

Visualización de Información

**IIC1005 - Computación: Ciencia y
Tecnología del Mundo Digital**

Fernando Florenzano Hernández
faflorenzano@ing.puc.cl

Contenidos

¿Qué es?

¿Por qué?

¿Cómo?

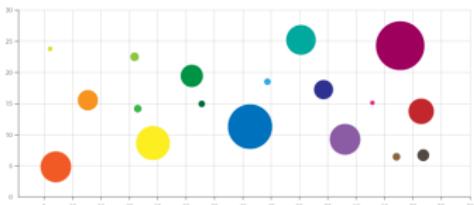
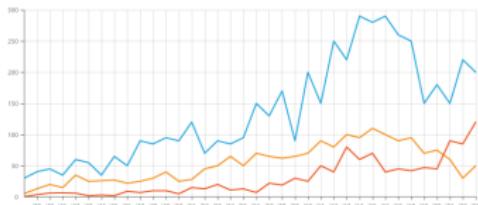
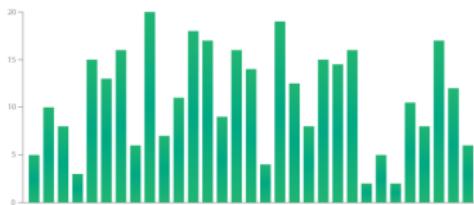
Contenidos

¿Qué es?

¿Por qué?

¿Cómo?

¿Qué es la visualización?



Fuente: [The Data Visualisation Catalogue](#)

¿Qué es la visualización?

Según el **Diccionario de la lengua española**, visualizar es:

- Visibilizar.
- **Representar** mediante **imágenes ópticas** fenómenos de **otro carácter**. ej., el curso de la fiebre o los cambios de condiciones meteorológicas mediante gráficas, los cambios de corriente eléctrica o las oscilaciones sonoras con el oscilógrafo, etc.
- **Formar** en la mente una **imagen visual** de un concepto abstracto.
- **Imaginar** con rasgos **visibles** algo que **no se tiene a la vista**.

¿Qué es la visualización?

- “Transformación de lo **simbólico** a lo **geométrico**”. [McCormick et al. 1987]
- “[...] encontrar la **memoria artificial** que mejor apoya a nuestros medios naturales de percepción.” [Bertin 1967]
- “El uso de representaciones visuales de datos, generados por computador, interactivos, para **amplificar nuestra cognición**.” [Card, Mackinlay, & Shneiderman 1999]

¿Qué es la visualización de información?

Según Munzner, 2014:

- “Computer-based visualization systems provide **visual representations of datasets** designed to help people carry out tasks more **effectively**.”
- “Visualization is suitable when there is a need to **augment human capabilities** rather than replace people with computational decision-making methods.”

¿Qué es IIC2026?

- Es una introducción al aprendizaje de los **principios de diseño gráfico** y de técnicas interactivas para **visualizar datos**.
- Está diseñado para entregar las herramientas necesarias para entender el estado del arte en **Visualización de Información**.
- Es un curso multidisciplinario que incorpora subcampos de la **computación**, la **estadística**, del **diseño gráfico** (e.g. teoría de color, tipografía), y de **psicología cognitiva**.
- Busca explicar cómo las **representaciones visuales** son una ayuda en el análisis y entendimiento de *datasets* altamente complejos, y cómo, además, diseñar e implementar visualizaciones efectivas usando modernas librerías *web-based*.

Contenidos

¿Qué es?

¿Por qué?

¿Cómo?

¿Por qué?

Como humanos, hemos producido **muchos** datos. ¿Cuántos?

Se estima, que para el 2018 se ha producido **33 zetabytes** de datos.

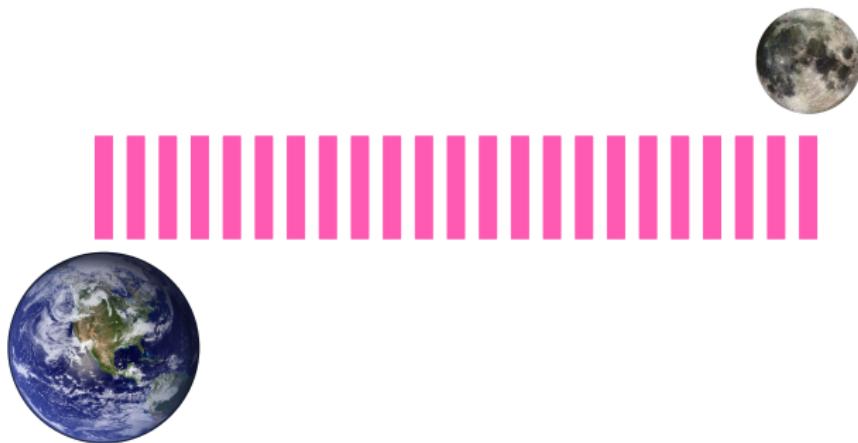
$1 \text{ zetabyte} = 10^{21} \text{ bytes} = 10^9 \text{ terabytes}$

Y se estima que para el 2025, se alcanzarán los **175 zetabytes**.

