

Introducción a la Visualización

Dr. Luis Miguel de la Cruz Salas



¿Por qué visualizar?

¿Por qué visualizar?

- No importa la audiencia, tampoco importa la historia que quieres contar, nada soporta mejor tus argumentos que la evidencia en forma de datos duros.
 - **Números fríos y calculadores.**
- En un reporte anual de una compañía o en un artículo de investigación, los datos son la herramienta más poderosa para comunicar un mensaje.
 - **Aunque a veces se pasan por alto.**

¿Por qué visualizar?

- Pero esos números solitos, no necesariamente hacen el impacto deseado cuando la audiencia no es capaz de interpretarlos.

Proc.	Time [s]
1	29278
16	178
25	78
64	16
256	2
400	1



Objetivos de la visualización

- **Análisis de los datos.**
 - Entender los datos.
 - Derivar más información de los datos originales.
 - Implica comprender la información.
- **Comunicar los hallazgos.**
 - Informar.
 - Implica simplificación.

¿Por qué visualizar?

Proc.	Time [s]	Speedup	Efficiency	Cost
1	29278			
16	178			
25	78			
64	16			
256	2			
400	1			

Aceleración: $s_p = t_1/t_p$

Eficiencia: $e_a = s_p/p$

Costo: $c_p = p * t_p$

Costo óptimo: $c_p = t_1$

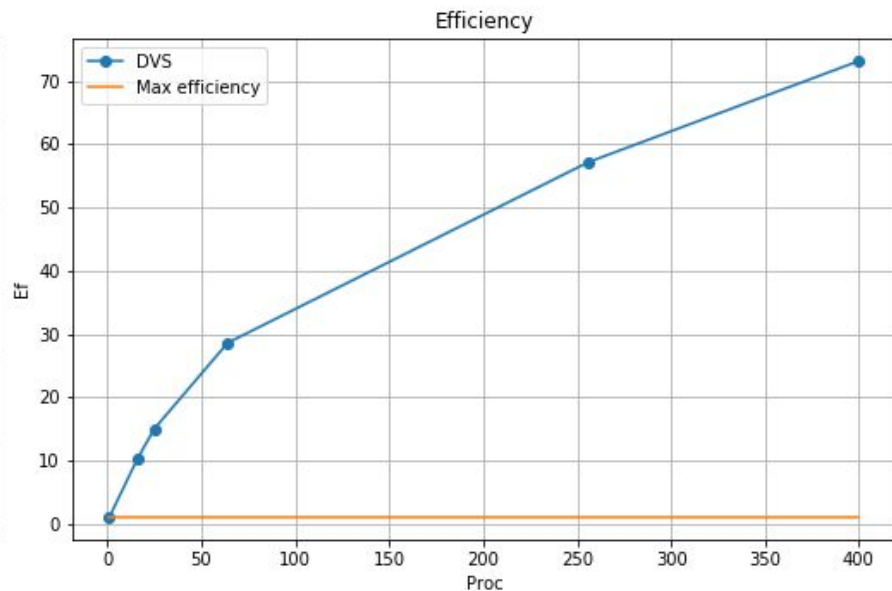
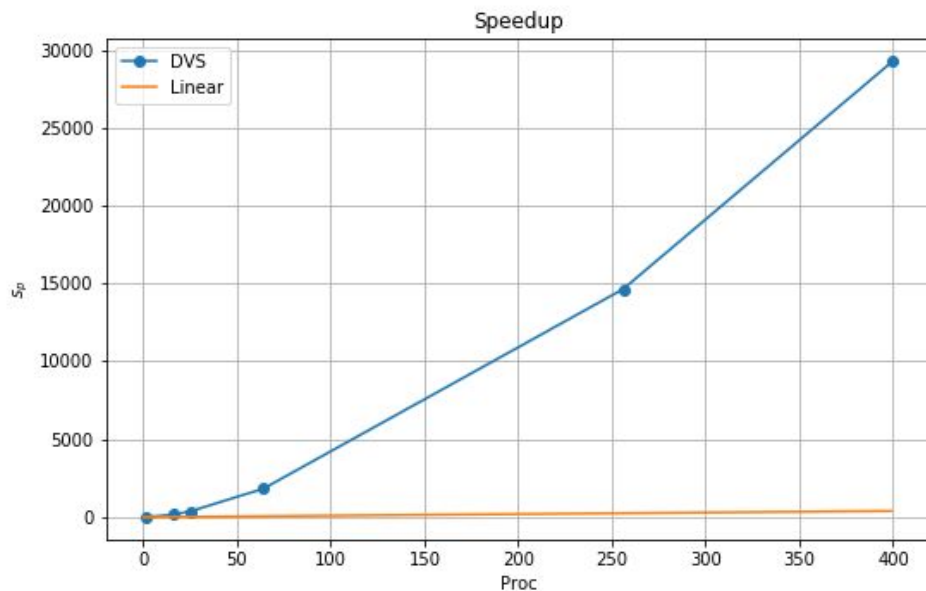
Aceleración relativa: $s_{p'} = t_p/t_{p'}$

