

# **Visualización de Información Buenas Prácticas y Evaluación**

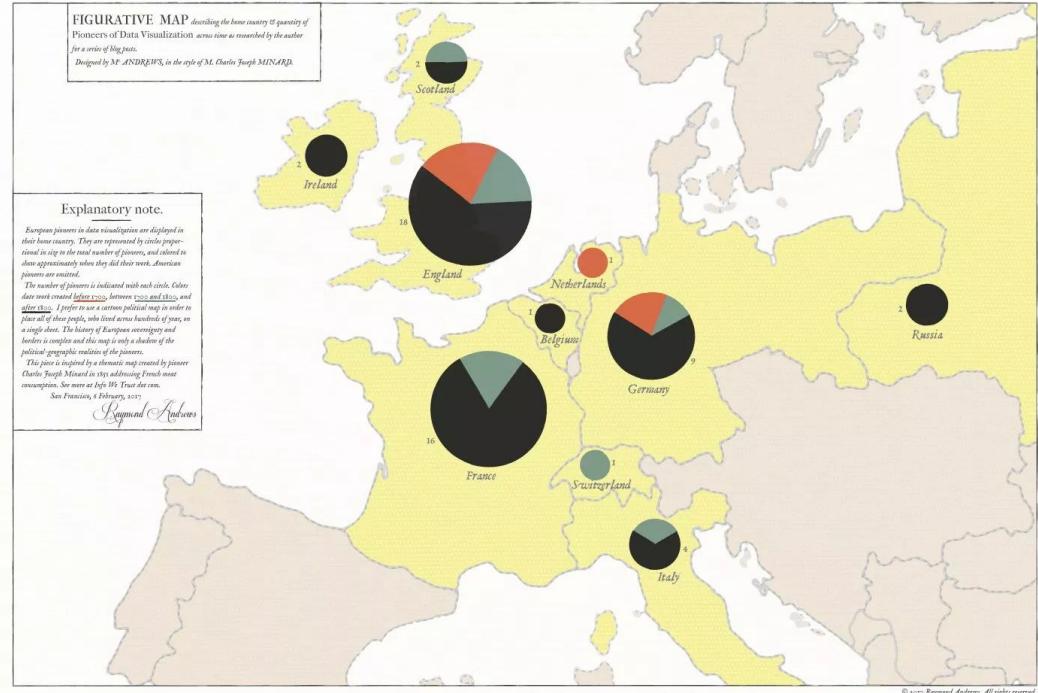
Daniela Opitz

[dopitz@udd.cl](mailto:dopitz@udd.cl)

Data Science Institute, Universidad del Desarrollo  
Edición 2023

# Buenas Prácticas y Evaluación

- Estas pueden pensarse como guías o principios que podemos seguir y considerar para crear visualizaciones de información más efectivas.
- Sintetizaremos conocimiento común del área, que se basa en estudios empíricos y relacionados con la percepción de estas herramientas.
- También es importante tener en cuenta que al ser guías, y no reglas o leyes, no siempre aplican para toda situación, **por lo cual no siempre es necesario realizarlas al pie de la letra.**



Vimos que los pie charts usualmente son una mala idea, pero en esta visualización de Minard funcionan bien - permiten ver la distribución de una variable en cada país, incluyendo el contexto geográfico.

# 3 Tipos de Buenas Prácticas

- **Conceptuales** - dependen de la teoría y la interpretación de ésta.
- **Estructurales** - cómo organizar y conectar la funcionalidad de la visualización.
- **Narrativas** - mostrar los datos es contar una historia. Siempre.

# Prácticas Conceptuales

# Prácticas Conceptuales

1. Función antes que forma
2. No todo necesita ser coloreado
3. No al 3D injustificado
4. No al 2D injustificado

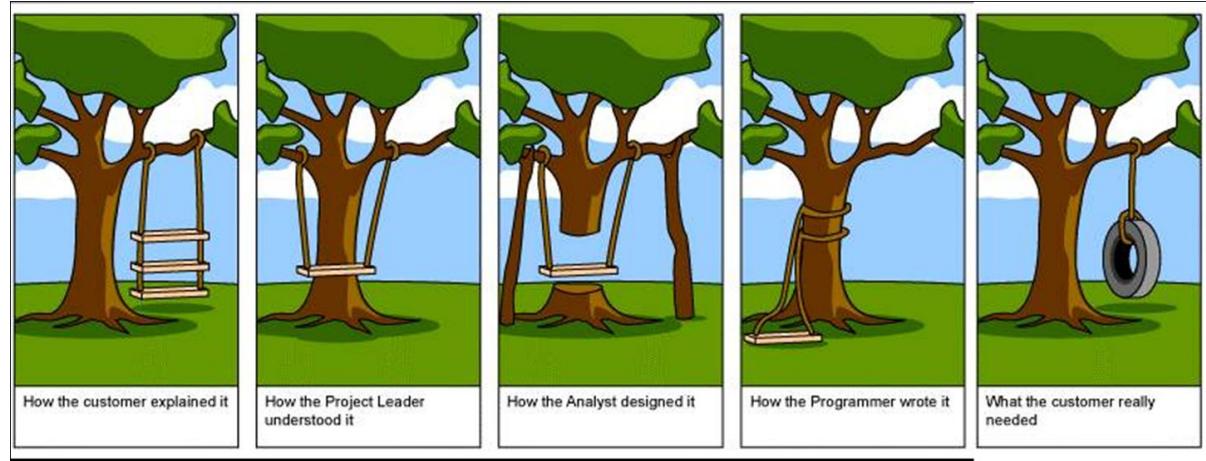
# Función antes que forma

Comienza enfocándote en la **funcionalidad**, pensando en las tareas a realizar.

*Forma:* “Necesitamos un dashboard”

*Función:* “Necesitamos tener datos actualizados de las tendencias de ventas de nuestros canales”

Es más fácil mejorar la estética de una visualización que la manera de trabajar con ella.



¿Cuál es la funcionalidad que resuelve la tarea de la imagen?

Distintas visiones sobre lo que pide un cliente por diferentes roles en un proyecto. Lo que necesitaba el cliente es distinto incluso de la manera en que solicitó el proyecto. Fuente: Cartoons Mix.

# No todo necesita ser coloreado

## Fertility Rate

Average number of children per woman over her lifetime  
Showing all countries for which complete data is available

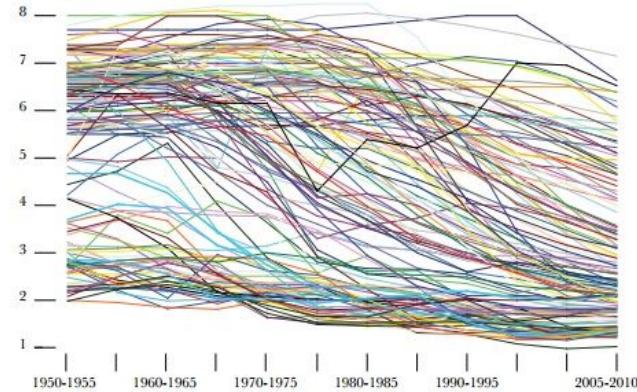


Figure 1.5 Too many lines obscure the message.

Fuente: The functional art, Alberto Cairo.

Por mostrarlo todo terminamos viendo nada.

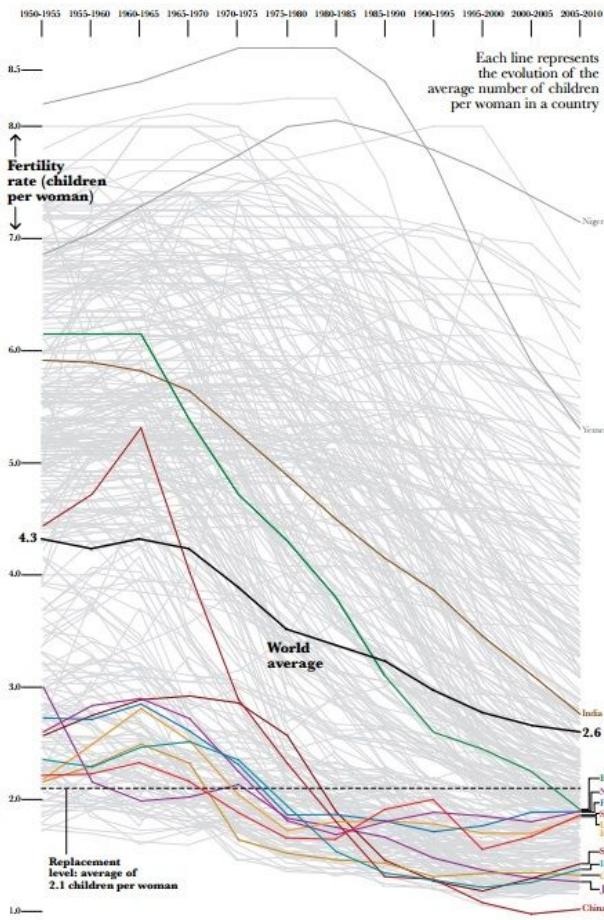
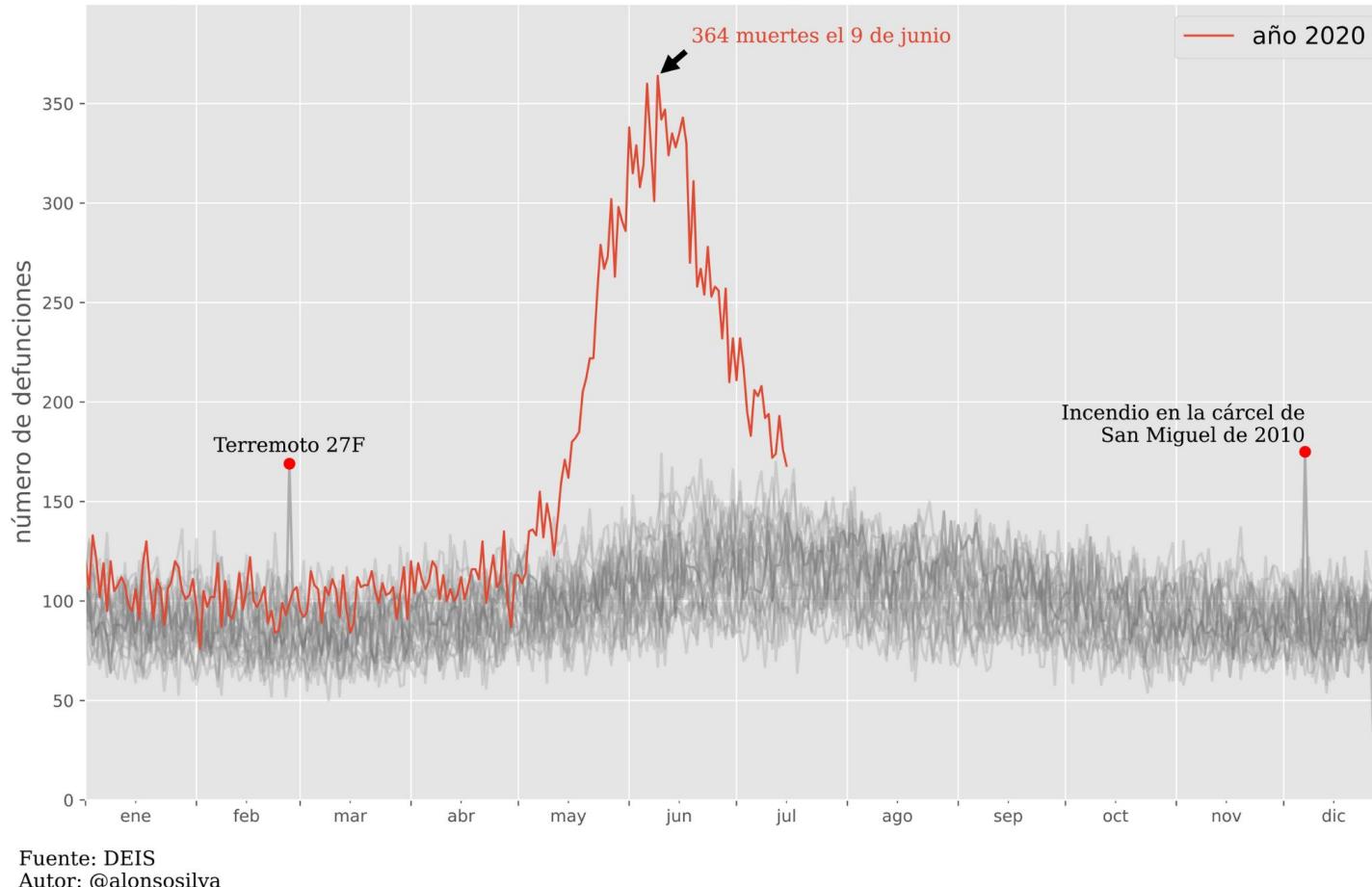


