

Rules of thumb

Visualización de Información
IIC2026

Profesor: Denis Parra

¿Qué es un *rule of thumb*?

- Es posible definirlo como un principio o una **guía**, basado en la **experiencia/práctica** más que en la teoría.
- La idea de esta clase es sintetizar el conocimiento, sacado desde **estudios empíricos** que se desarrollan en esta área.

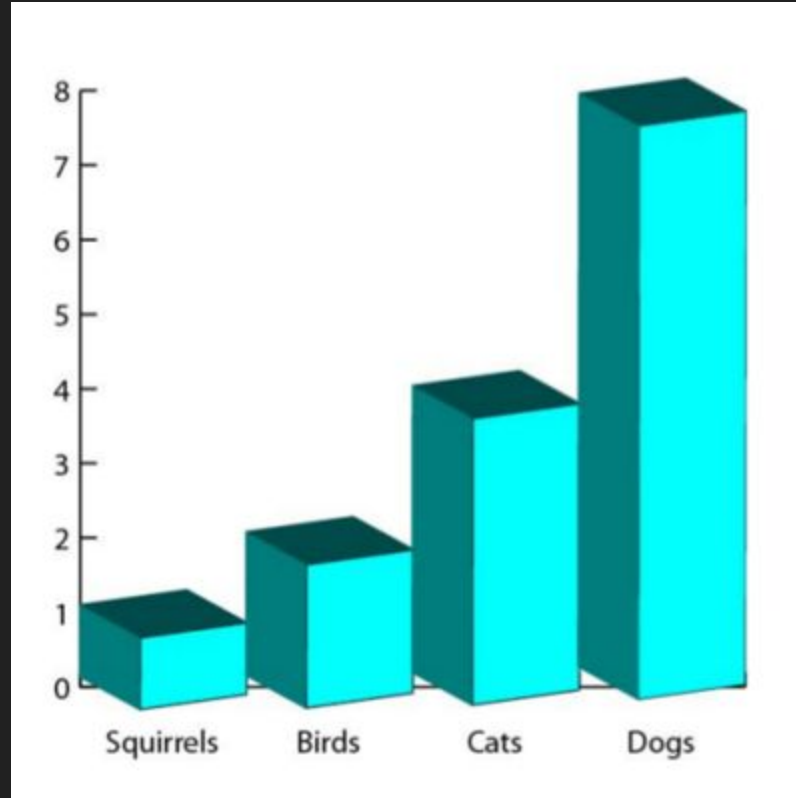
Data ink ratio (Tufte)

$$\text{data ink ratio} = \frac{\text{data ink}}{\text{total ink used}}$$

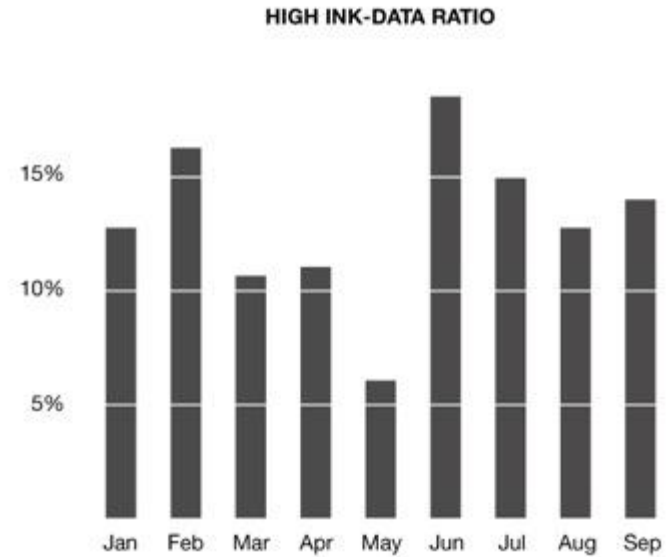
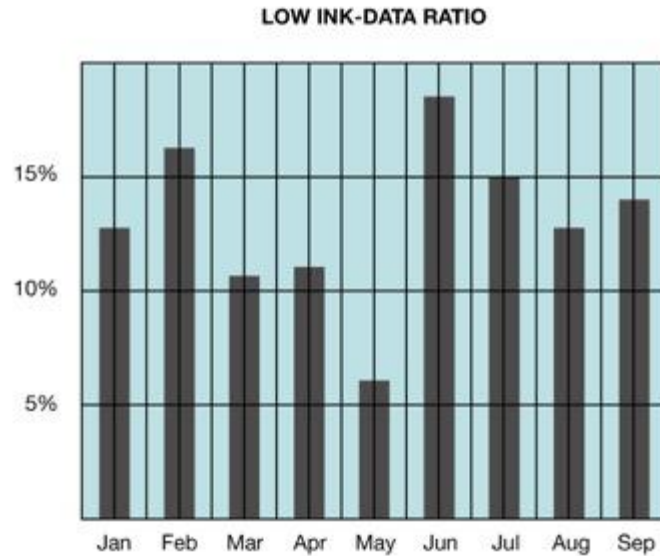
En nuestras visualizaciones, buscaremos **maximizar** este *ratio*, para que cada marca/canal que usemos tenga una razón de existir.

Llevándolo a un extremo (i.e. *ratio* = 1), cada pixel debe estar justificado.

Data ink ratio (Tufte)



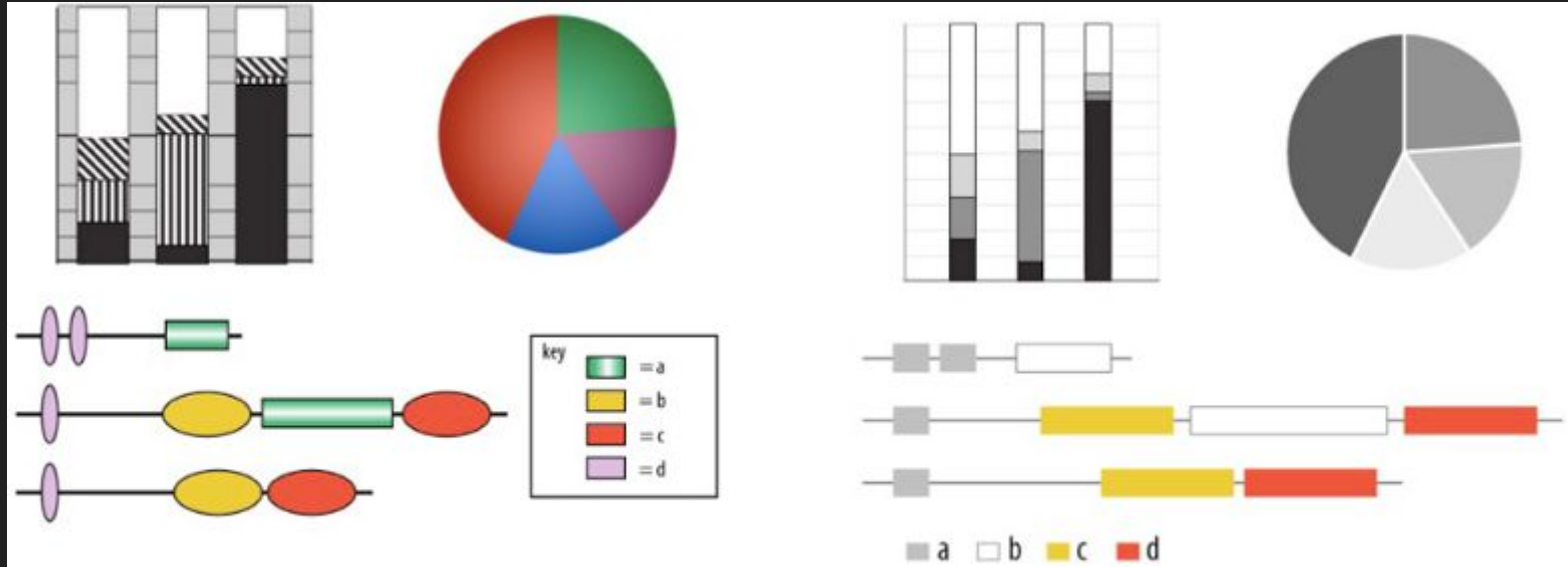
Data ink ratio (Tufte)



Data ink ratio (Tufte)

Remove
to improve
(the **data-ink** ratio)

Más ejemplos



Quitamos colores, formas, texturas que no aportan información.