Eficiencia relativa: $e_{p'} = p's_{p'}/p$

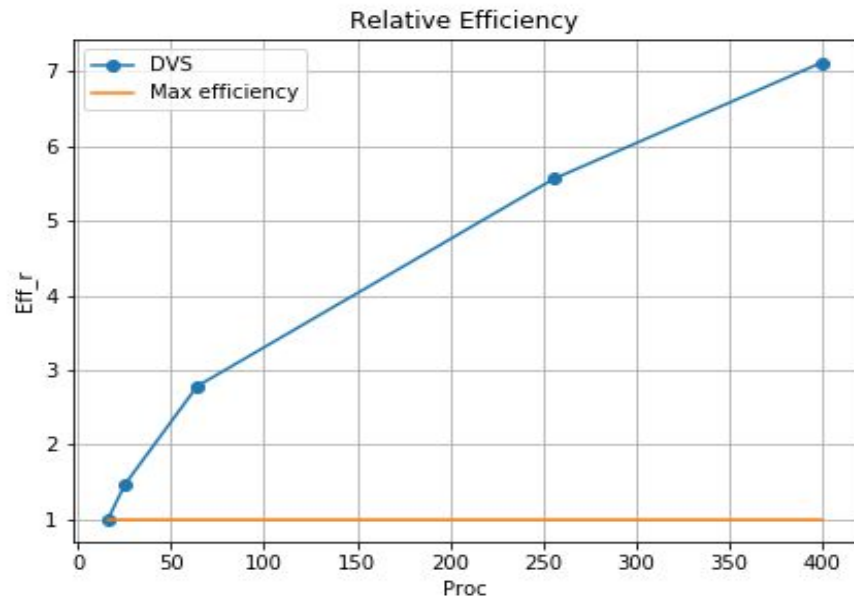
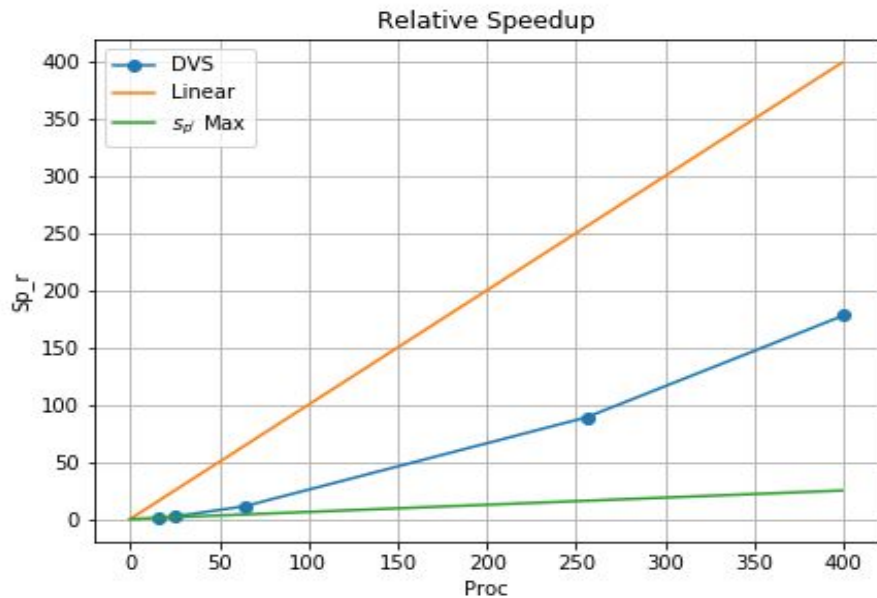
Máx. acel. rel.: $s_{p'} = p/p'$