Figure 1.6 Highlighting the relevant, keeping the secondary in the background.

# Defunciones diarias en R.M. de 2000-2020

El gráfico muestra el número de defunciones diarias en R.M. En gris, todos los años hasta 2019; en rojo, el año 2020.



Fuente: DEIS

Autor: @alonsoSilva

El color también se puede utilizar para desviar la atención, quizás no engañosa directamente, pero sí de manera dudosa.



Mac

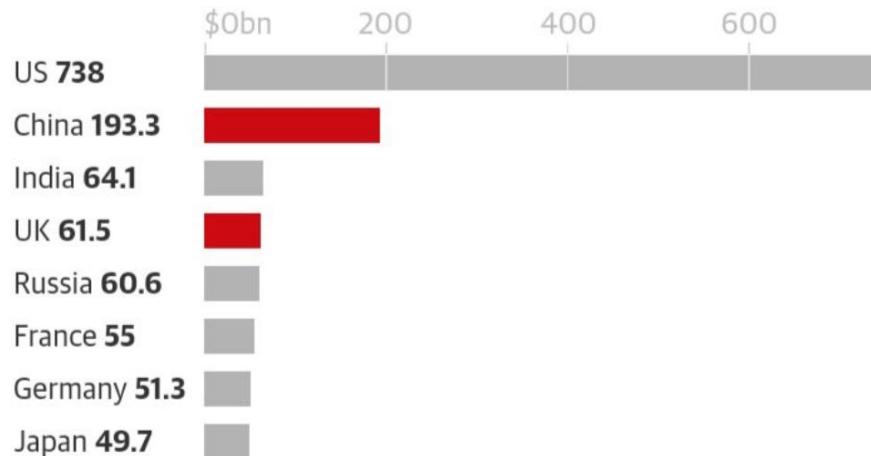
@GoodPoliticGuy

...

yea that's the main thing I noticed too

[Traducir Tweet](#)

### **China's defence budget in 2020 was three times that of the UK**



Guardian graphic | Source: IISS, the Military Balance 2021

# No al 3D injustificado

De acuerdo a los estudios de percepción, la profundidad no es un buen canal para codificar información.

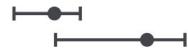
La información en un plano tiene una mejor codificación: se corresponde 1 a 1 en la variación de la información, y contiene los canales más eficientes a la hora de trabajar con datos de magnitud.

## ④ Magnitude Channels: Ordered Attributes

Position on common scale



Position on unaligned scale



Length (1D size)



Tilt/angle



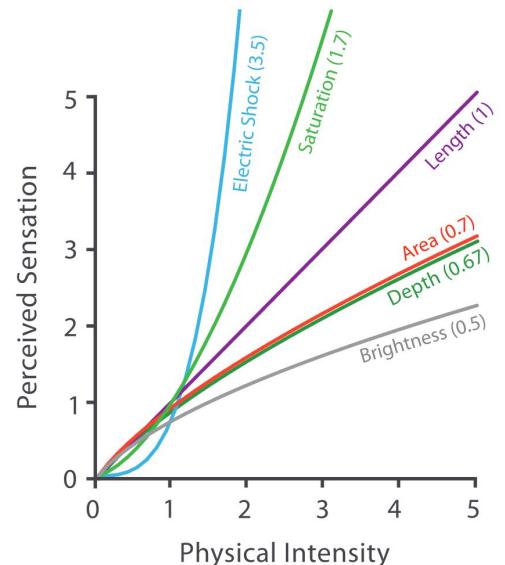
Area (2D size)



Depth (3D position)



Steven's Psychophysical Power Law:  $S = I^N$



# No al 3D injustificado (Oclusión)

La oclusión oculta información: los objetos más cercanos a la cámara pueden tapar a los que están más lejos.

Este problema se puede disminuir utilizando técnicas complejas de interacción, sin embargo, la oclusión persiste - es inherente a esta codificación visual.

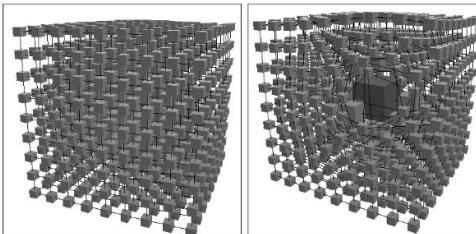


Figure 18: The radial Gaussian distortion around the central focus in 3D on the left, and with the visual access distortion on the right

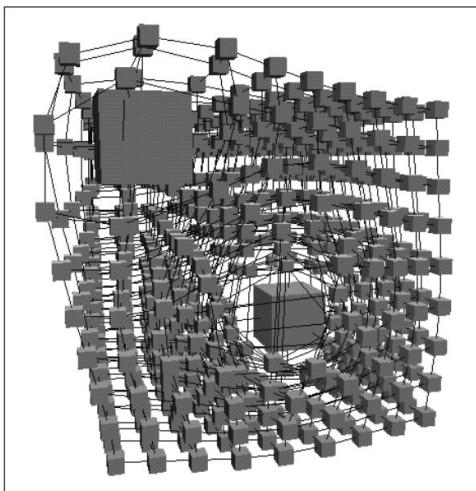


Figure 19: The visual access function may be applied simultaneously to more than one focus

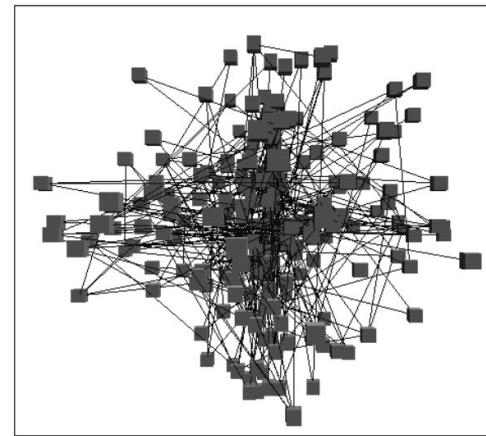


Figure 20: Undistorted random polar graph layout

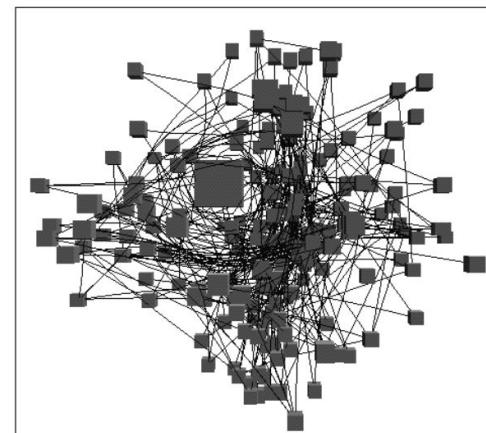


Figure 21: Random graph revealing focal node (from back of graph) with visual access distortion

Carpendale, M. S. T., Carpendale, T., Cowperthwaite, D. J., & Fracchia, F. D. (1996, October). Distortion viewing techniques for 3-dimensional data. In *Proceedings of the 1996 IEEE Symposium on Information Visualization (INFOVIS'96)* (p. 46). IEEE Computer Society.

# No al 3D injustificado (Perspectiva)

La perspectiva interfiere con los canales que utilizan el tamaño para codificar un atributo.

¡Se pierde el poder de utilizar un plano para visualizar!