Fuente: IDC - The Digitization of the World

Muchos, muchos datos

Si almacenáramos **175 zetabytes** en solo DVDs, entonces tendrías suficientes DVDs como para crear **23** pilas que lleguen a la **Luna**.



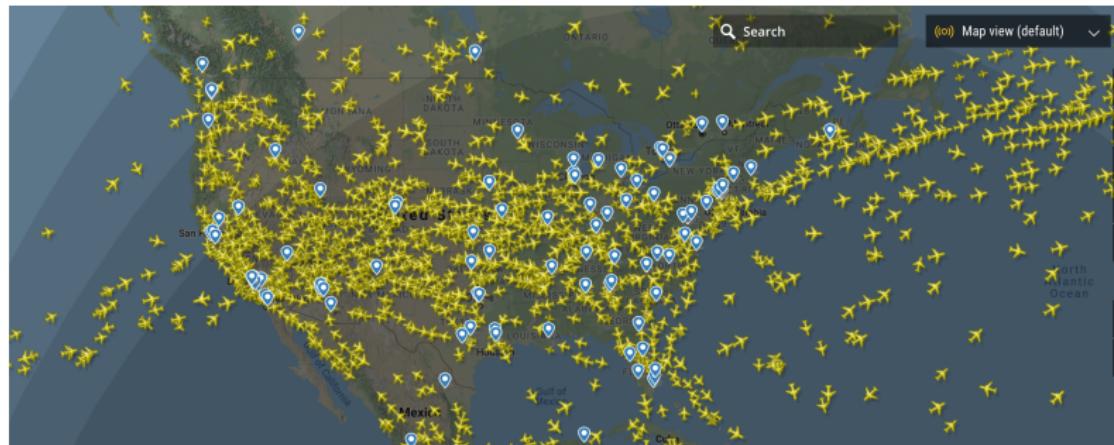
Muchos, muchos datos

Si los juntáramos todos en una sola pila, esta sería capaz de dar la vuelta a la **Tierra 222 veces**.

Si se pudiera descargar todo a una velocidad de **25 Mb/s**, una sola persona se demoraría **1.8 billones de años** en hacerlo. Si todas las personas en la tierra ayudan, entonces se demorarían **81 días**.

¿De dónde provienen tantos datos?

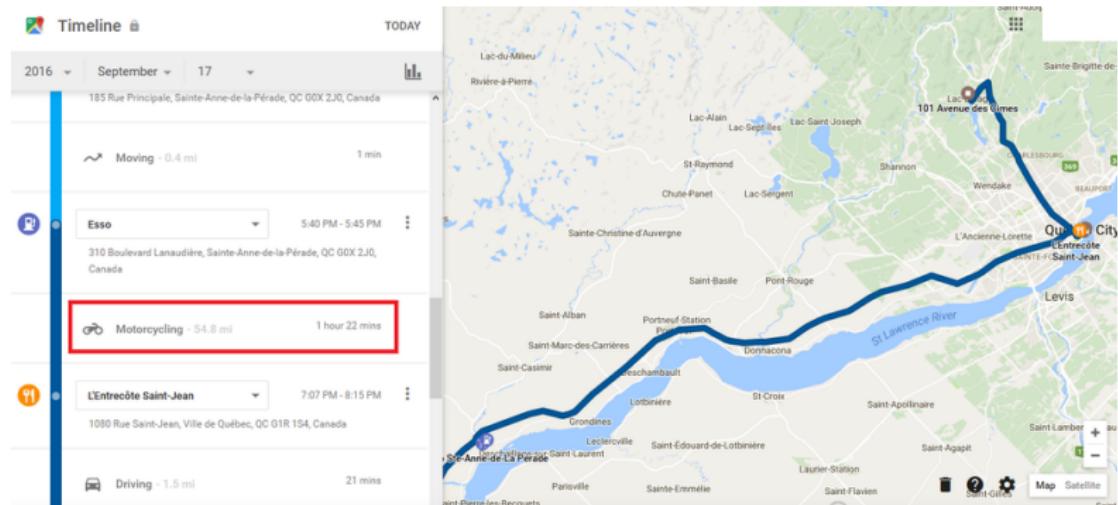
Sensores físicos: datos en **vivo** de vuelos.