$p' = 16$

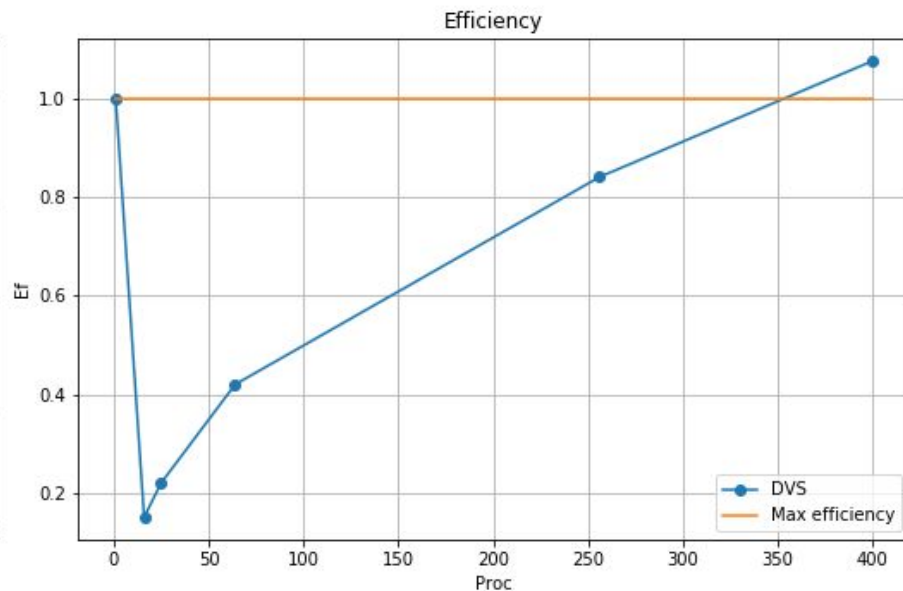
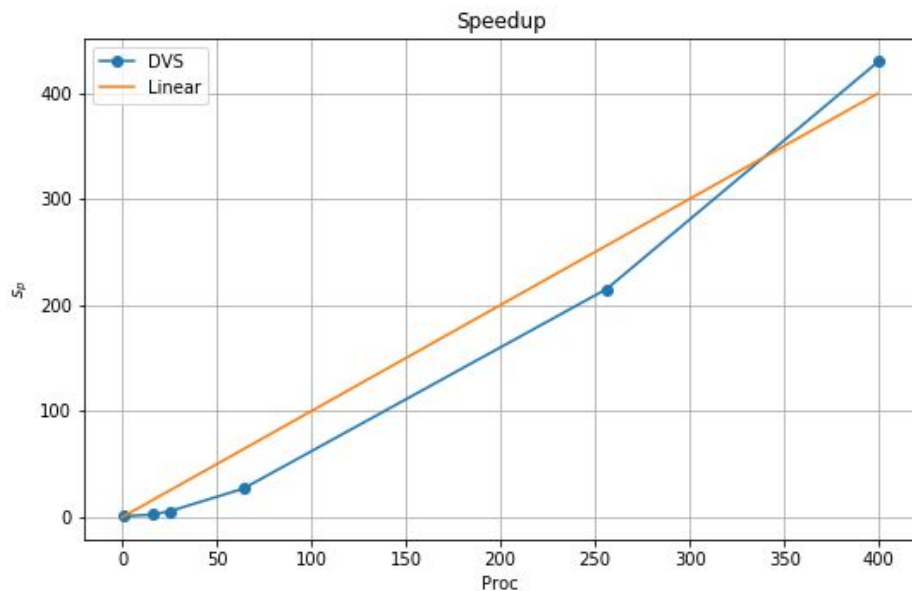
Análisis absoluto



Análisis relativo



Análisis sin valores atípicos.



Time = 430 [s], con 1 procesador.

¿Por qué visualizar?

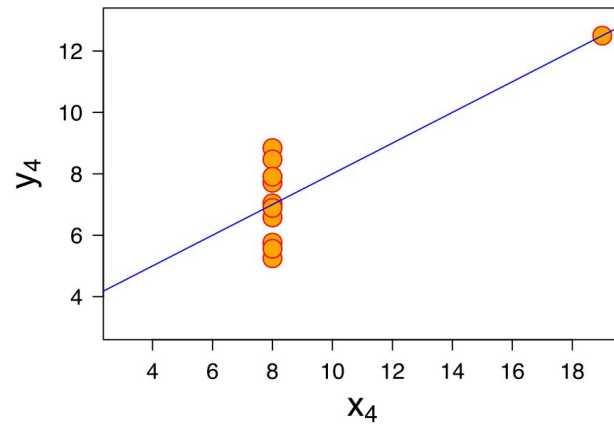
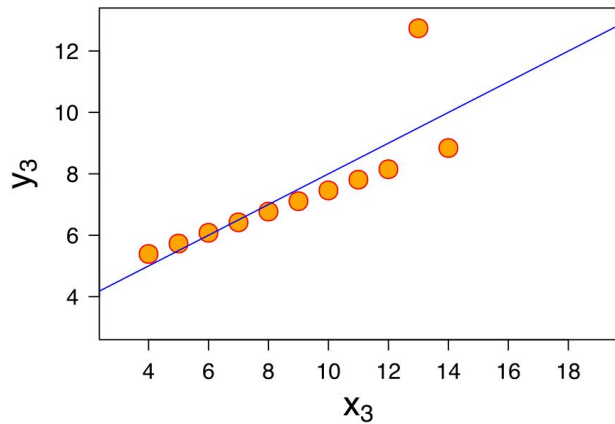
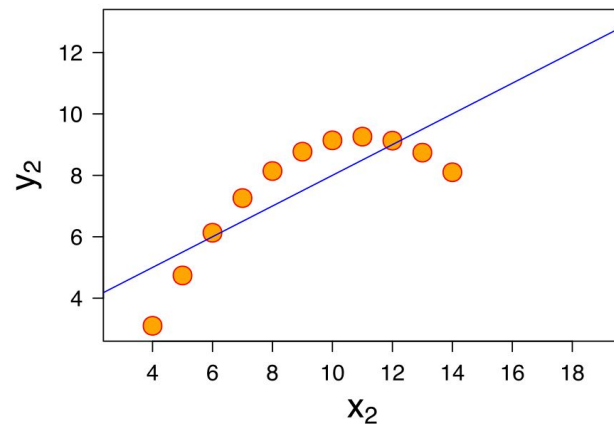
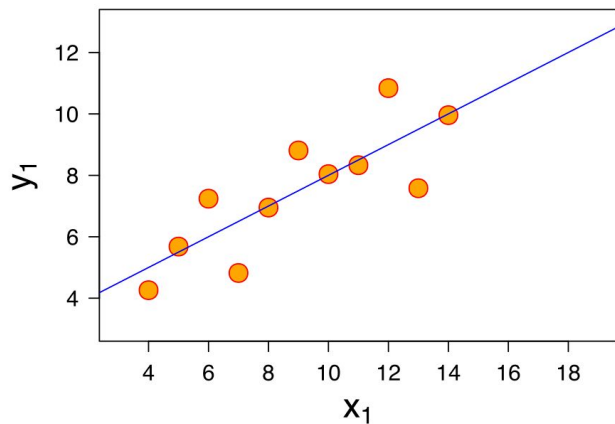
- Ejemplo: Anscombe's quartet¹
 - Consiste de cuatro conjuntos de datos que tienen casi las mismas propiedades estadísticas, pero tienen una distribución diferente, **que solo se ve cuando se grafican**.