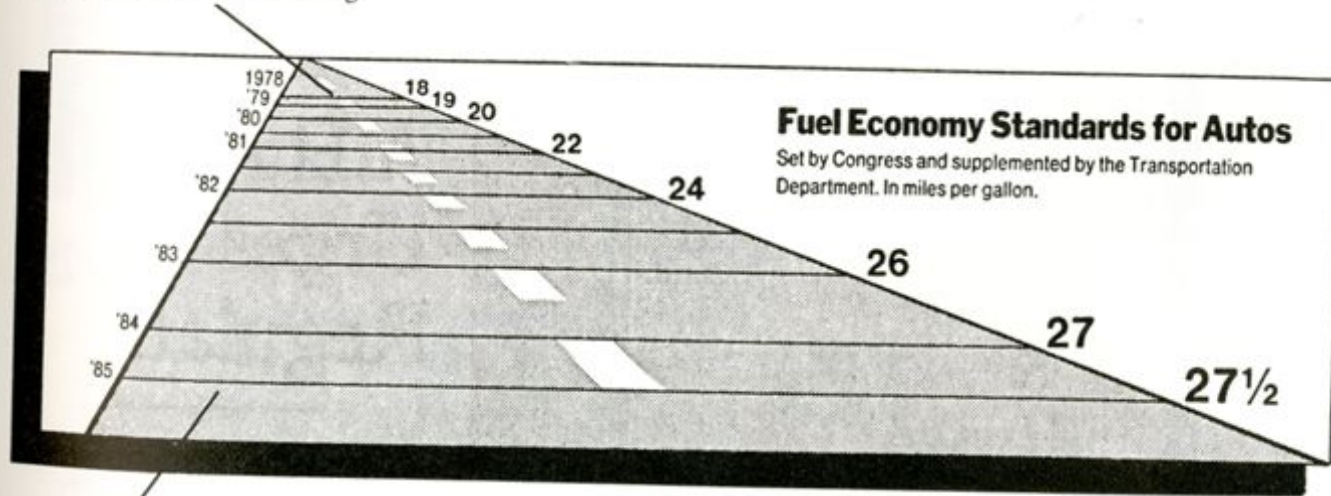
Lie factor (Tufte)

$$\text{lie factor} = \frac{\text{size effect in graphic}}{\text{size effect in data}}$$

- La tasa de cambio entre los datos debe ser **fielmente reflejada** por el efecto que se muestra gráficamente.
- En este caso, deberíamos apuntar a un factor de 1 (i.e. mismo efecto).

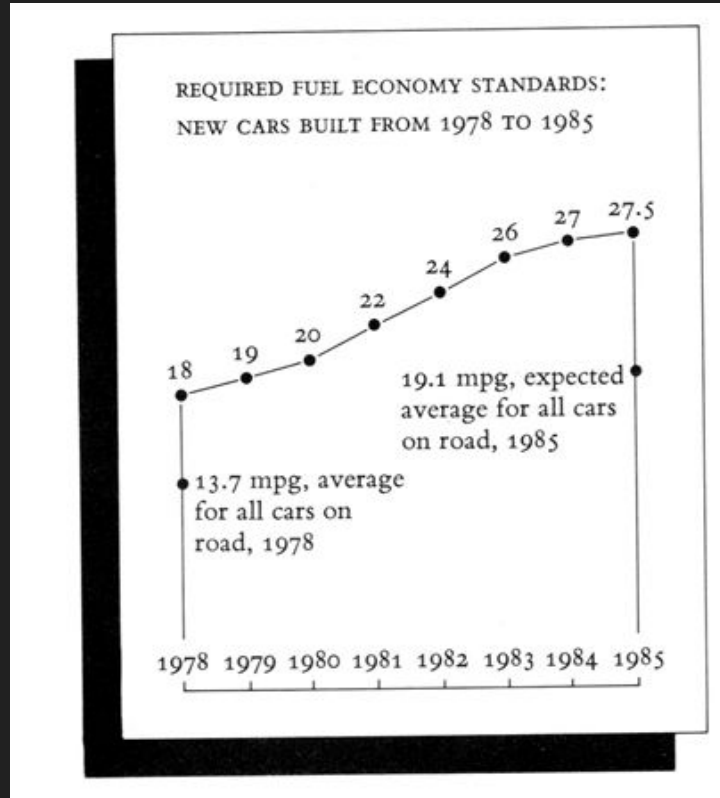
Lie factor (Tufte)

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.