Fuente: [flightradar24](#)

¿De dónde provienen tantos datos?

Nosotros producimos datos: **Google Maps Timeline**.



Fuente: [Reddit post](#)

Muchos, muchos datos

*“What information consumes is rather obvious: **it consumes the attention of its recipients.***

*Hence a wealth of information creates a **poverty of attention**, and a need to allocate that attention **efficiently** among the overabundance of information sources that might consume it.”*

Herbert Alexander Simon (1916-2001)
Premio Nobel de Economía
Premio Turing

Muchos, muchos datos

*"The ability to take data—to be able to **understand** it, to **process** it, to **extract value** from it, to **visualize** it, to **communicate** it's going to be a hugely important skill in the next decades, not only at the professional level but even at the educational level for elementary school kids, for high school kids, for college kids. Because now we really do have essentially **free and ubiquitous data**. So the complimentary scarce factor is the ability to understand that data and **extract value** from it."*

Hal Varian
Google's Chief Economist (2009)

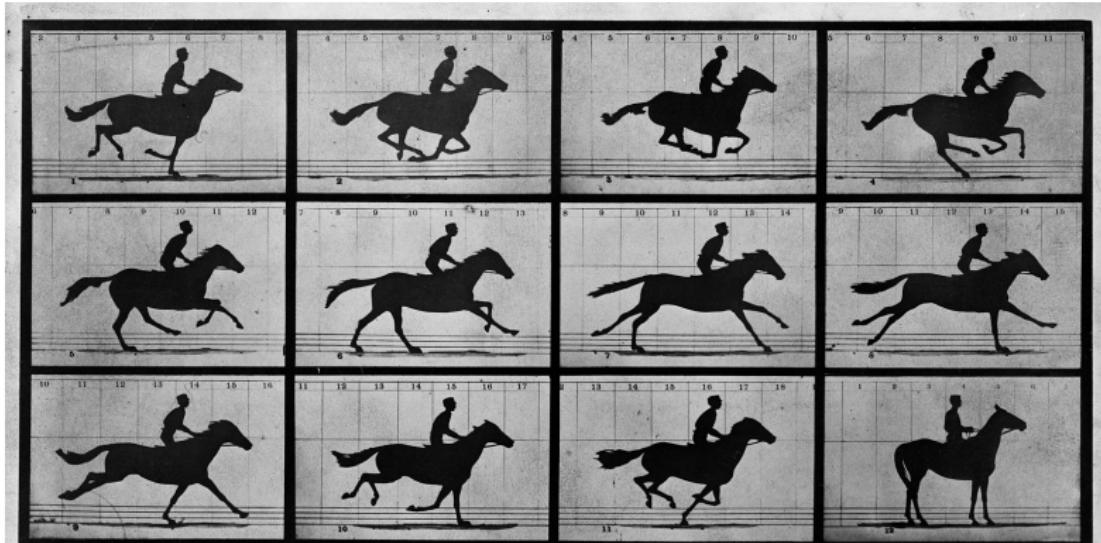
Tres funciones de las visualizaciones

- Registrar información
 - Fotografías, planos, etc...
- Apoyar razonamientos sobre la información (analizar)
 - Procesar y calcular
 - Razonar acerca de los datos
 - Retroalimentación e interacción
- Transmitir información a otras personas
 - Compartir y persuadir
 - Colaborar y revisar
 - Enfatizar aspectos importantes de los datos

Registrar información



Registrar información: Eadweard Muybridge



Copyright, 1878, by MUYBRIDGE.

MORSE'S Gallery, 417 Montgomery St., San Francisco.

THE HORSE IN MOTION.