¹ Francis Anscombe, (1973). "Graphs in Statistical Analysis". [*American Statistician*](#). 27 (1): 17–21. [doi:10.1080/00031305.1973.10478966](#). JSTOR [2682899](#).

	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	x	y	x	y	x	y	x	y
1	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
2	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76

Property	Value	Accuracy
Mean of x	9	exact
Sample variance of x : σ^2	11	exact
Mean of y	7.50	to 2 decimal places
Sample variance of y : σ^2	4.125	± 0.003
Correlation between x and y	0.816	to 3 decimal places
Linear regression line	$y = 3.00 + 0.500x$	to 2 and 3 decimal places, respectively
Coefficient of determination of the linear regression : R^2	0.67	to 2 decimal places

10	7.0	7.02	7.0	7.20	7.0	6.72	8.0	7.91
11	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89



¿Por qué visualizar?

- Anscombe demostró la importancia de graficar los datos antes de analizarlos y el efecto de los valores atípicos.
 - La intención fue cambiar la impresión de que “los cálculos numéricos son exactos, pero los gráficos aproximados”.

¿Por qué visualizar?

- El núcleo de una buena investigación se encuentra en el tratamiento cuidadoso y respetuoso de los datos.
 - Calibrar los instrumentos, planear las encuestas, limpiar la información, revisar cuidadosamente los algoritmos de procesamiento, estudiar el comportamiento de los modelos matemáticos, etc.
- Sin embargo, este tratamiento cuidadoso no siempre se hace cuando se realiza una visualización.

Beneficios de la visualización

1. Agrega credibilidad a la historia.
2. Fomenta la interacción con la audiencia
3. Incrementa el impacto.
 - Cuando se proporcionan los datos para apoyar los hallazgos, se demuestra que se hace investigación, eso da credibilidad y por lo tanto la audiencia confía.
 - Y si se conjuntan con visualizaciones bien diseñadas, claras, limpias y convincentes, entonces la audiencia interactúa y eso lleva a más y mejores descubrimientos.

¿Cuándo usar la visualización?

- Para responder preguntas.
- Comunicación rápida y efectiva.
 - Internamente en un grupo de trabajo.
 - Hacia un cliente o colega de otra institución.
- Agregar objetividad a un argumento.
 - Definir el rumbo de un negocio o una investigación.
- Hacer nuevos descubrimientos.



Moraleja

- Los datos deben visualizarse para realmente alcanzar el objetivo:
 - Impactar a la audiencia a la que nos dirigimos.
- La presentación de los datos debe hacerse de manera visual y atractiva.
 - Mejora la historia y la experiencia de la audiencia.

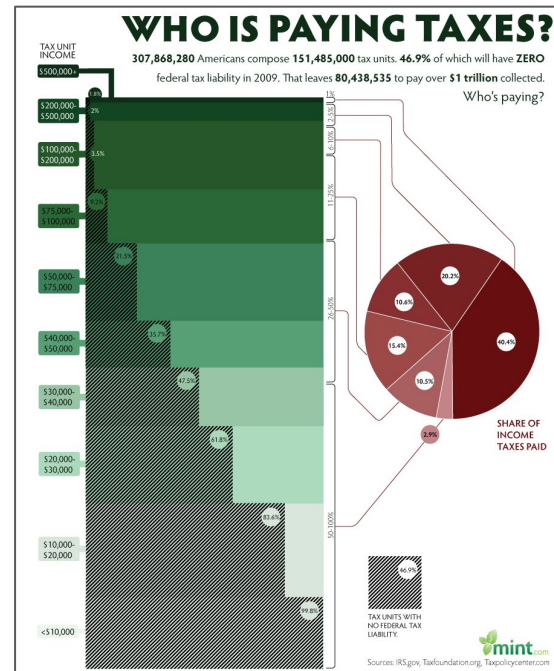
La visualización es la clave.