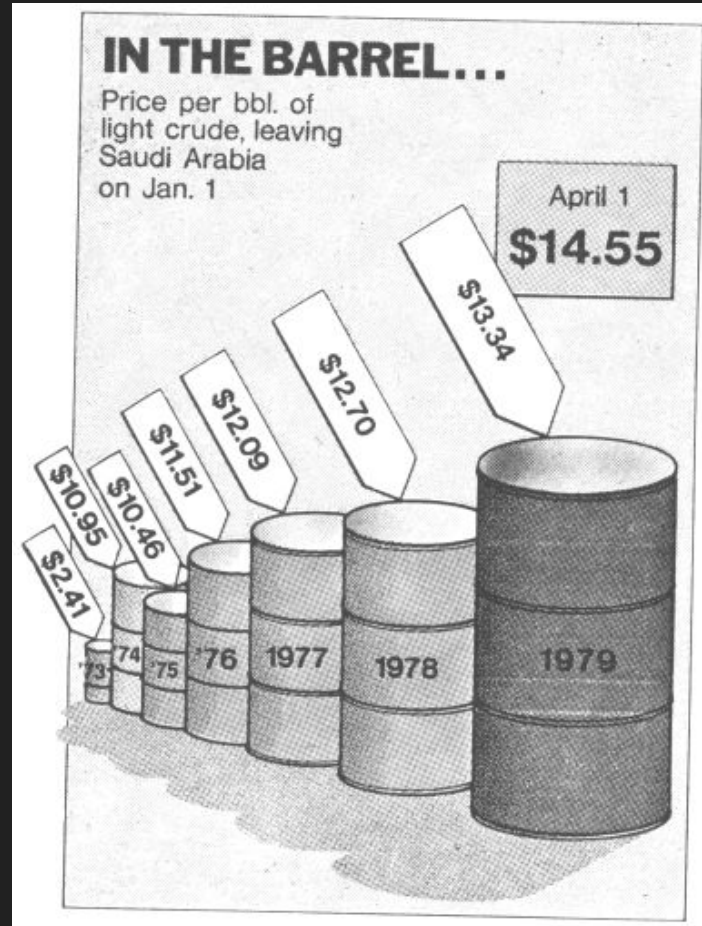


This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

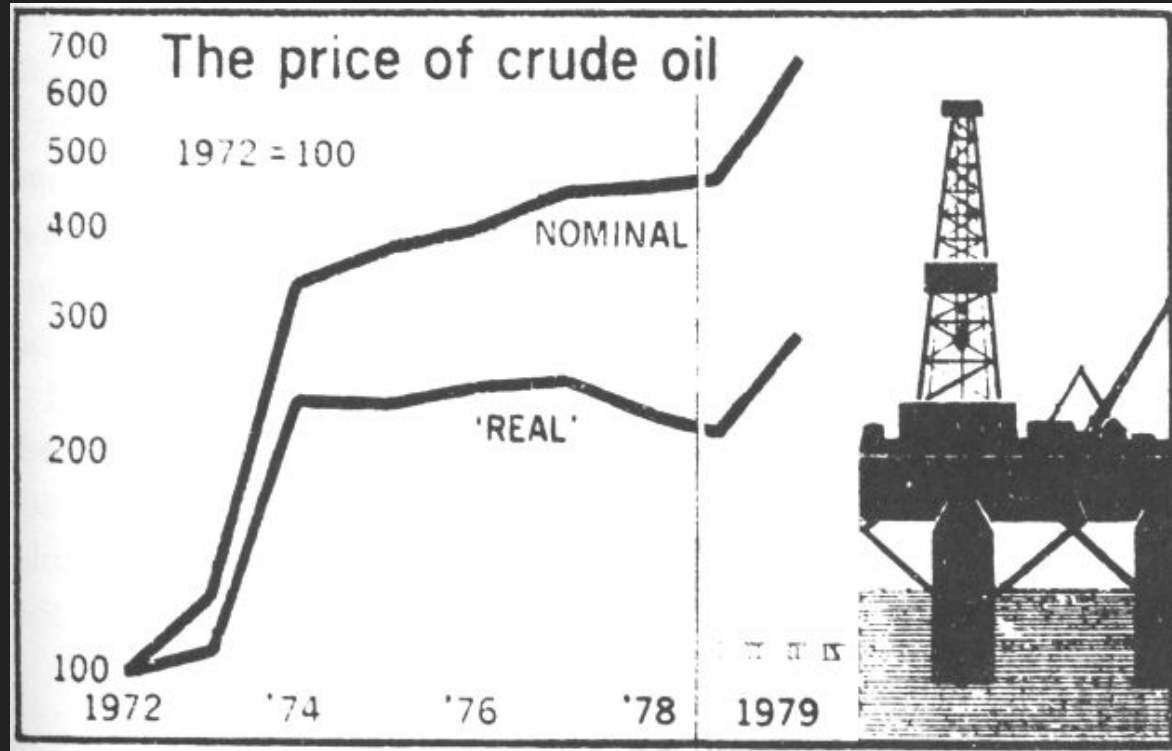
Lie factor (Tufte)



Lie factor (Tufte)

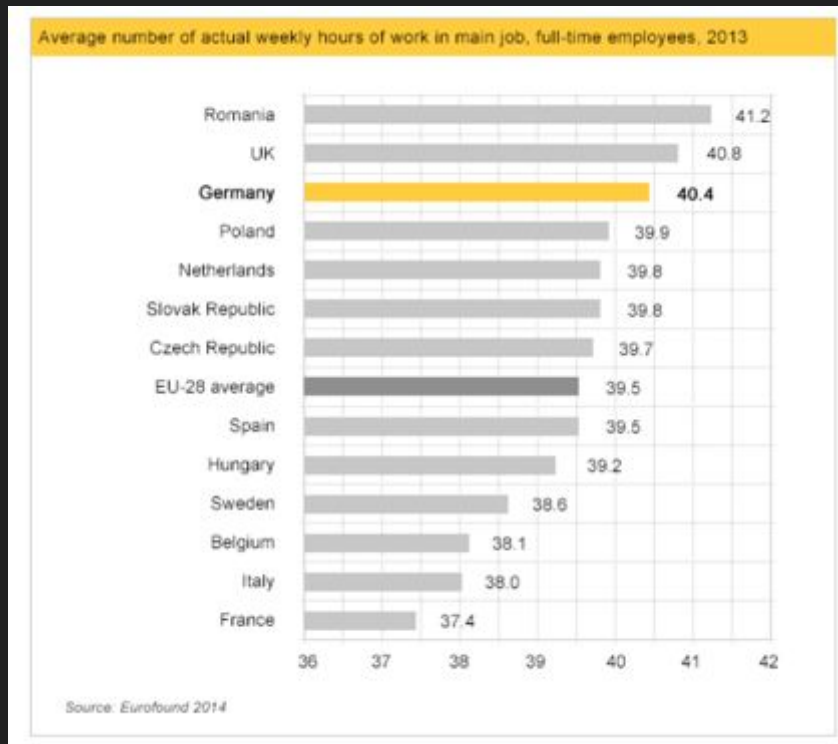


Lie factor (Tufte)



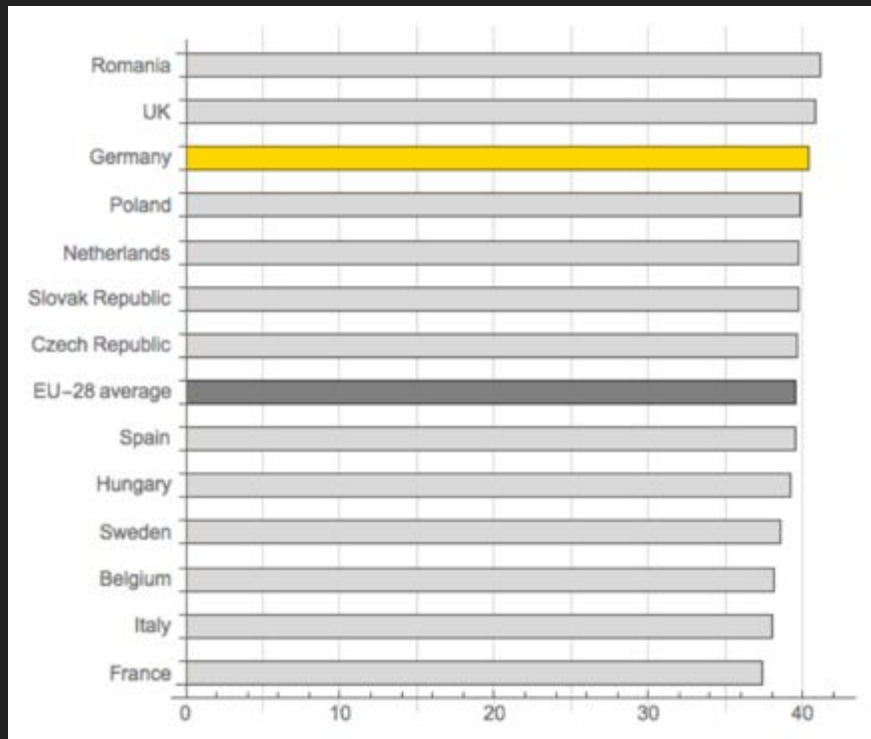
Ejes engañosos

- Relacionado con el *lie factor*, los ejes de este diagrama de barras **no comienzan en cero**.
- A raíz de esto, pareciera que los alemanes trabajan casi el triple que los franceses, siendo que sólo hay un factor de 1,08 de diferencia.