Ejemplo: documentos que están más lejos pueden ser igual de importantes en una interfaz de búsqueda, sin embargo, al estar lejos podemos ignorarlos o bien asumir que no son relevantes.

Mukherjea, S., Hirata, K., & Hara, Y. (1996, October). Visualizing the results of multimedia web search engines. In *Proceedings IEEE Symposium on Information Visualization'96* (pp. 64-65). IEEE.

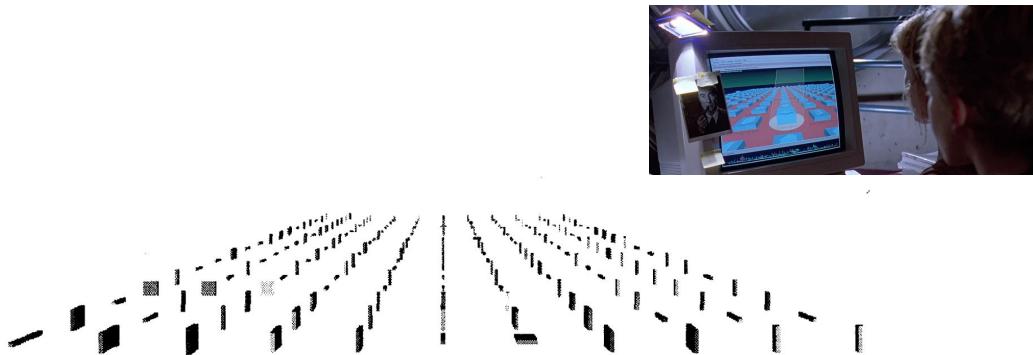


Figure 1: A bird's eye view of a tabular visualization of the results of a text search. Cubes representing retrieved documents are placed from left to right and front to back based on their total number of keyword matches. The width, height and depth of the cubes are matched to the frequency of individual keywords.

# Texto no Orientado

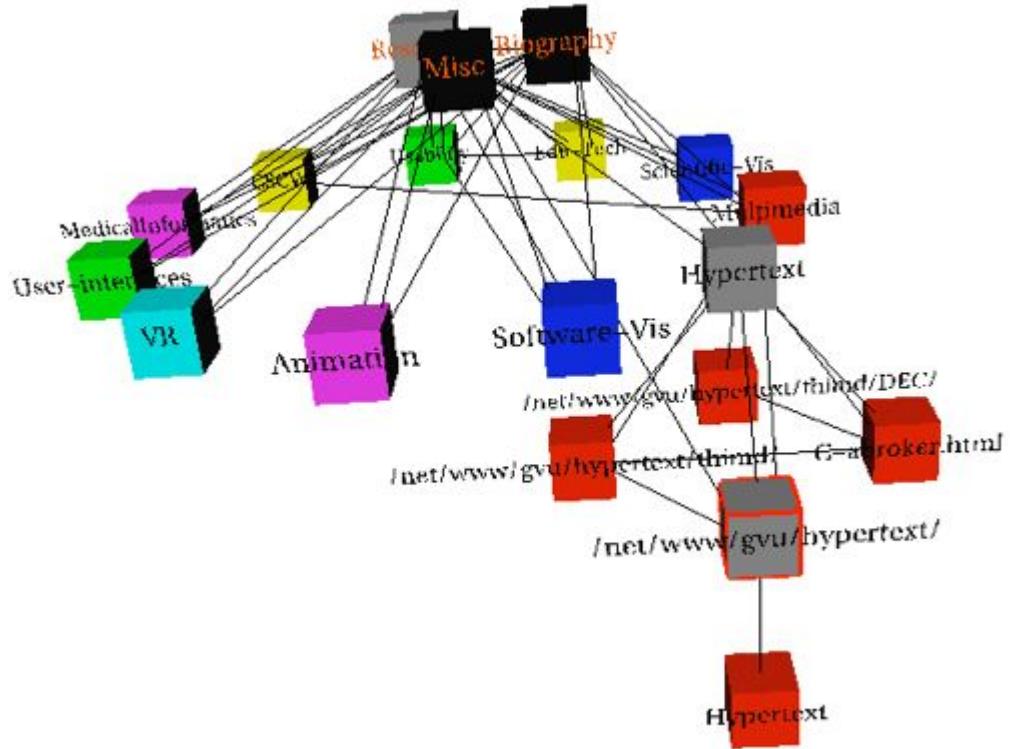
La legibilidad de un texto en 3D cambia.

No solamente hay cambios en el tamaño debido a la profundidad, sino también debido a la rotación.

También hay problemas de rendering debido a la resolución: ¡En 3D el texto se ve peor que en un plano 2D!

Recomiendo leer:

Grossman, T., Wigdor, D., & Balakrishnan, R. (2007, April). Exploring and reducing the effects of orientation on text readability in volumetric displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 483-492). ACM.



Mukherjea, S., & Foley, J. D. (1995). Visualizing the world-wide web with the navigational view builder. *Computer networks and ISDN Systems*, 27(6), 1075-1087.

# ¿Cuándo se justifica usar 3D?

Cuando la tarea **es** estudiar la forma en tres dimensiones, los beneficios de usar 3D superan los costos en la eficiencia de la visualización.

En esta visualización de volumen la visualización directa (a) no es útil. Por mucho que interactuemos con ella, no lograremos entender la forma de los datos. En (b) se utiliza un cálculo de líneas representativas de la forma, que junto a la interacción permiten entender qué es lo que se visualiza.

Li, L., & Shen, H. W. (2007). Image-based streamline generation and rendering. *IEEE Transactions on Visualization and Computer Graphics*, 13(3), 630-640.

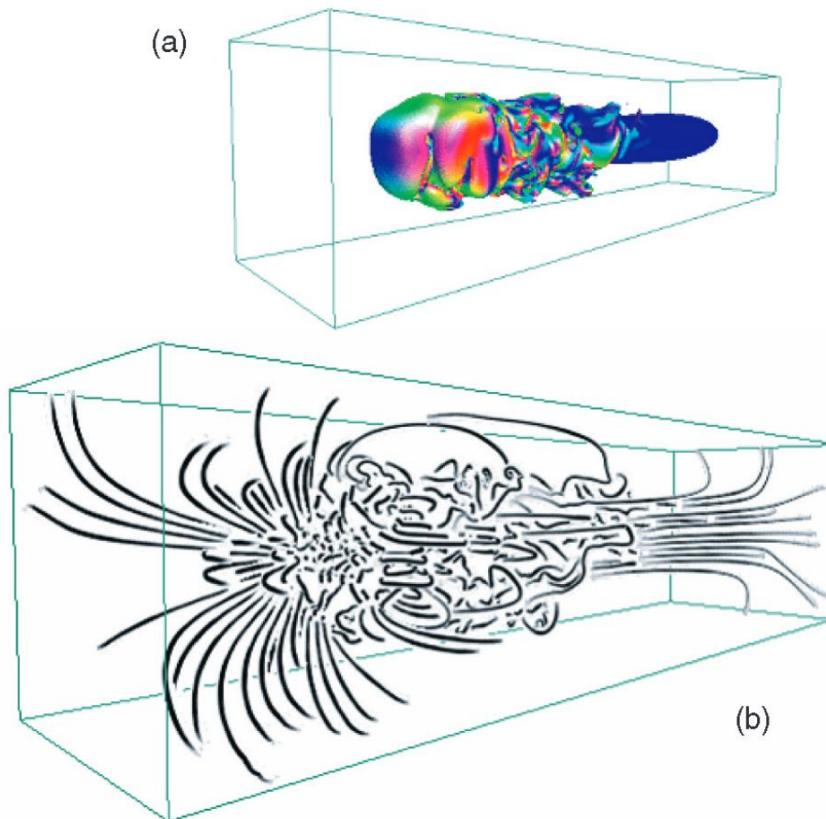


Fig. 6. (a) An isosurface of velocity magnitude colored by using the velocity  $(u, v, w)$  as  $(r, g, b)$ . (b) Streamlines generated from the isosurface with our image space method.

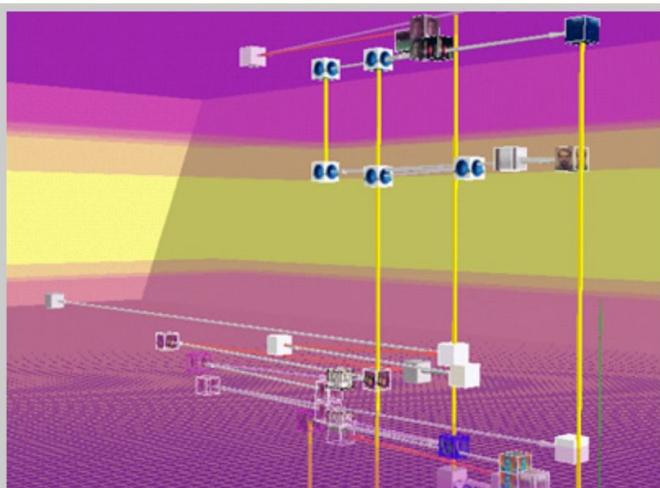
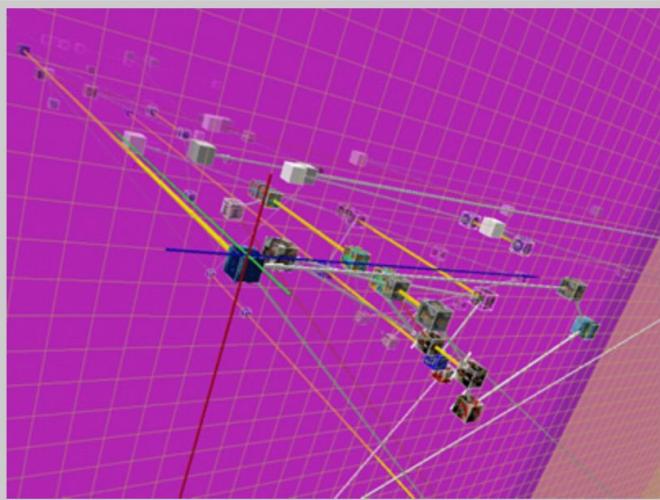
# ¿Cuándo se justifica usar 3D?