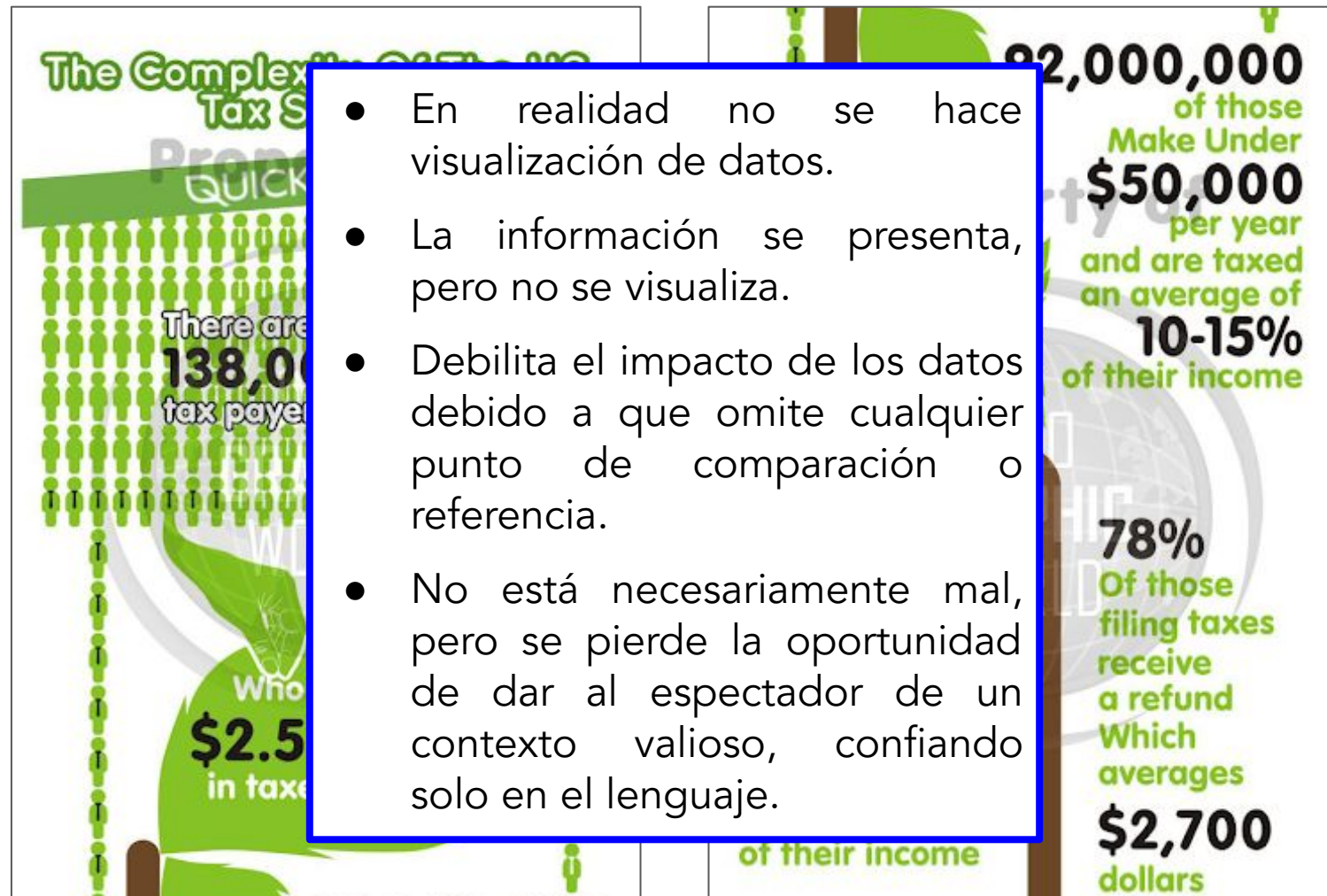
Encontrar la mejor técnica de visualización suele ser un arte.

¿Cuál de estas dos infografías, hace que los datos se entiendan mejor?

¿Cuál provee más contexto?