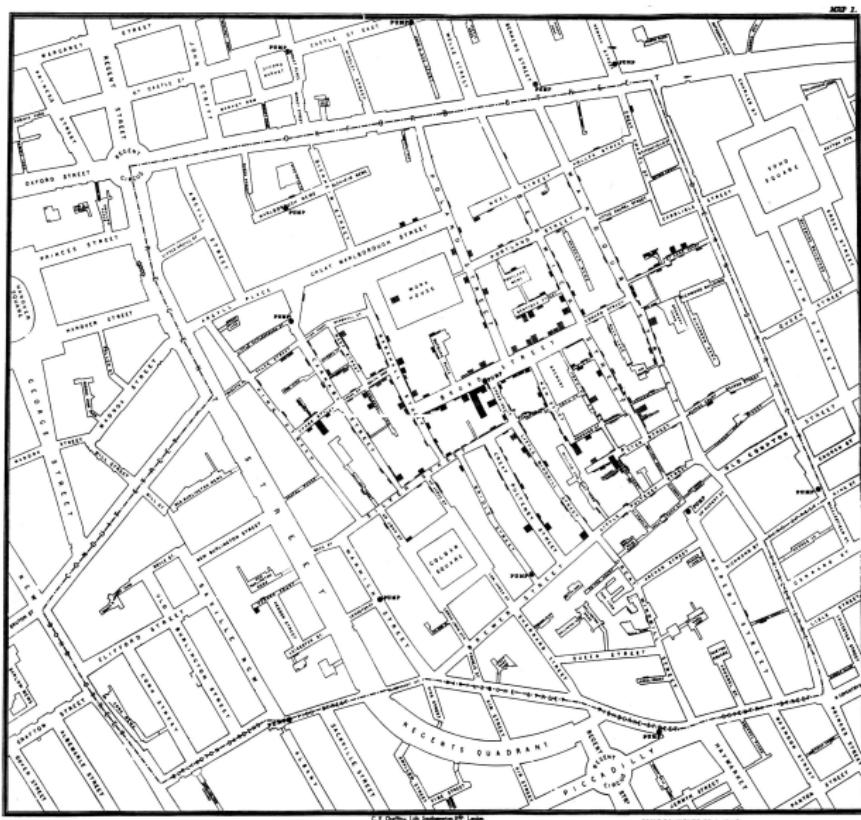
Illustrated by
MUYBRIDGE,

"SALLIE GARDNER," owned by LELAND STANFORD; running at a 1.40 gait over the Palo Alto track, 19th June, 1878.

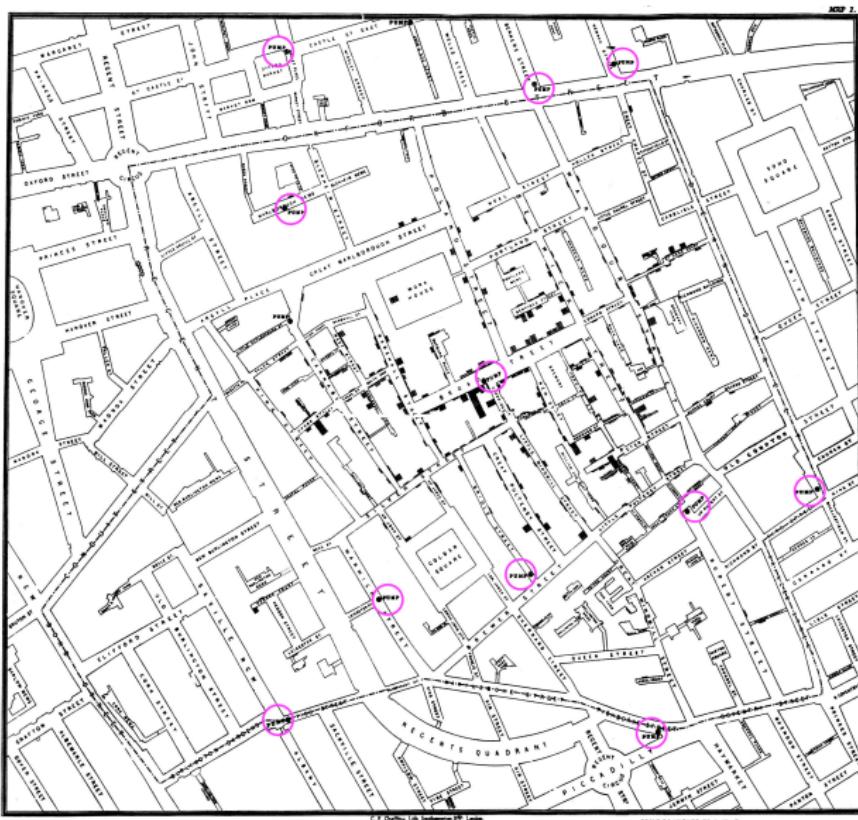
The negatives of these photographs were made at intervals of twenty-seven inches of distance, and about the twenty-fifth part of a second of time; they illustrate consecutive positions assumed in each twenty-seven inches of progress during a single stride of the mare. The vertical lines were twenty-seven inches apart; the horizontal lines represent elevations of four inches each. The exposure of each negative was less than the two-thousandth part of a second.