El uso de 3D es legítimo para datos que son 3D+.

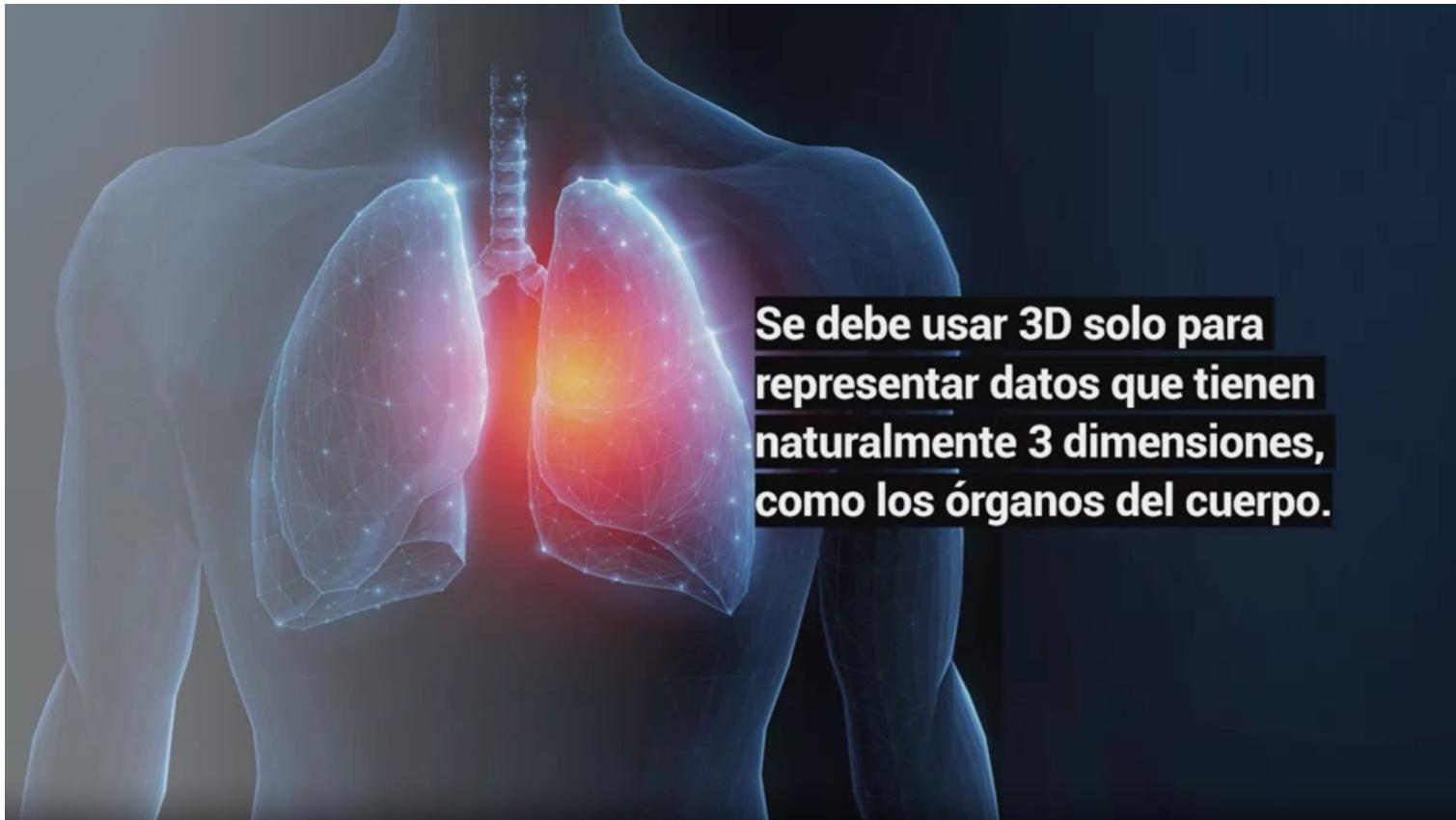
Ahora bien, para datos abstractos, se necesita una buena justificación. No es lo mismo una visualización de datos espaciales que de datos abstractos para los que no tenemos una referencia.

¿Con qué hay que tener cuidado? Cuando los datos son principalmente puntos (point clouds) o redes.

Frécon, E., & Smith, G. (1998, October). Webpath-a three dimensional web history. In *Proceedings IEEE Symposium on Information Visualization* (Cat. No. 98TB100258) (pp. 3-10). IEEE.



# ¿Cuándo se justifica usar 3D?



**Se debe usar 3D solo para representar datos que tienen naturalmente 3 dimensiones, como los órganos del cuerpo.**

# Un ejemplo de uso no justificado

En esta visualización, tenemos una serie temporal por cada día del año.

Cada día contiene el perfil de consumo eléctrico de un edificio.

La visualización en 3D permite mostrar todos los días del año en un solo gráfico. Sin embargo, **el volumen de cada día hace que la comparación entre períodos de tiempo sea prácticamente imposible debido a la oclusión y a la perspectiva.**

Van Wijk, J. J., & Van Selow, E. R. (1999). Cluster and calendar based visualization of time series data. In Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis' 99) (pp. 4-9). IEEE.

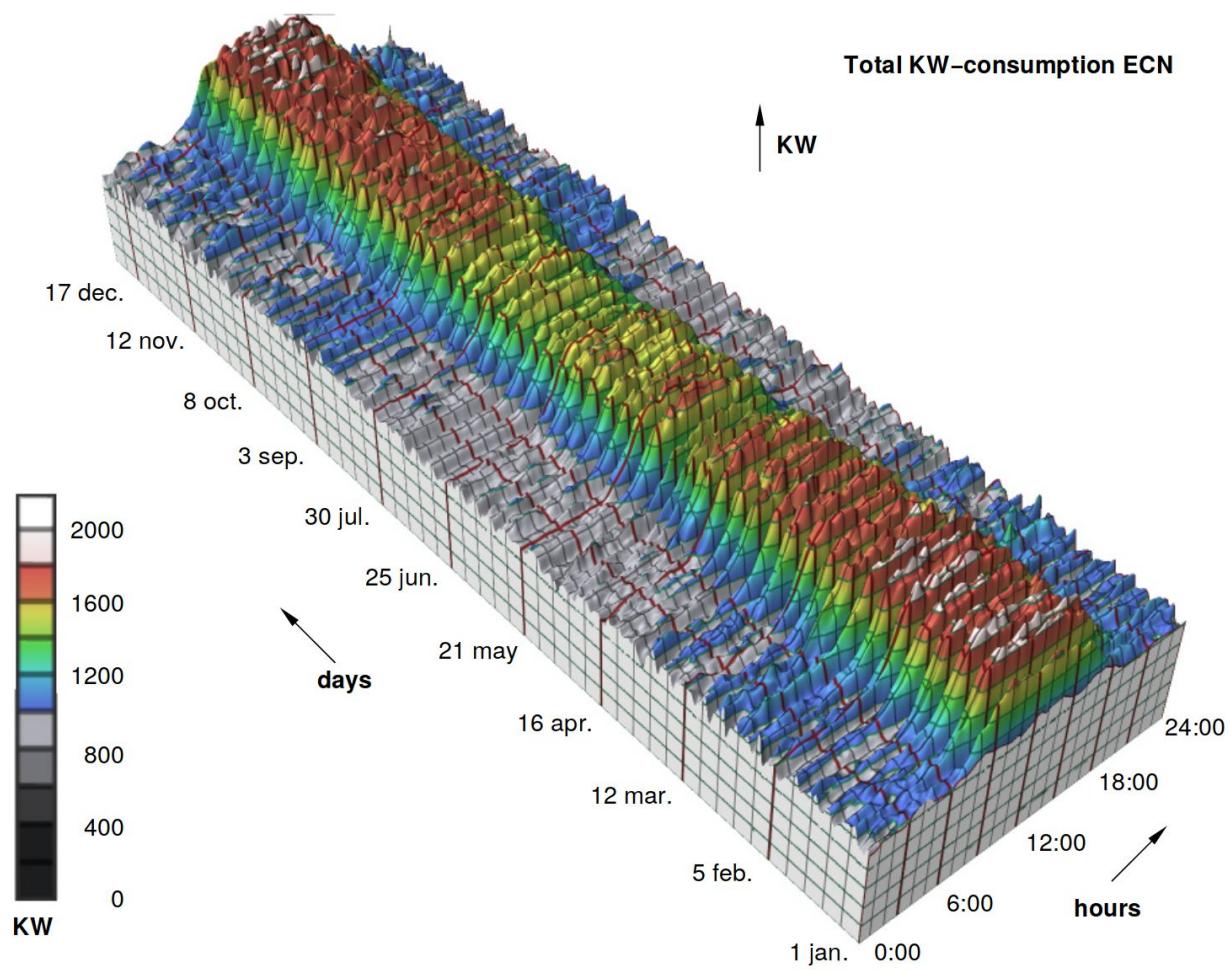


Figure 1: Power demand by ECN, displayed as a function of hours and days

# Solución

En el caso anterior, los autores decidieron transformar los datos para crear una visualización eficiente.

Primero, determinaron que en el dominio del problema (electricidad) era relevante definir **patrones de consumo eléctrico**. Así, hicieron clustering sobre las series temporales para poder decir que un día específico cumplió con cierto patrón.