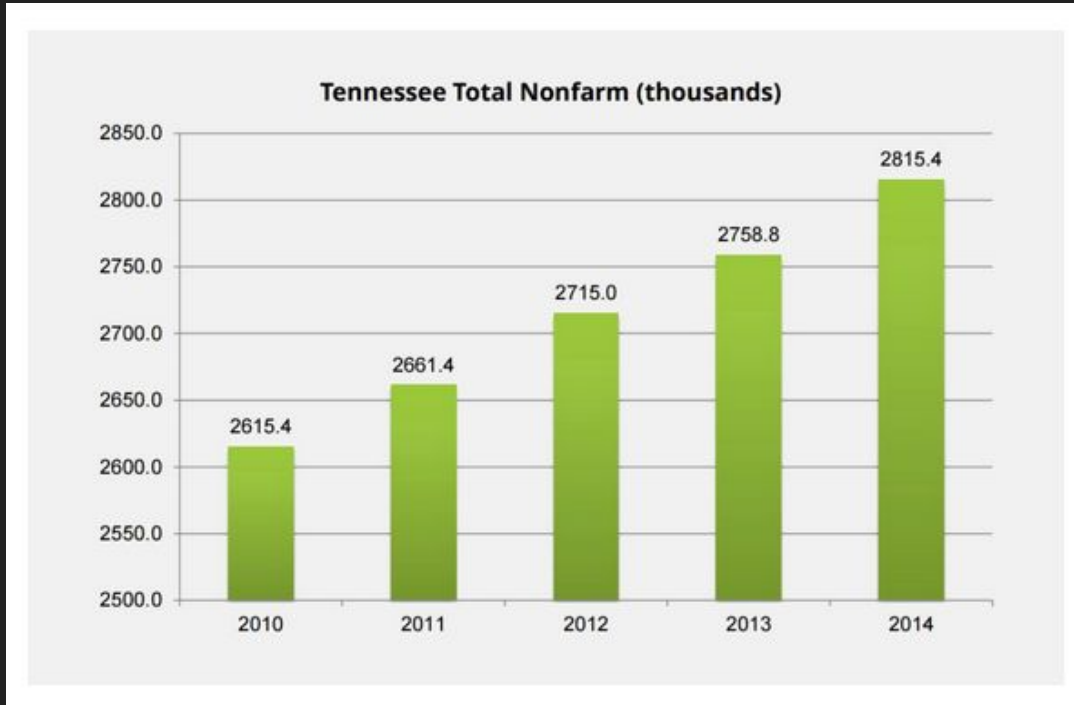


Ejes engañosos

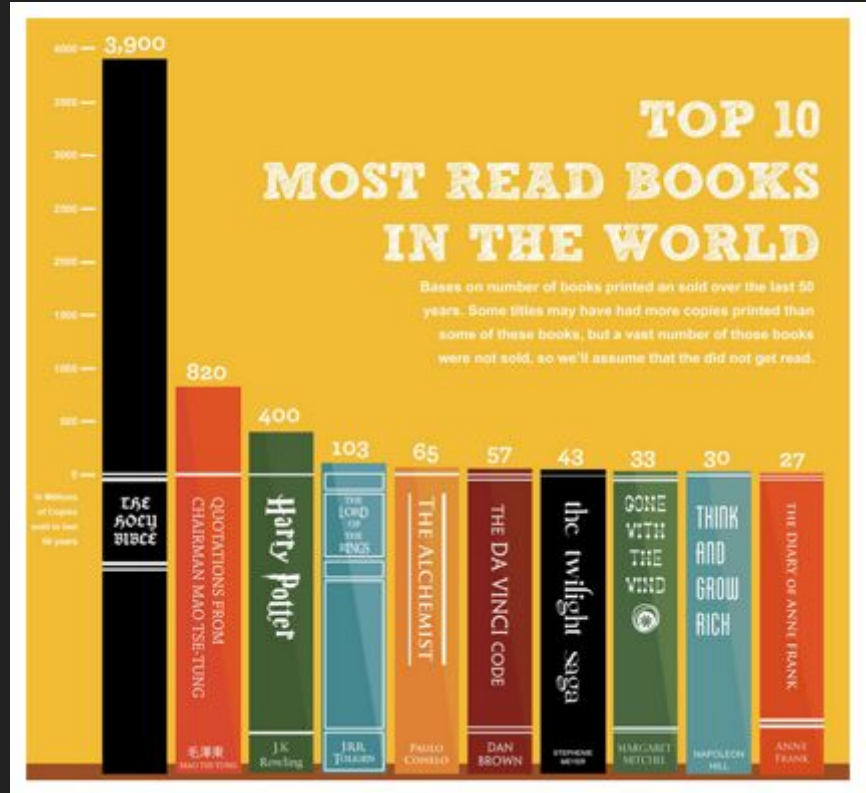
- Aquí tenemos el mismo *dataset* representado con ejes que parten desde cero.
- Ahora las diferencias de las barras se ajustan acorde a las diferencias en el *dataset*.



Ejes engañosos (más ejemplos)



Ejes engañosos (más ejemplos)



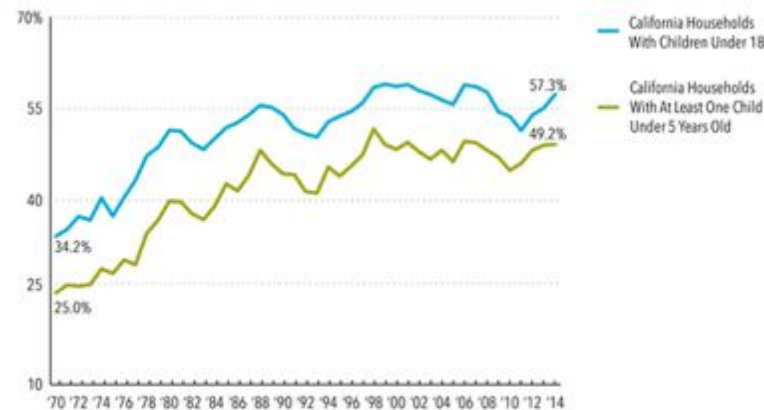
Otro ejemplo



Ejes engañosos (más ejemplos)

- Este gráfico de línea no parte con su eje en cero. Sin embargo, no hay problema con eso, puesto que, a diferencia de las barras, este *line graph* busca contar **otra historia**.
- El diagrama de barra se enfoca en la **diferencia de magnitud** entre las categorías, mientras que acá buscamos mostrar el **cambio de la variable dependiente** (eje y).

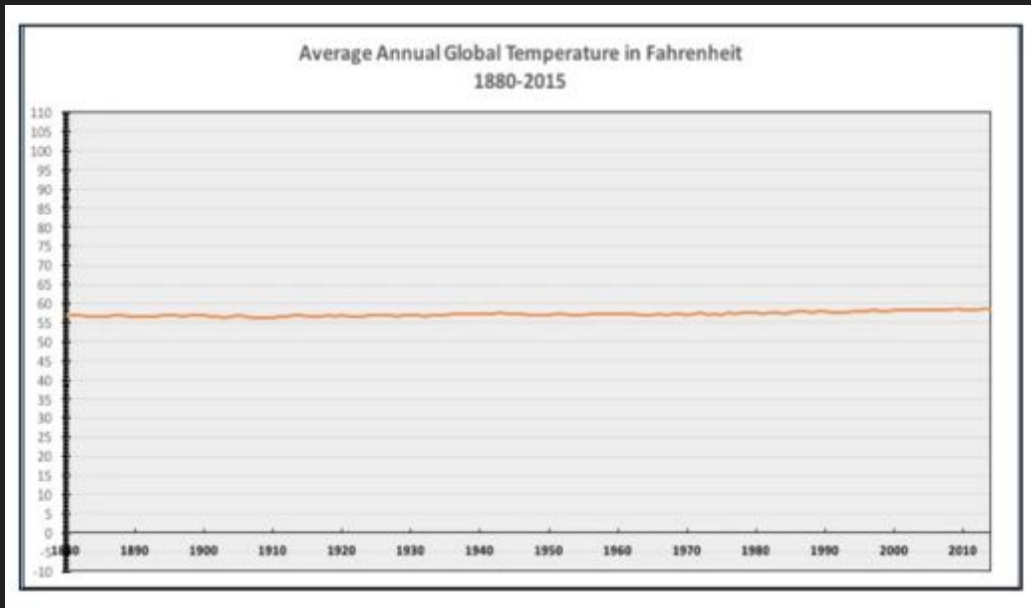
Percentage of California Households Where All Parents Work, 1970 to 2014



