagonal. ¿Qué tarea permitirá eso?



# Análisis de casos - Para la casa

Algunas tareas posibles son:

- **Descubrir** cómo distribuye la puntuación de amor a lo largo del mundo.
- **Descubrir** si existe una correlación entre las puntuaciones de amor (*Percent Feeling Love*) y felicidad (*Happiness Score*).
- **Navegar** para encontrar **un país** en base a su valor de puntuaciones de amor (*Percent Feeling Love*) y felicidad (*Happiness Score*). [*No sé qué país encontraré, pero si donde buscarlo por su puntaje en el gráfico de puntos*]
- **Identificar extremos** en la diferencia de las puntuaciones de amor y felicidad.
- **Resumir la puntuación de amor** en todo el mundo.

# Taller 1

Resolución de dudas y trabajar en el taller.

---



---

# Visualización de Información y Analítica Visual

— Hernán Valdivieso López ([hfvaldivieso@uc.cl](mailto:hfvaldivieso@uc.cl)) —

---