Luego se visualizan esos patrones (series temporales superimpuestas en un gráfico 2D) y su distribución en el año (vista de calendario 2D). Como resultado, **¡es posible conocer los perfiles de uso energético y su distribución durante el año!**

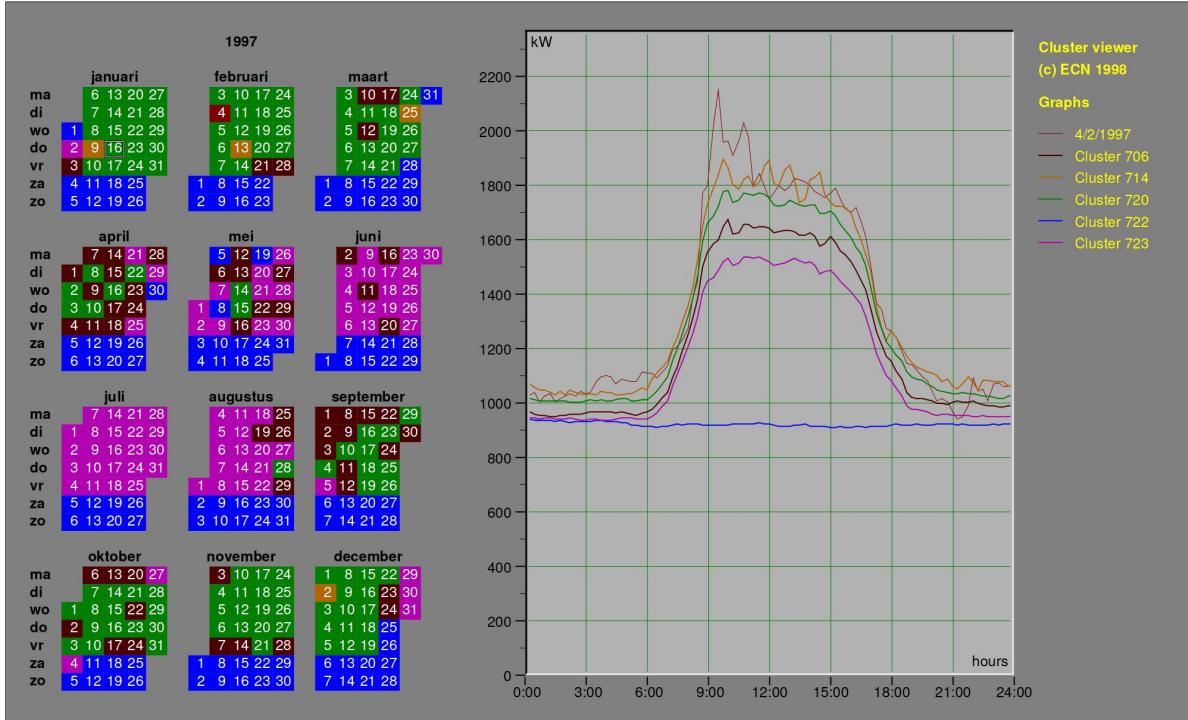


Figure 5: Cluster analysis of power demand by ECN

Van Wijk, J. J., & Van Selow, E. R. (1999). Cluster and calendar based visualization of time series data. In Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis' 99) (pp. 4-9). IEEE.

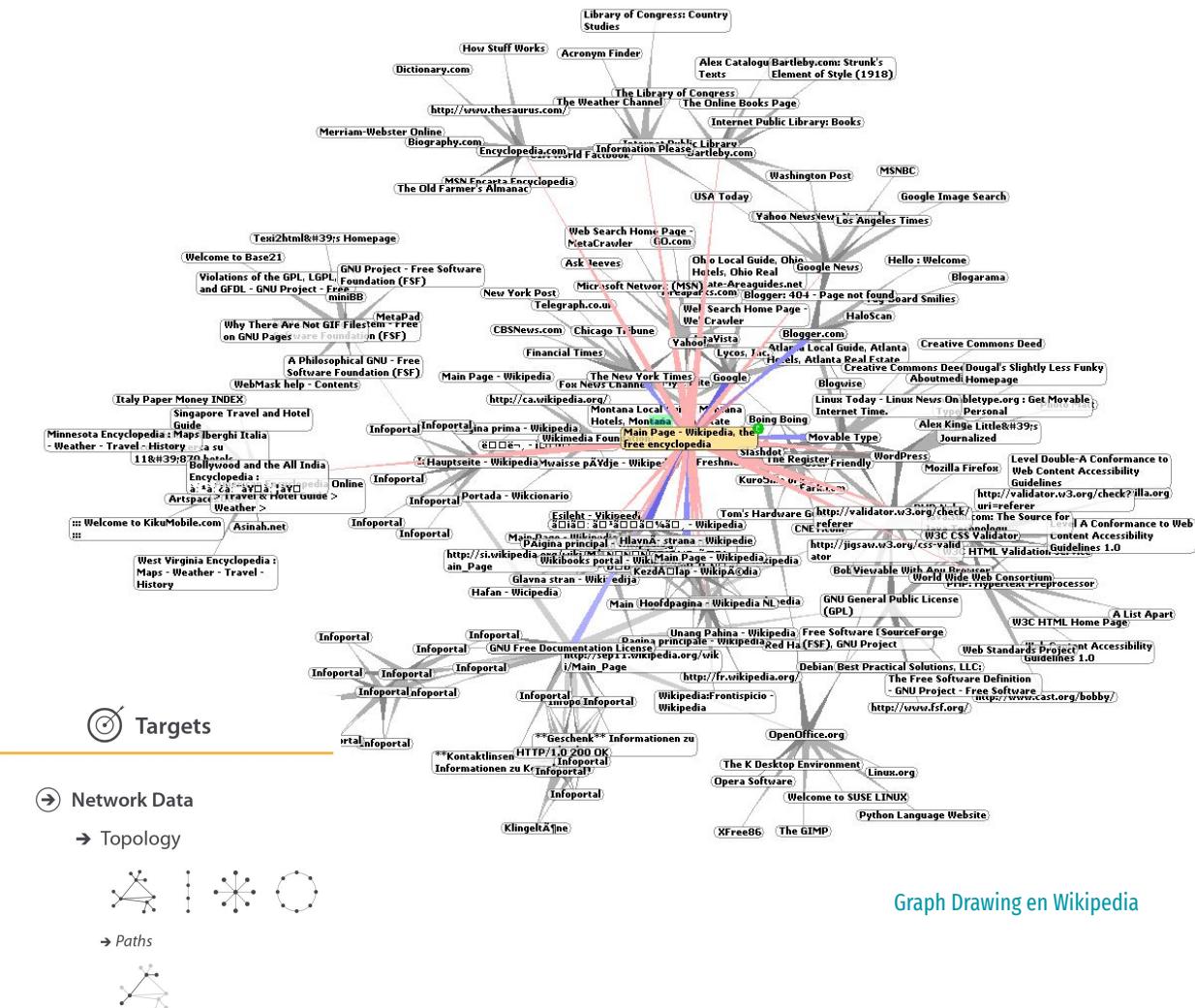
# 2D sin justificar

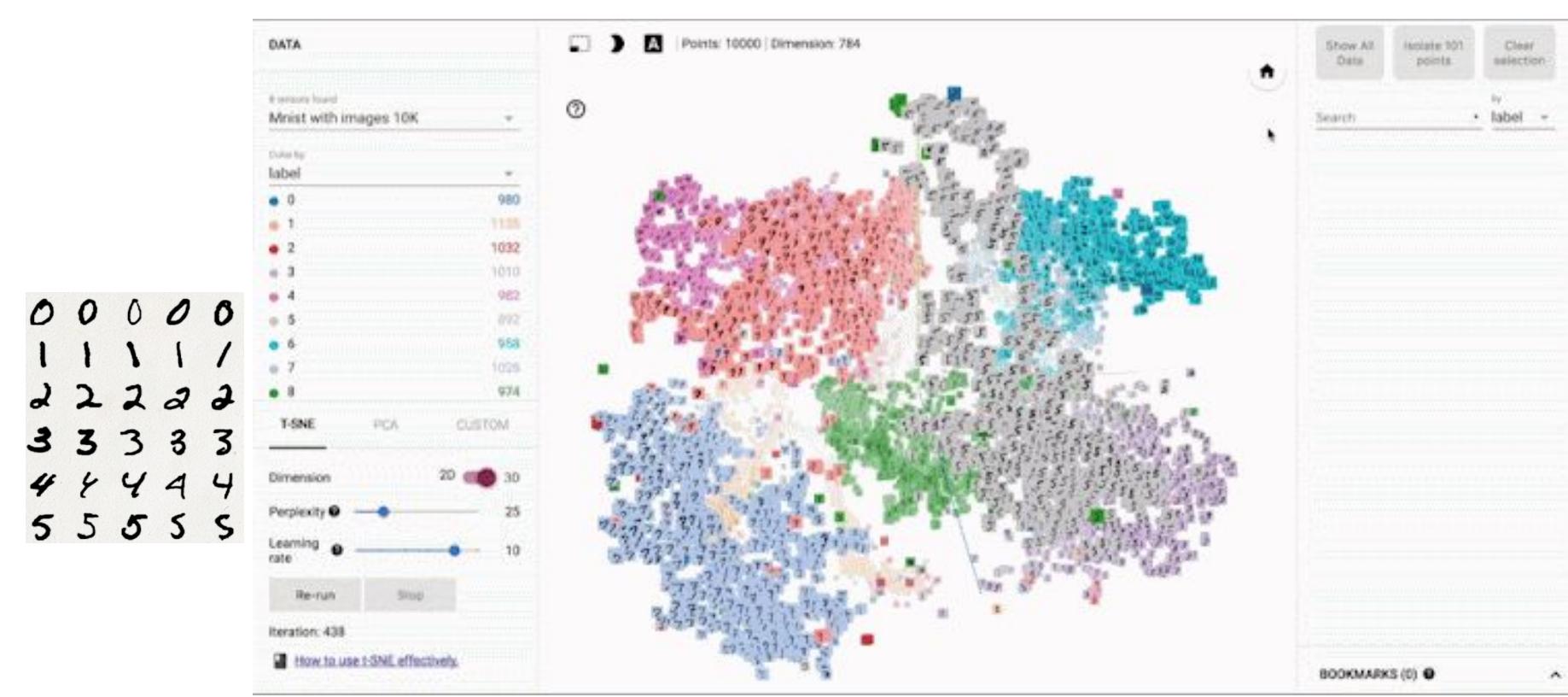
Piensa si los datos de la red requieren 2D, o si bien se puede utilizar una representación alternativa.

Esto es crucial si leer texto es importante para la tarea a ejecutar.

Debido a cómo funcionan los algoritmos de organización de redes (como force placement), la densidad de la red puede dificultar la lectura de las etiquetas de cada nodo.