Note: A "household where all parents work" includes single-parent households and dual-earner households. Parents include stepparents and adoptive parents.
Source: Budget Center analysis of US Census Bureau data

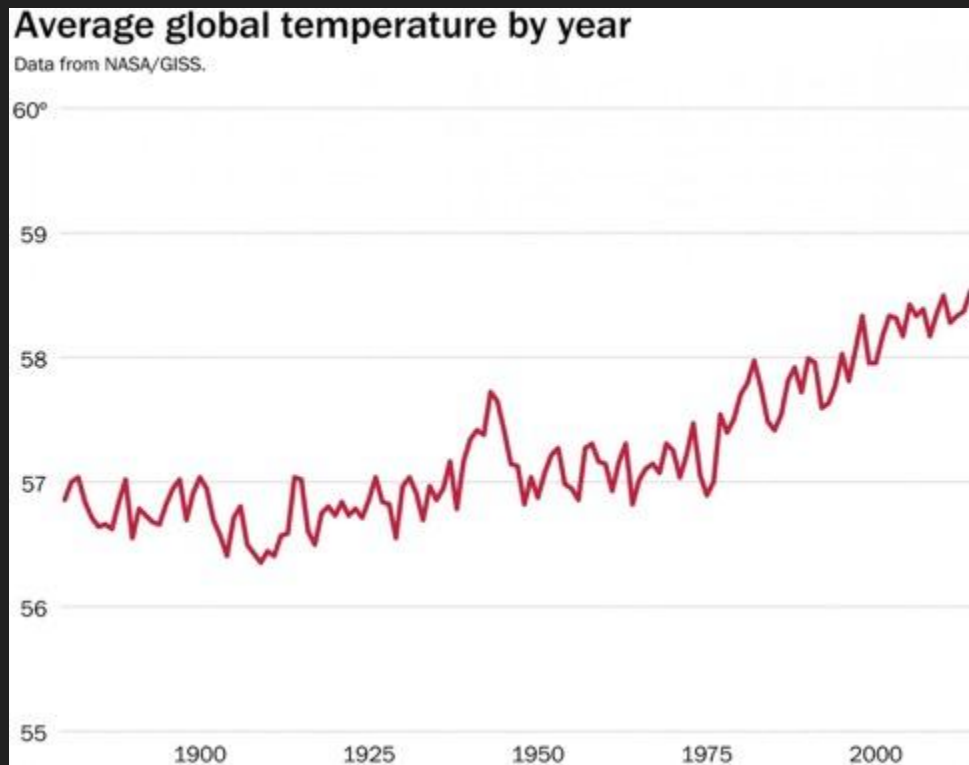
Ejes engañosos (más ejemplos)

- Incluso, mostrar el cero en el eje puede ser engañoso, ya que se intentan ocultar la **tasa con que ocurren los cambios**.
- En este caso, este gráfico no muestra los cambios, sino la **magnitud absoluta** que es inconsistente con la historia que estaban tratando de narrar.



Ejes engañosos (más ejemplos)

- Esta es una representación más correcta de lo que ocurre con la temperatura promedio global de nuestro planeta, ya que no es importante la magnitud, sino el **cambio**.

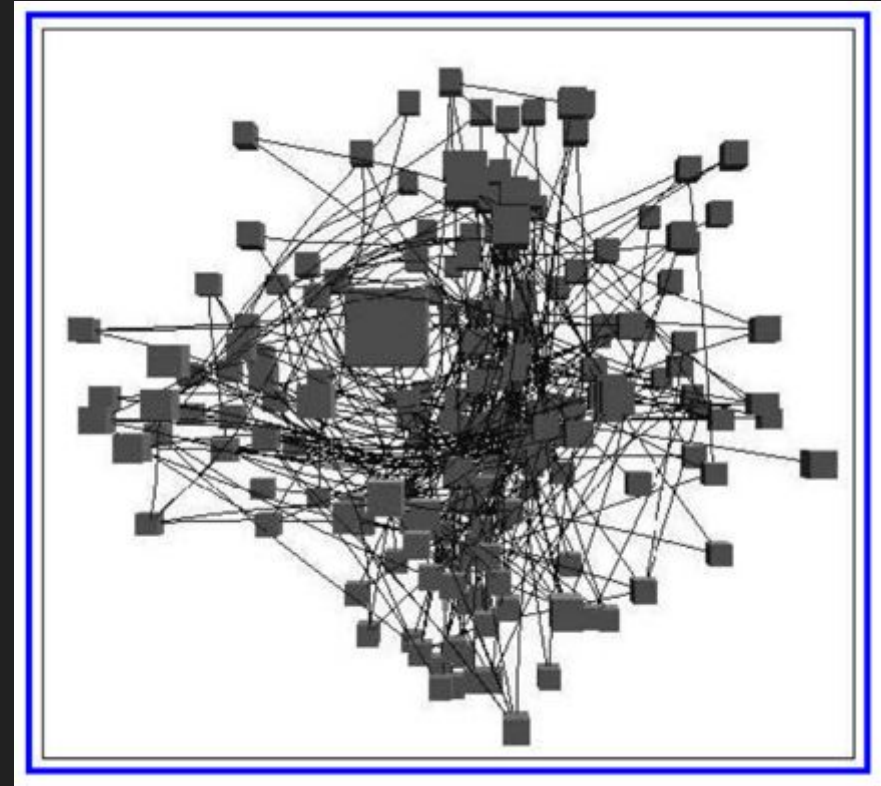


No al 3D injustificado

- Generalmente, las personas creen que si en dos dimensiones se ve bien, entonces en tres debe ser mejor aún —claro, después de todo, vivimos en un mundo tridimensional.
- Sin embargo, existen muchas **dificultades** relacionadas al *encoding* de información usando una tercera dimensión espacial (i.e. **profundidad**), ya que tiene diferencias importantes con respecto a las otras dos dimensiones.
- El uso de una tercera dimensión sí se justifica cuando el usuario debe ejecutar tareas (e.g. percepción de formas 3D) que están relacionadas con una estructura que **inherentemente tiene tres dimensiones**.