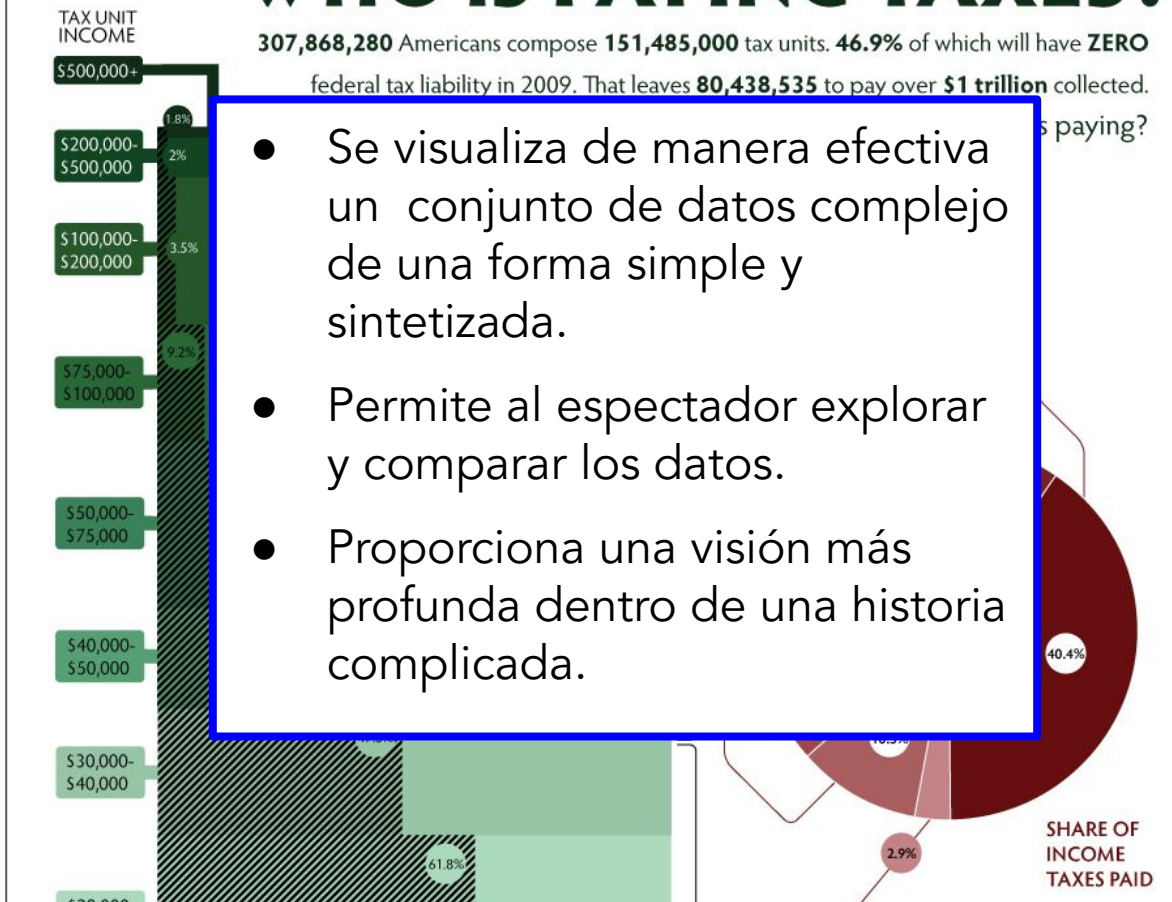
¿Cuál es más creíble?







WHO IS PAYING TAXES?



De los datos a las imágenes

Entender, explicar y aplicar lo siguiente:

- Propiedades de los datos.
 - Tipos de datos, taxonomía de los datos,
- Propiedades de las imágenes.
 - Teoría y semiología de las imágenes y gráficos, teoría del color, percepción humana, efectos visuales, ...
- Reglas de mapeo de los datos a las imágenes.
 - Esquemas de codificación, rasgos sobresalientes, ...