Si la tarea es principalmente topológica, entonces el uso de 2D para la red está justificado: los costos son menores a los beneficios.





Muchos de los problemas mencionados se pueden aliviar con un buen diseño de interacción.

Fuente: <https://ai.googleblog.com/2016/12/open-sourcing-embedding-projector-tool.html>

<http://projector.tensorflow.org/>

# Prácticas Estructurales

**Visual Information Seeking Mantra** (“el mantra de la búsqueda visual de información”)  
propuesto por Ben Shneiderman.

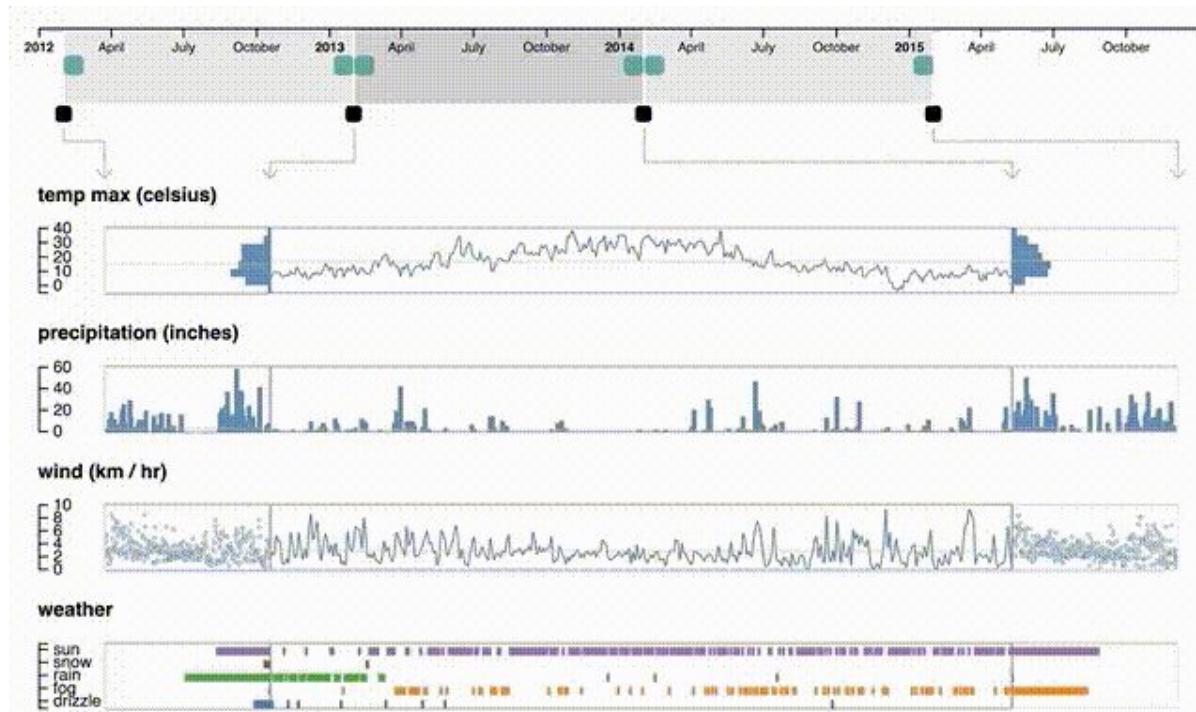
**“Vista global primero -> filtrar y hacer zoom -> detalles a medida que se necesiten”**  
(en inglés: *overview first, zoom and filter, then details-on-demand*).

# Focus+Context e Interactividad

Existe un tipo de técnica llamada focus+context que ejemplifica parte de este paradigma interactivo, en la que una visualización tiene al menos dos partes: una donde se muestra a grossó modo el dataset (context), y una donde se muestra con mucho detalle el contenido de la parte de interés (focus).

Desde la vista general procedemos a una vista donde los datos se visualizan con más detalle, pero con un criterio que los ha filtrado.

En un sistema interactivo, esto permite decidir dónde fijarse, qué llama la atención, dónde puede ser necesario profundizar.



<https://gotz.web.unc.edu/research-project/periphery-plots/>

# Mejor ver que memorizar

:Cognición externa vs memoria interna!

Es fácil comparar elementos desplazando nuestro foco de atención entre vistas paralelas.

Es difícil comparar elementos si solamente veo uno y tengo que recordar el otro.

En una animación debemos recordar lo que hemos visto. Se suele utilizar en narrativa utilizando visualización, o para mostrar la transición entre dos estados. Si hay muchos estados, cada uno con varios cambios - es preferible utilizar **small multiples**, como en el ejemplo, más una visualización que recalque los cambios.



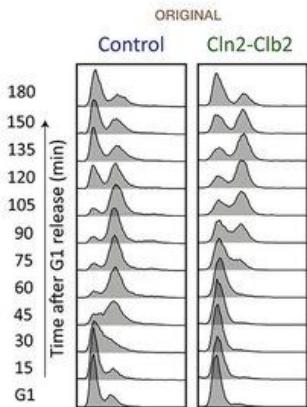
Fig. 2: The Cerebral display of the TLR4 graph ( $V=91$ ,  $E=124$ ) with associated LPS and LPS+LL-37 time series. The small multiples show an overview of all 8 experimental conditions. The most noticeable differences between the LPS and the LPS+LL-37 condition occur at hour 4. By selecting the hour 4 conditions, the main window shows the computed difference between the two conditions.

Barsky, A., Munzner, T., Gardy, J., & Kincaid, R. (2008). Cerebral: Visualizing multiple experimental conditions on a graph with biological context. *IEEE transactions on visualization and computer graphics*, 14(6), 1253-1260.

# Mejor ver que memorizar

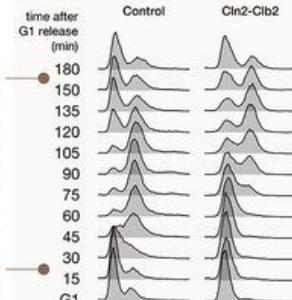
## WHAT ARE THE IMPORTANT DIFFERENCES?

The reader won't always know where to look or know what is important. Emphasizing the quantity of interest makes the purpose of the figure clear.



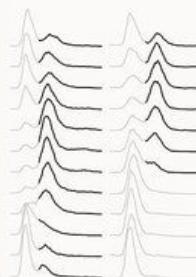
Courtesy: Deniz Pirnici Ercan & Frank Uhlmann, Francis Crick Institute, UK

REDESIGN 1  
Remove boxes.  
Reduce spacing to improve comparisons, taking care to avoid creating clutter.

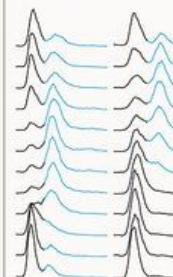


Avoid vertical axis labels at all cost. Don't restate the obvious. A time axis (especially with arrow head) is redundant since values obviously establish an axis direction upwards (dictated by convention).

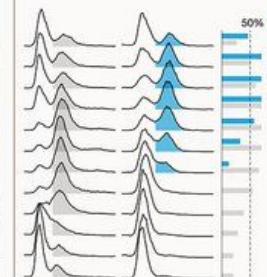
REDESIGN 2  
The appearance of the second peak is what is important in this figure. Hint it with darker stroke (possibly mildly thicker).



REDESIGN 3  
Hue even more powerfully separates the left and right peaks.



REDESIGN 4  
Now it is obvious that the ratio of areas of the second peak is the property of interest. Area histogram and ratios are rough approximations.



Simple functions of previously drawn quantities can be expressed using the encoding scheme.

Here "ratio of areas" would help viewer by computing useful ratios if more fundamental quantities are shown graphically.

Here, the area under the curve is worth showing because it demonstrates the distribution of area over time.

It is not necessary to show the ratio graphically, since the numbers themselves are easy to compare and interpret.

Un ejemplo de tener que memorizar es evaluar las diferencias entre dos gráficos. **A pesar de que se muestren en paralelo, tenemos que mirar de un lado a otro y recordar para medir la diferencia.** Una manera de evitar usar la memoria **es destacar visualmente la diferencia relevante para el análisis.** Fuente

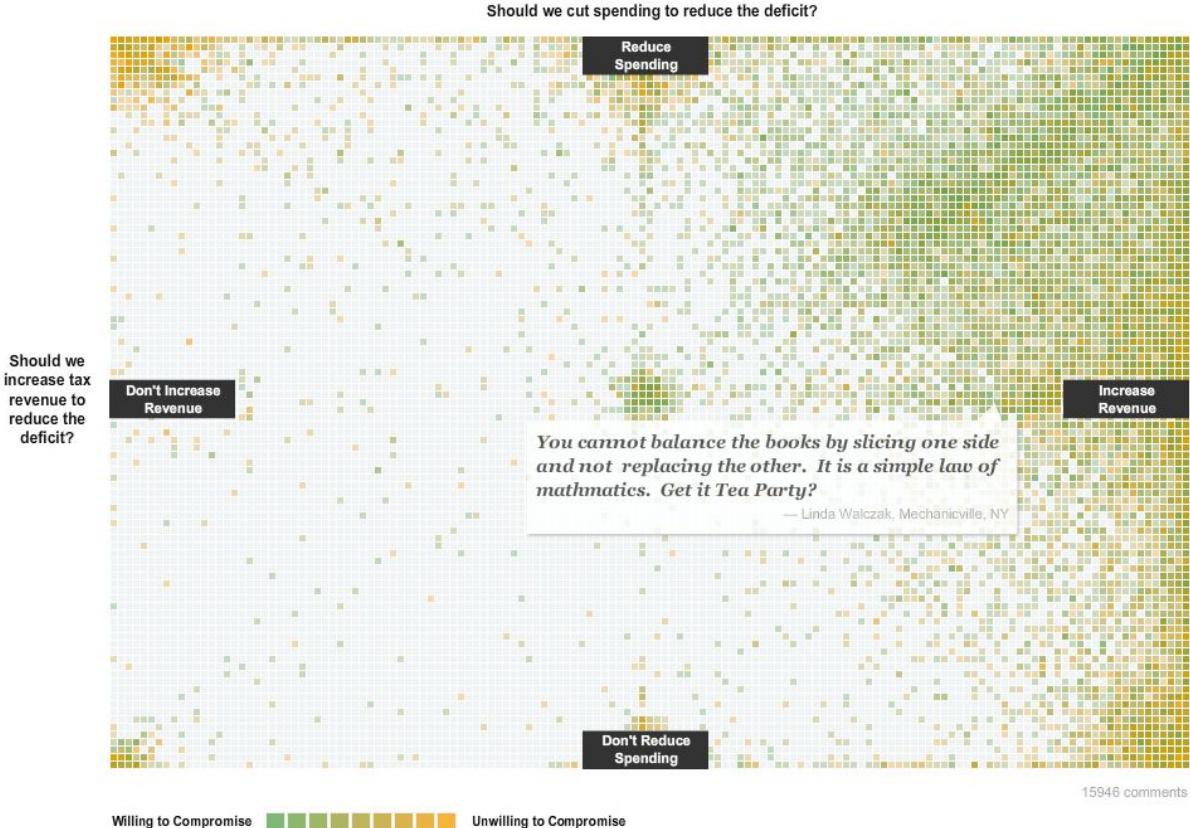
Redesigns by Martin Krzywinski <http://mkweb.bcgsc.ca>