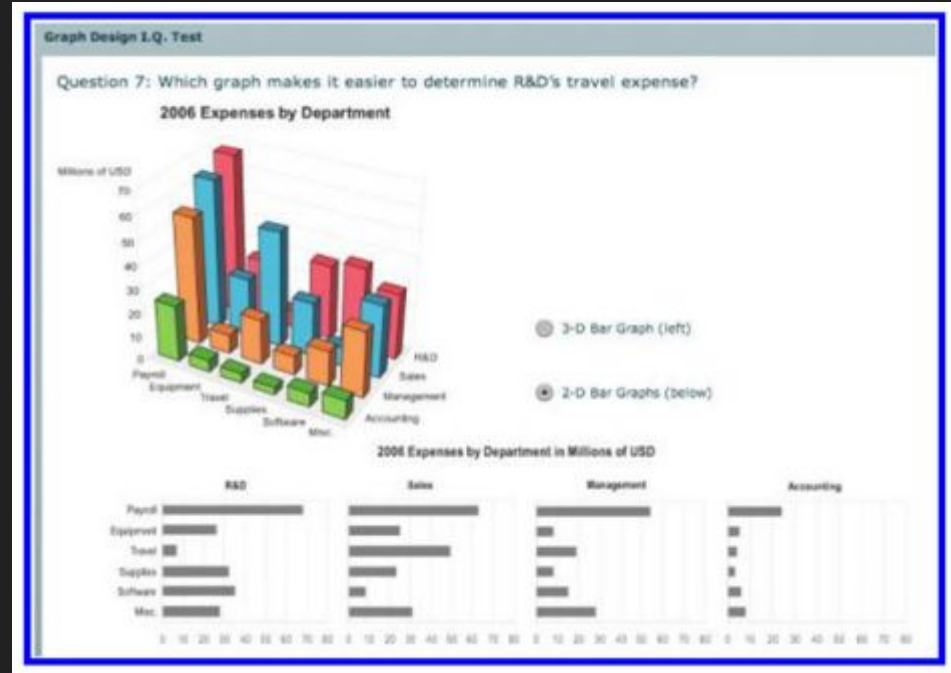
No al 3D injustificado (oclusión)

- El problema de este grafo con nodos y aristas es el de **oclusión**, en donde nodos quedan ocultos detrás de otros.
- Si bien es posible agregar algún tipo de navegación interactiva, el costo asociado a esto es el **tiempo**.

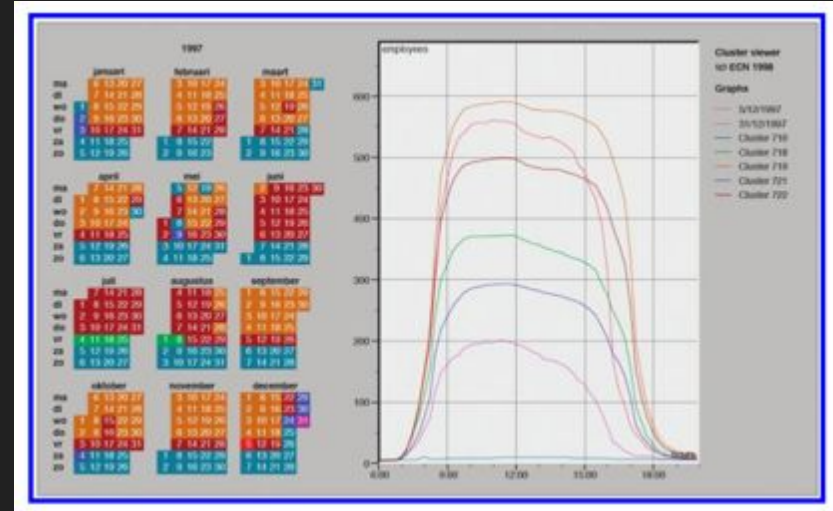
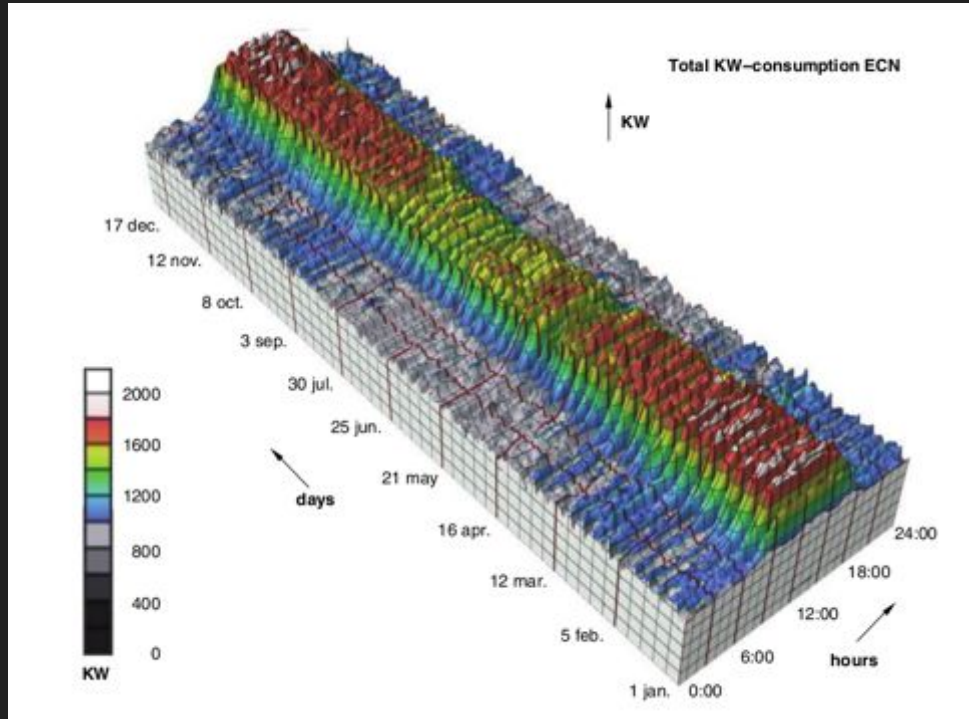


No al 3D injustificado (distorsión por perspectiva)

- Los objetos que están a mayor distancia **se perciben más pequeños** (e.g. línea de tren).
- En este ejemplo, tenemos dos variables categóricas y otra numérica; sin embargo, usar 3D **no es una buena opción**.
- Por la perspectiva (y también por oclusión), **cuesta comparar** los tamaños de las barras.



No al 3D injustificado (buscar alternativas)



No al 3D injustificado (buscar alternativas)

- En el ejemplo de la izquierda, hay problemas de **oclusión** y **distorsión**. Además, sólo es posible notar que existe un cambio en las horas de trabajo y la variación por estación entre verano e invierno.
- En el de la derecha, se crearon nuevos datos a partir de un *clustering* sobre las curvas más parecidas, logrando un promedio entre ellas. Aquí no hay oclusión ni distorsión entre las curvas, lo que permite una rápida comparación de ellas.
- Asimismo, el calendario ya es una marca tradicional y exitosa para mostrar patrones temporales.

No al 2D injustificado

- De forma análoga, mostrar datos en plano **también debe ser justificado**, comparado con la alternativa de una única dimensión (e.g. una lista).
- Las listas tienen varias ventajas: pueden exponer información (como etiquetas de texto) en un espacio mínimo; además, las listas son una herramienta excelente para **tareas de lookup**, cuando están ordenadas de forma apropiada.
- Por ejemplo, buscar un *label* específico será **más fácil en una lista**, que una representación 2D de *node-link*, en donde el usuario tendrá que buscar nodo por nodo. Sin embargo, si la tarea realmente requiere entender la **estructura topológica** de la red, entonces, en ese caso, sí vale la pena mostrar las relaciones en el plano.

“Overview first, details on demand”

- Esta frase, que es de Ben Shneiderman [1996], es un citado *guideline* que hace énfasis en la interacción de dos requisitos en una visualización: tener un *overview* y conocer los *detalles*.
- Un *idiom* que entrega un *overview* permite darle al usuario cierta *conciencia* (o conocimiento) sobre el espacio de información. Según el *what-why-how analysis* de Tamara, vendría a tomar el objetivo de *summarize*.
- Un objetivo típico en el *overview design* es mostrar todos ítems del *dataset* de forma simultánea, *sin la necesidad de navegar* (con *panning* o *scrolling*).