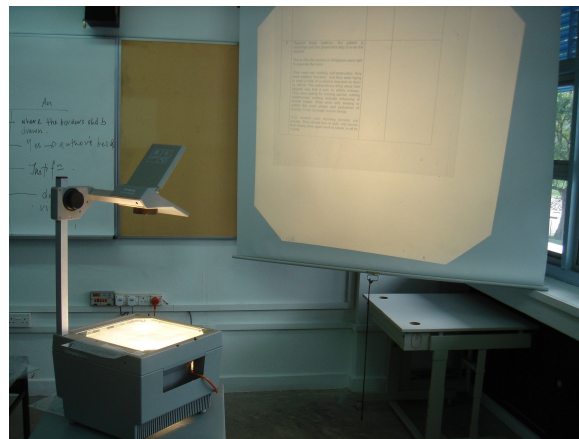
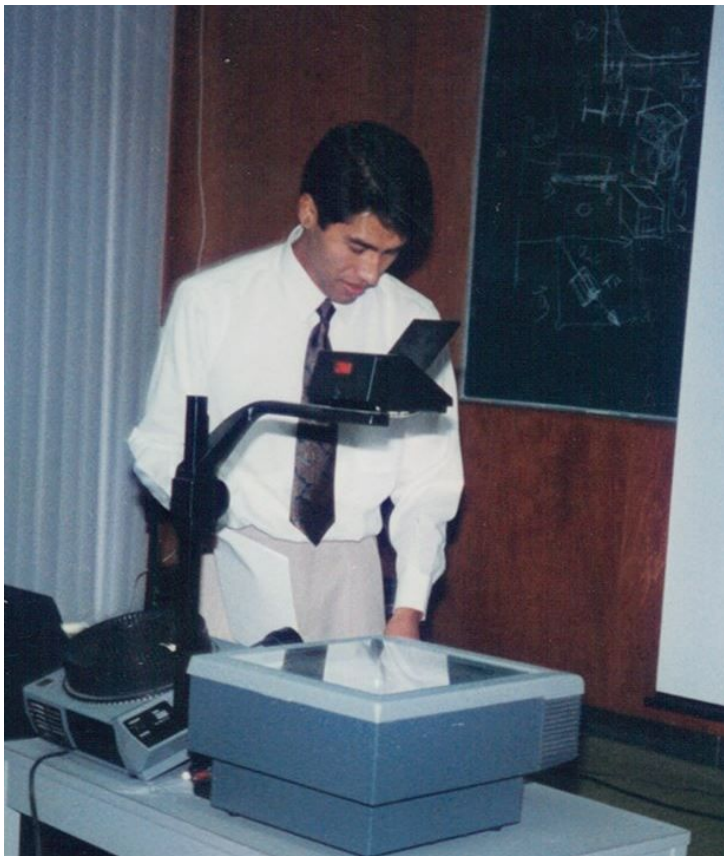
Tipos de despliegues e interfaces

Tipos de pantallas

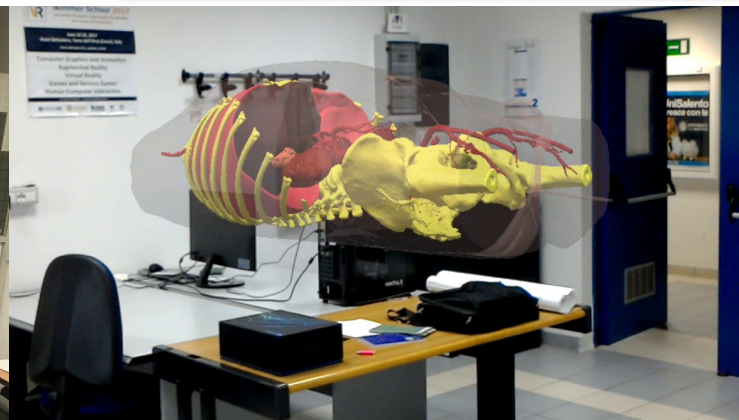
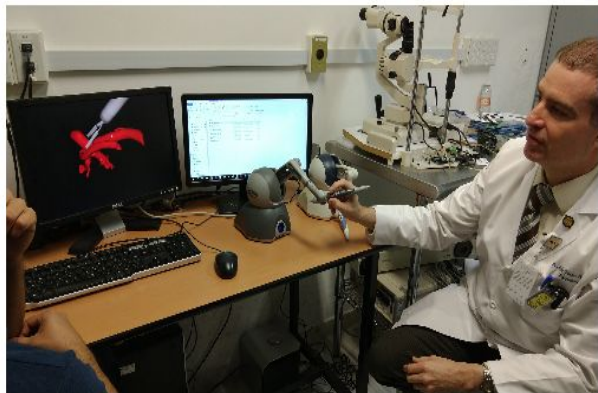
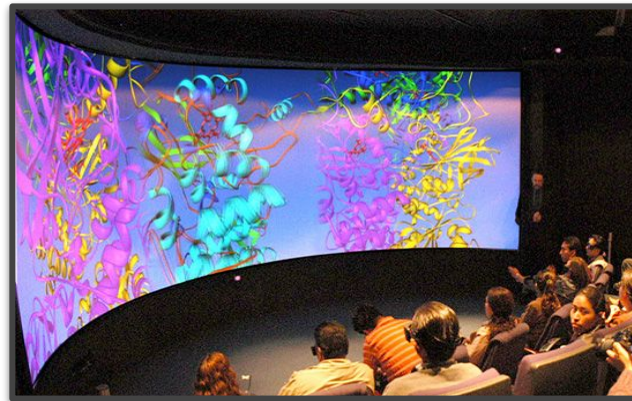
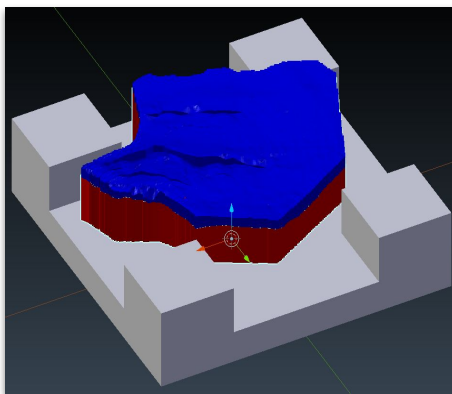
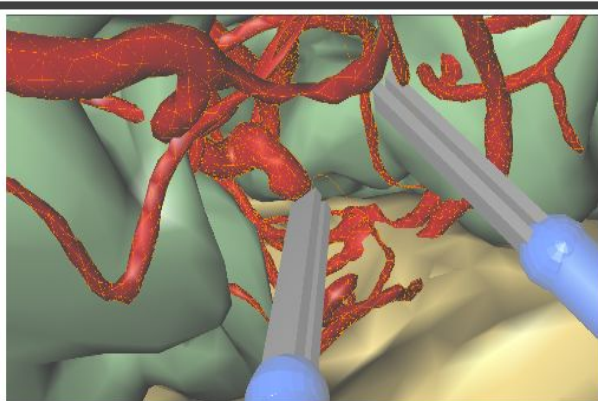
- CRT
- VGA
- SVGA
- PLASMA
- LCD
- LED
- 4K ULTRA HD
- TOUCH SCREEN
- MULTITOUCH
- AMOLED
- GORILLA GLASS