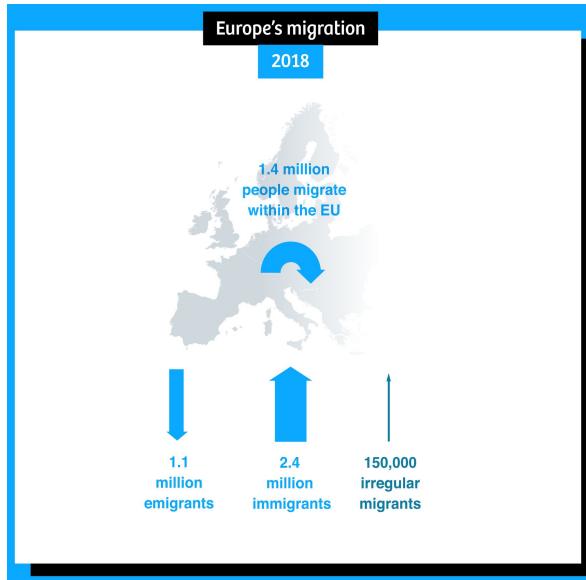
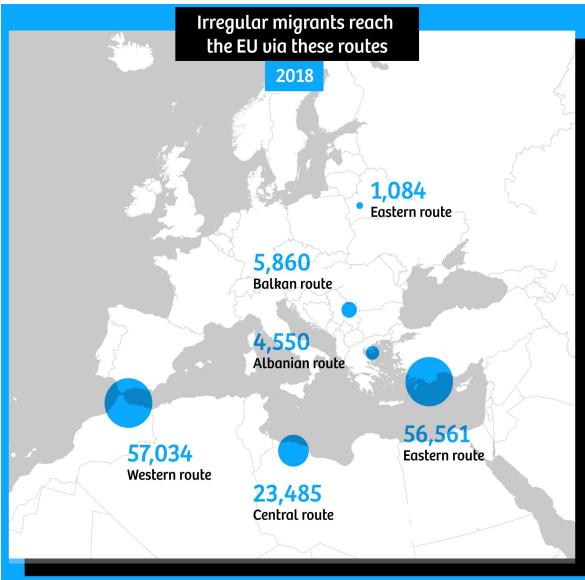
# Prácticas Narrativas

# Toda historia se cuenta desde una perspectiva

Esta visualización de tipo heatmap permitía a las personas que leen el periódico posicionar su opinión respecto al gasto público de los Estados Unidos entre dos ejes: reducir gasto o no hacerlo, incrementar impuestos o no hacerlo.

Sin embargo, la opción de incrementar gasto no está disponible, así como la de reducir impuestos. No es lo mismo decir no reducir gasto que incrementar gasto, y ciertamente la posición en el eje y variará en función de como sea etiquetado ese eje en el heatmap.





Estos tres mapas visualizan los mismos datos. ¿Entregan el mismo mensaje?

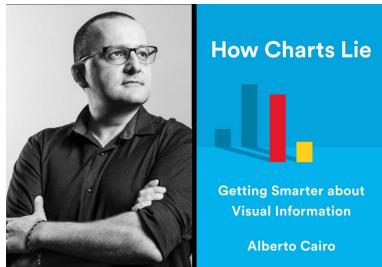
Fuente:

<https://thecorrespondent.com/664/how-maps-in-the-media-make-us-more-negative-about-migrants/738023272448-bac255ba>

# ¡Es fácil engañar!

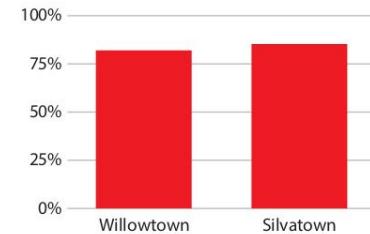
La codificación visual es tan potente que incluso cuando estamos conscientes de las falacias visuales, al ver un gráfico que utiliza esas técnicas somos propensos/as a creer en lo que dice.

Aquí tres ejemplos a seguir para presentar una visualización en la que se puede confiar.

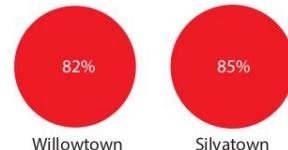


## Correct versions

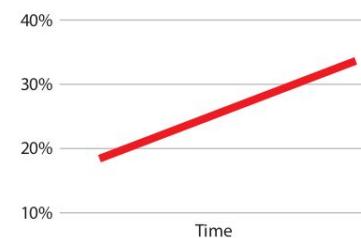
Percentage of population with access to safe drinking water



Percentage of population with access to safe drinking water



Access to safe drinking water by minority ethnic group

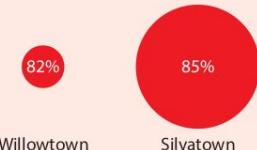


## Deceptive versions

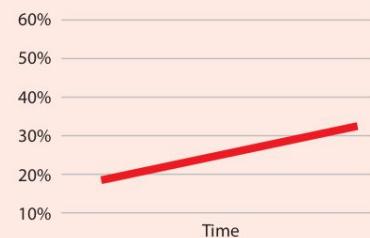
Percentage of population with access to safe drinking water