“Overview first, details on demand”

- Cuando el *dataset* es suficientemente grande, alguna forma de **reducción** es necesaria para mostrar todo de una misma vista: *filtering* y *aggregation*.
- También es común tener la vista con **detalles junto a la del overview**, logrando el *idiom* de *focus-and-context* (como un *minimap*).

Responsiveness is required

- La **latencia** de interacción importa de forma considerable en el diseño de un sistema de visualización.
- Nuestra reacción a la latencia **no ocurre** en un *continuum* de tiempo; no obstante, nuestro **nivel de irritación crece** mientras las operaciones van tomando más y más tiempo.

Time Constant	Value (in seconds)
perceptual processing	0.1
immediate response	1
brief tasks	10

Table 6.1. Human response to interaction latency changes dramatically at these time thresholds. After [Card et al. 91, Table 3].

Responsiveness is required (visual feedback)

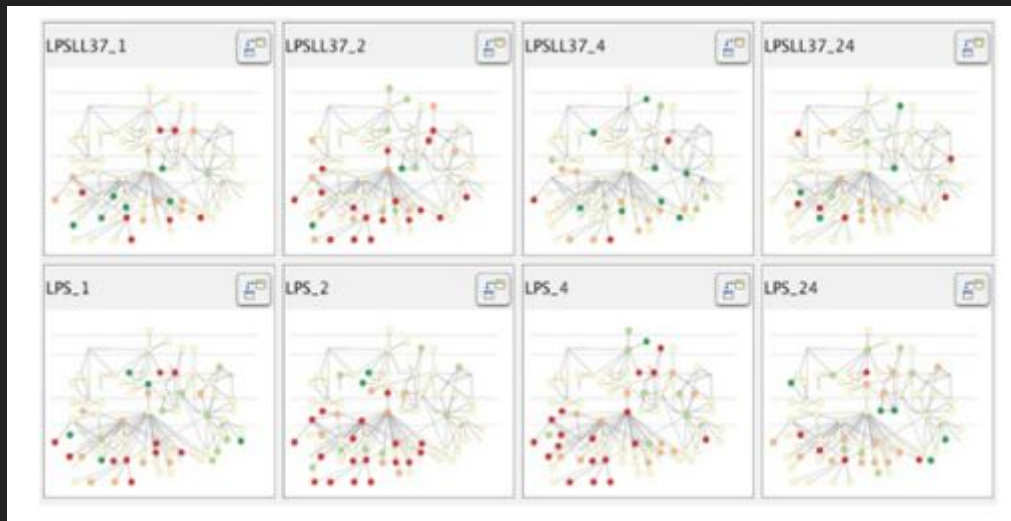
- Desde el punto de vista del usuario, la latencia de la interacción es el tiempo entre una acción y algún tipo de *feedback* del sistema, indicando que la operación fue completada.
- El tipo de *feedback* más natural es *ver un cambio en el sistema* en sí, en vez de un *approach* de imprimir notificaciones en consola o algún *popup dialog* que interfiere con el flujo de la exploración.
- Por ejemplo, hacer un *highlight* de un elemento seleccionado es una buena forma de confirmar que la operación fue exitosa.
- Además, si es que una operación está tomando *más tiempo* de lo que el usuario esperaría, una barra de progreso *debería* ser mostrada al usuario.

Responsiveness is required (costos)

- La interacción ofrece mucho **poder**, pero también tiene **costos**.
- Los beneficios se basan en que los usuarios pueden explorar un **espacio de información mucho mayor** de los que podrían ser entendidos con una imagen estática. Sin embargo, esto tiene un costo de **tiempo humano** y su **atención**.
- Detectar automáticamente los **features de interés**, para llamar la atención del usuario mediante un *visual encoding*, es un objetivo para el diseñador. Pero si la tarea pudiese ser resuelta completamente de forma automática, entonces no habría necesidad de una visualización, en primer lugar.
- Siempre habrá, entonces, un *trade-off* entre encontrar **aspectos automatizables** y dejar al humano y su cerebro para **encontrar patrones**.

Eyes beat memory

- Es más fácil usar **external cognition** que nuestra **memoria interna**.
- Por lo tanto, es más fácil comparar, moviendo nuestros ojos de lado a lado, que hacerlo tratando de recordar algo que vimos recientemente.
- Ejemplo: un gráfico con diferentes instancias, variando el color según el experimento.



Comparación de Mapas



An Evaluation of Interactive Map Comparison Techniques

Maria-Jesus Lobo, Emmanuel Pietriga and Caroline Appert.

In CHI '15: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 3573-3582, ACM, 2015. Honorable mention Award.



Técnicas comparadas

An Evaluation of Interactive Map Comparison Techniques

María-Jesús Lobo
INRIA ; INRIA Chile (CIRIC)
maria-jesus.lobo@inria.fr

Emmanuel Pietriga
INRIA ; INRIA Chile (CIRIC)
emmanuel.pietriga@inria.fr

Caroline Appert
Univ Paris-Sud ; CNRS ; INRIA
caroline.appert@lri.fr

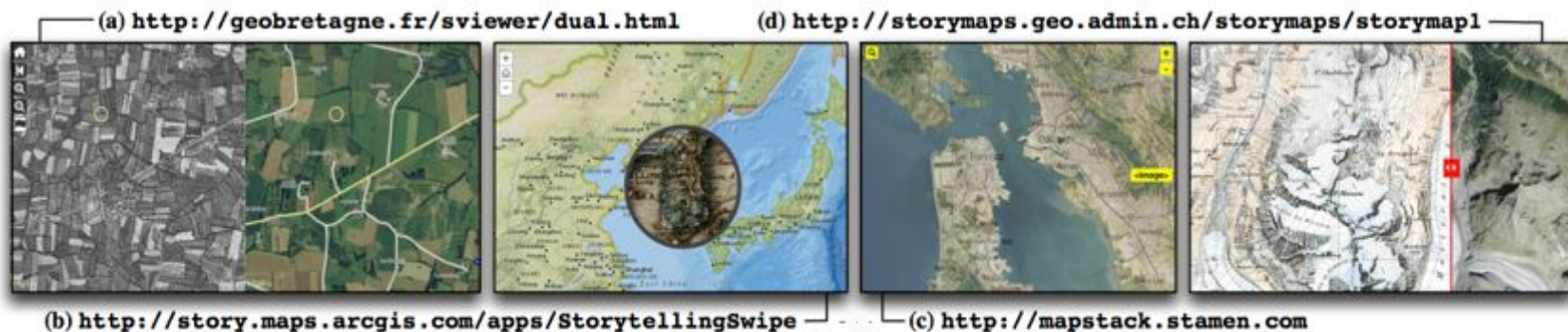


Figure 1. The four main composition strategies: (a) synchronized juxtaposed views, (b) magic lens, (c) translucent overlay, (d) swipe.