Prehistoria



Realidad virtual, aumentada, mixta





Identificando historias en los datos





Hacer y responder preguntas como las siguientes

- ¿Qué preguntas se desean responder?
 - ¿En dónde debo invertir mis ahorros?
- ¿Qué datos se necesitan para responder a las preguntas?
 - Rendimientos de los últimos años de las casas de bolsa
- ¿Dónde podemos encontrar esos datos?
 - Reportes financieros, banco de México, ...
- ¿Es posible generar información complementaria?
 - Correlacionar información de varias fuentes.



Abastecerse de información

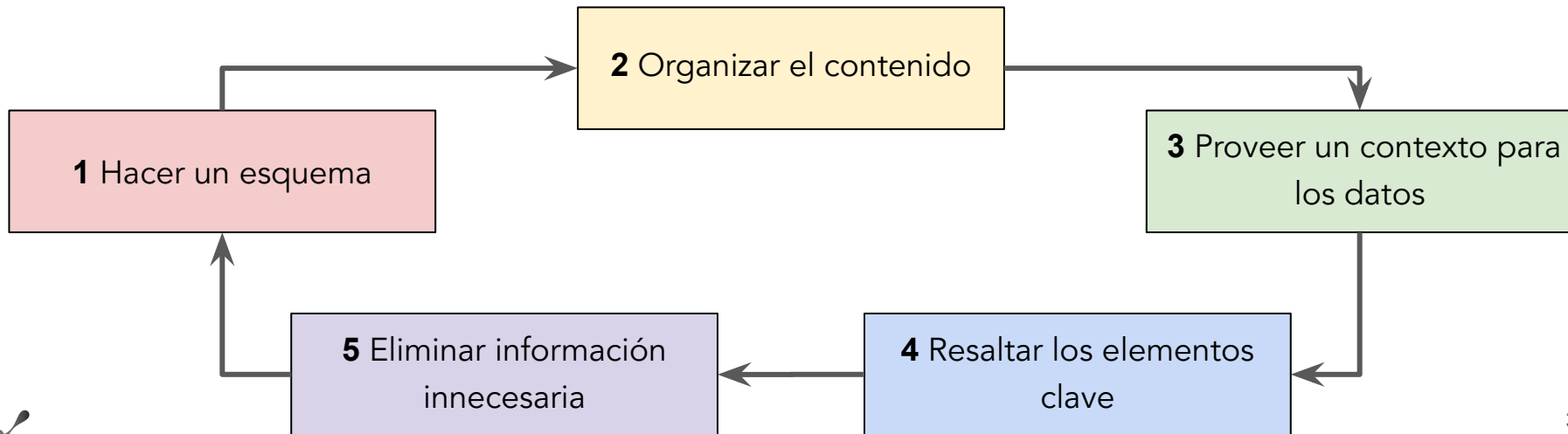
- Recolectar conjuntos de datos que proporcionen un marco sólido para la investigación y que permitan responder las preguntas.
 - En la actualidad este paso puede ser abrumador porque existen muchas fuentes de información.
 - Considere solamente aquellas fuentes de información que provean de un marco sólido, cohesivo y consistente.
- Algunas fuentes:
 - INEGI, Repositorios Conacyt, Google Scholar, US Census Bureau, Data.gov, Google Public Data Explorer, SocialMention.

Análisis de la información

- La comunicación más poderosa sucede cuando los datos realmente muestran una historia.
 - Cuando se hace el análisis de los datos, ciertos elementos van a resaltar.
 - Estos elementos definen el marco de la historia. Por ejemplo:
 - Tendencias
 - Correlaciones
 - Valores atípicos

Elaborar la historia

- Definir cuánta y qué información se debe incluir para que la historia se cuente correctamente.
 - **Objetivo: motivar e impactar a la audiencia.**



How to Convince Someone When Facts Fail :Why worldview threats undermine evidence



Credit: Izhar Cohen