Percentage of population with access to safe drinking water



Access to safe drinking water by minority ethnic group

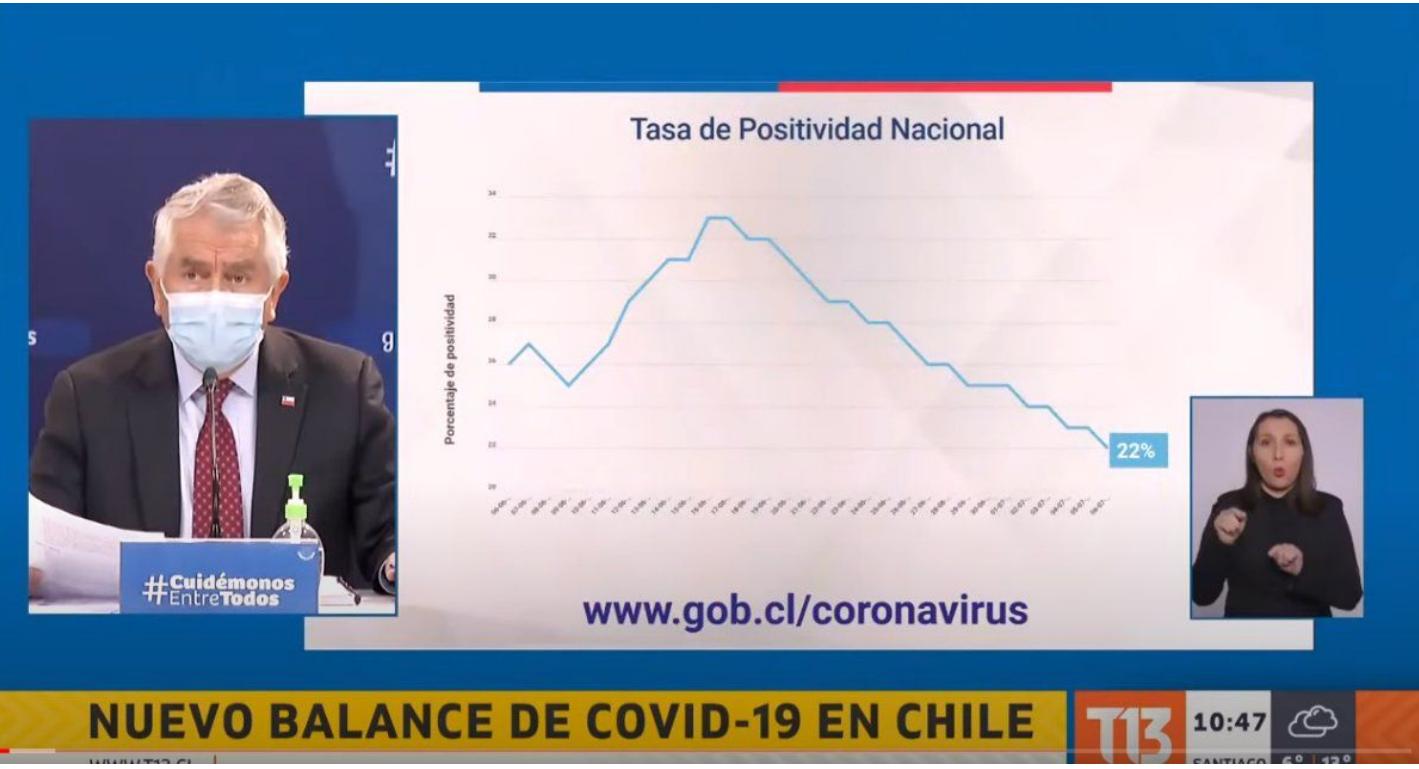




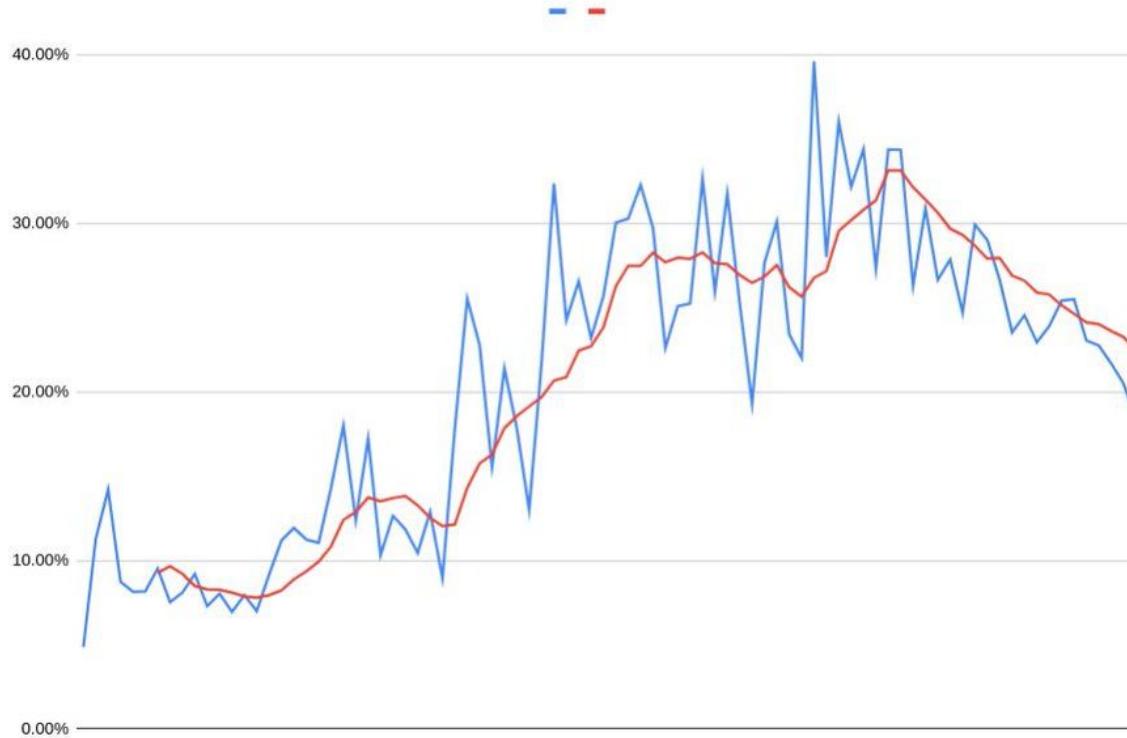
¿Por qué esta visualización desinforma?



Un gráfico de barras que no distorsiona nuestra percepción al tener 0 como base y mostrar todos los años. Fuente: CNN.



¿Por qué esta visualización desinforma?



Tasa de positividad de COVID-19 en Chile el 7 de Julio de 2020. Fuente: [@juancriolivares](#).

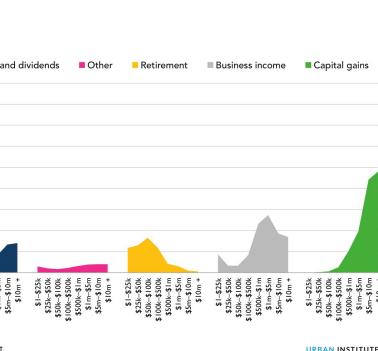
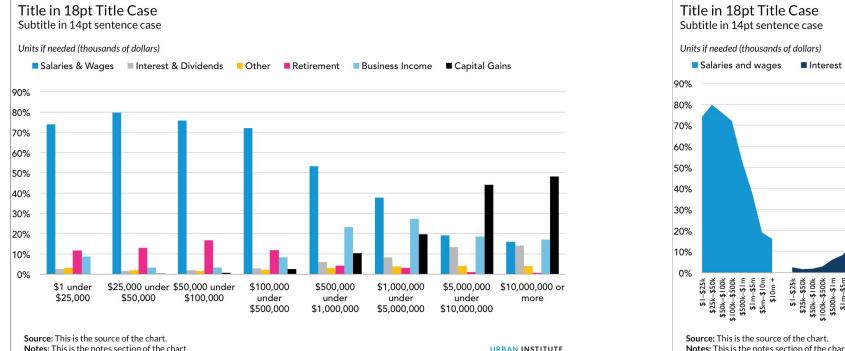
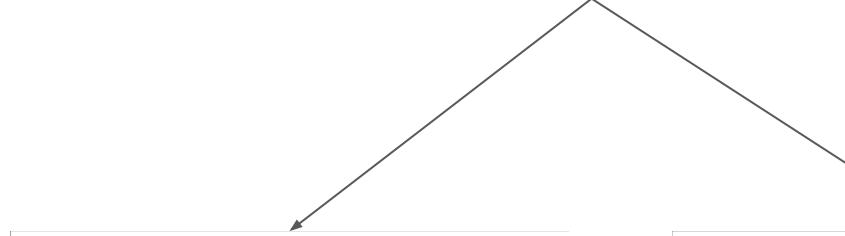
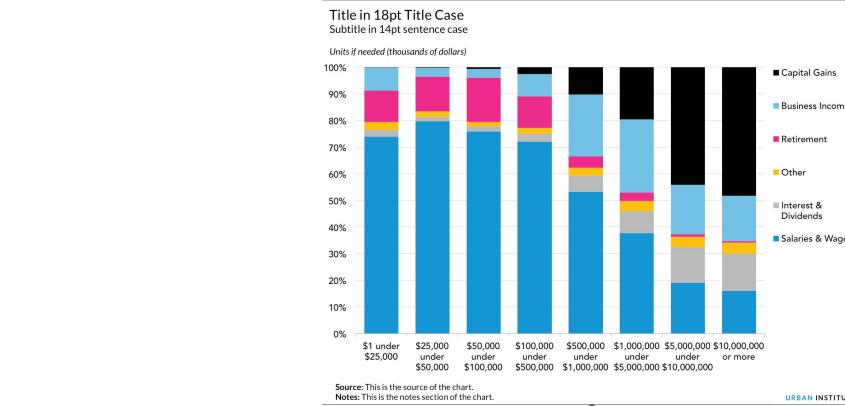
# Pautas de Alineamiento

Una organización suele tener un conjunto de *guidelines* o pautas de alineamiento con buenas prácticas de presentación y análisis de datos. Así el trabajo y sus resultados se presentan de manera **consistente y efectiva**.

Estas guías suelen incorporar criterios a la hora de diseñar la codificación visual de una visualización.

## Ejemplo:

<https://urbaninstitute.github.io/graphics-styleguide/>



# Evaluación

# Evaluar: Validar la Efectividad de Un Sistema

¿Qué significa que un sistema esté validado?

- ¿En términos de *software engineering*? Que el sistema funcione adecuadamente.
- ¿En términos de las tareas a realizar? Que el sistema permita realizar las tareas de manera efectiva (resultados correctos, mejores que alternativas, etc.).
- ¿En términos de desempeño? Que el sistema permita realizar las tareas de manera eficiente (en menos tiempo, con mayor precisión, etc.).

La única manera de saber si elegimos bien es realizando una **evaluación**. Hay herramientas formales para ello, provenientes de áreas como **User Research**.

- Ojo: **cada persona interpreta una visualización de manera diferente**, el hardware puede no ser el mismo para dos usuarios/as diferentes.

# ¿Cómo saber si una visualización es adecuada?

Además de abstracción de datos/tareas, hay dos capas adicionales relevantes para la visualización:

**Dominio:** todo lo relacionado con el contexto de las personas que usarán la visualización.

**Algoritmia:** la eficiencia en la implementación de la visualización, por ej., si es un sistema interactivo.

En cada capa es importante evaluar si se ha hecho un buen trabajo, porque errores en una se propagan a las otras.

## Domain situation

Observe target users using existing tools

## Data/task abstraction

## Visual encoding/interaction idiom

Justify design with respect to alternatives

## Algorithm

Measure system time/memory

Analyze computational complexity

Analyze results qualitatively

Measure human time with lab experiment (*lab study*)

Observe target users after deployment (*field study*)

Measure adoption

# ¿Cómo saber si una visualización es adecuada?

También es importante el orden en el que se evalúan los problemas.

Este esquema menciona las amenazas en cada capa y la manera de evaluar y validar que se ha hecho un buen trabajo.

Al avanzar a través de las capas **dominio - datos/tareas - codificación visual - implementación** debemos estar constantemente evaluando si lo estamos haciendo bien. ¿Por qué? Porque los errores se propagan y no se detectan hasta que sea demasiado tarde.

- ! **Threat** Wrong problem
- ✓ **Validate** Observe and interview target users

- ! **Threat** Wrong task/data abstraction

- ! **Threat** Ineffective encoding/interaction idiom
- ✓ **Validate** Justify encoding/interaction design

- ! **Threat** Slow algorithm
- ✓ **Validate** Analyze computational complexity

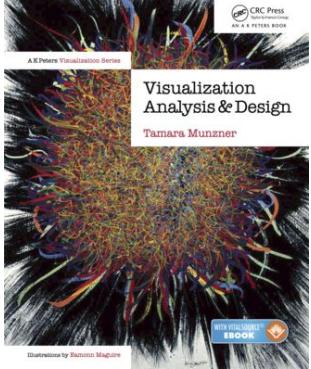
**Implement system**

- ✓ **Validate** Measure system time/memory

- ✓ **Validate** Qualitative/quantitative result image analysis  
*Test on any users, informal usability study*
- ✓ **Validate** Lab study, measure human time/errors for task

- ✓ **Validate** Test on target users, collect anecdotal evidence of utility
  - ✓ **Validate** Field study, document human usage of deployed system
- 
- ✓ **Validate** Observe adoption rates

# ¿Preguntas?



Esta clase incluye material del libro **Visualization Analysis & Design** de Tamara Munzner.  
<http://www.cs.ubc.ca/~tmm/vad/book/>