Fundamentación de las comparaciones

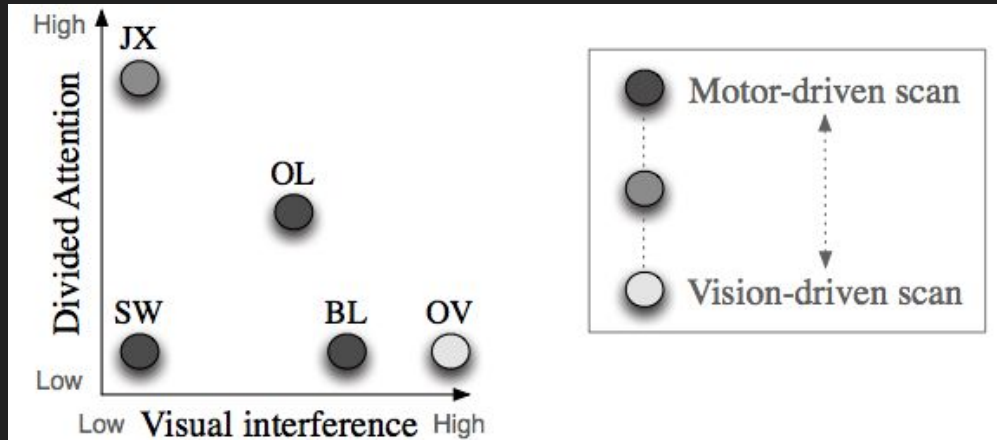
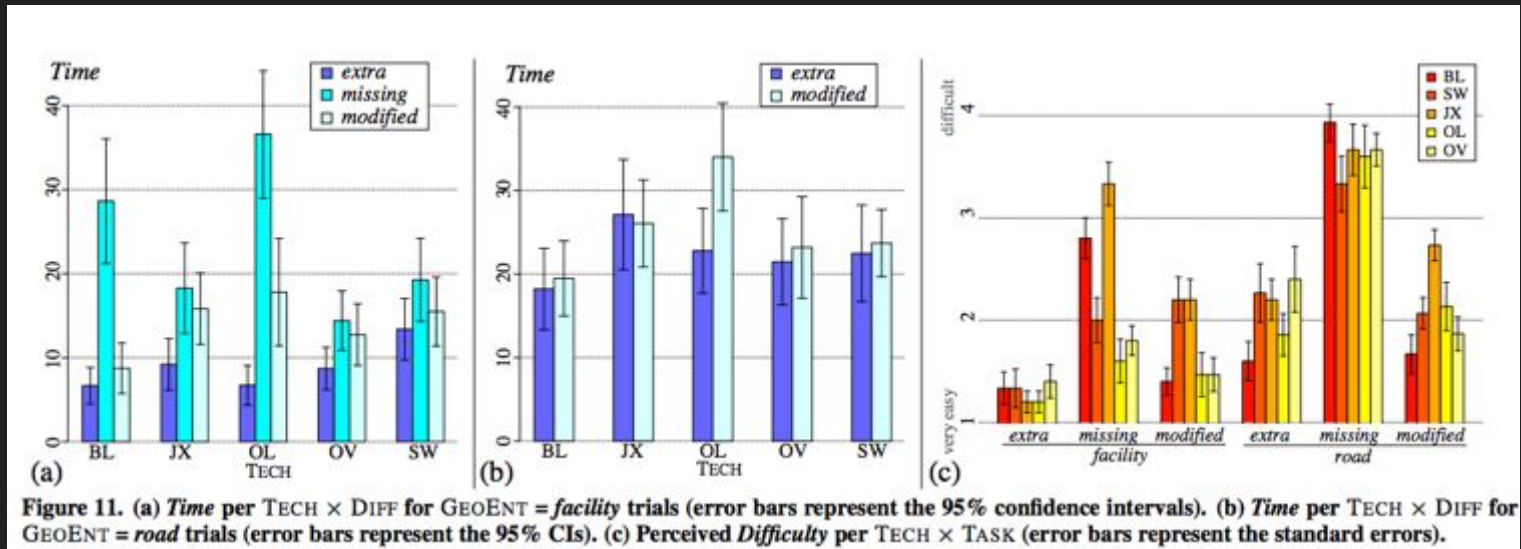


Figure 5. Coarse characterization of the techniques evaluated in the following study, in terms of visual interference and divided attention (which should both be minimized) and type of scanning. A motor-driven scanning strategy requires users to reposition elements on screen with their pointing device to compare different regions. A vision-driven strategy relies more on visual search and does not require so much interaction. This is a continuum: none of the techniques is either purely vision-driven or purely motor-driven.

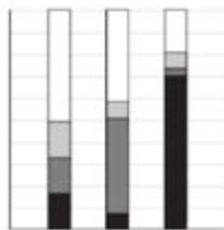
Resultados



Primero el fondo, luego la forma

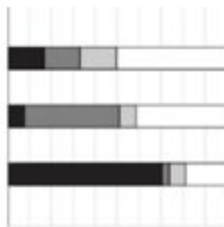
- Las mejores visualizaciones deben destacar tanto en **funcionalidad** como en **forma**: deben ser **efectivas** y **agradables** al ojo humano.
- Sin embargo, es mejor enfocarse, primero, en conseguir un diseño **efectivo** y quizá tosco, porque es posible refinarlo en forma más tarde, mientras se mantiene la efectividad.
- Al contrario, dado un diseño bello pero inefectivo, probablemente se tendrá que **rehacer desde cero**.

Consistencia



□ A
■ B
■ C
■ D

■ A
■ B
■ C
□ D



■ A
■ B
■ C
□ D

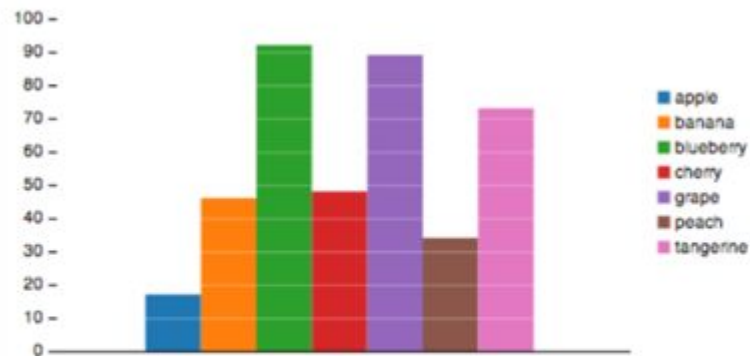
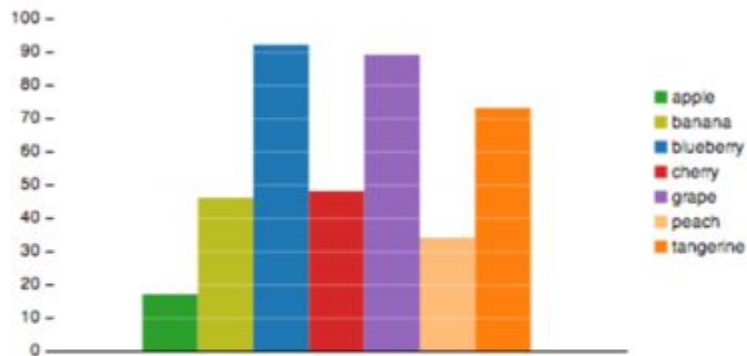
□ A
■ B
■ C
■ D



○ A
○ B
○ C
● D

□ A
■ B
■ C
■ D

Consistencia



Selecting Semantically-Resonant Colors for Data Visualization — S. Lin et al.

Más *guidelines*

- Las unidades deben ser **estandarizadas** (por ejemplo, el dinero)
- Las dimensiones del gráfico **no deberían exceder** la de los datos
- Los datos deben ser mostrados en su **contexto**
- Ojo con el **daltonismo**
- Ojo con la **tipografía**
- Gráficos **autoexplicativos** (mensajes claros, sin muchas abreviaciones)

Referencias

- “The Visual Display of Quantitative Information” — Edward Tufte
- “VAD” — Tamara Munzner
- [WTFViz](#)
- [Calling Bullshit: in the age of big data](#)