AUTOMATIC ELECTRO-PHOTOGRAPHIC

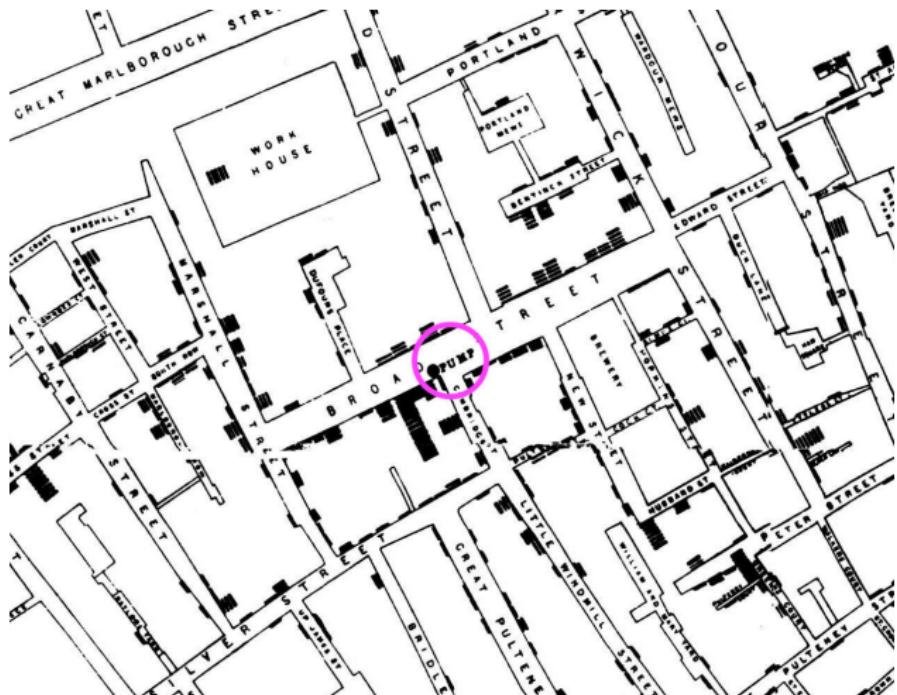
Apoyar razonamiento: John Snow



Apoyar razonamiento: John Snow



Apoyar razonamiento: John Snow

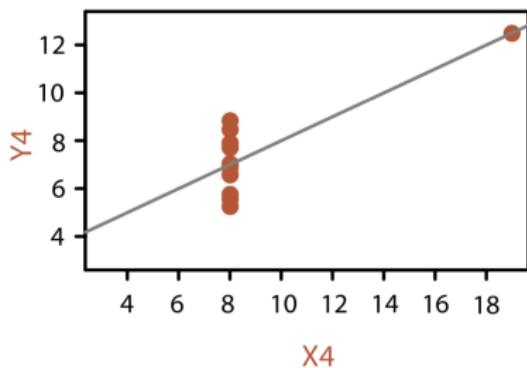
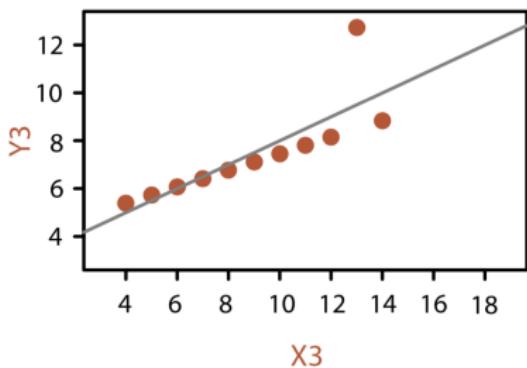
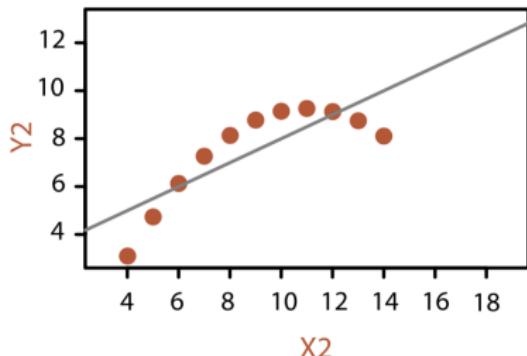
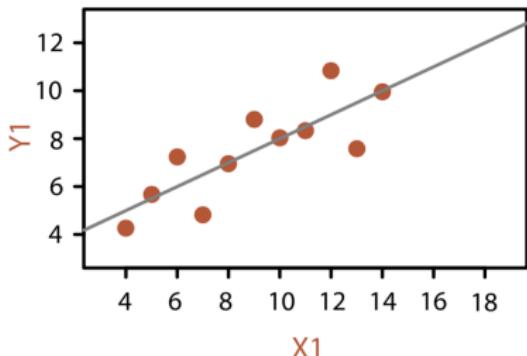


Apojar razonamiento

Anscombe's Quartet: Raw Data

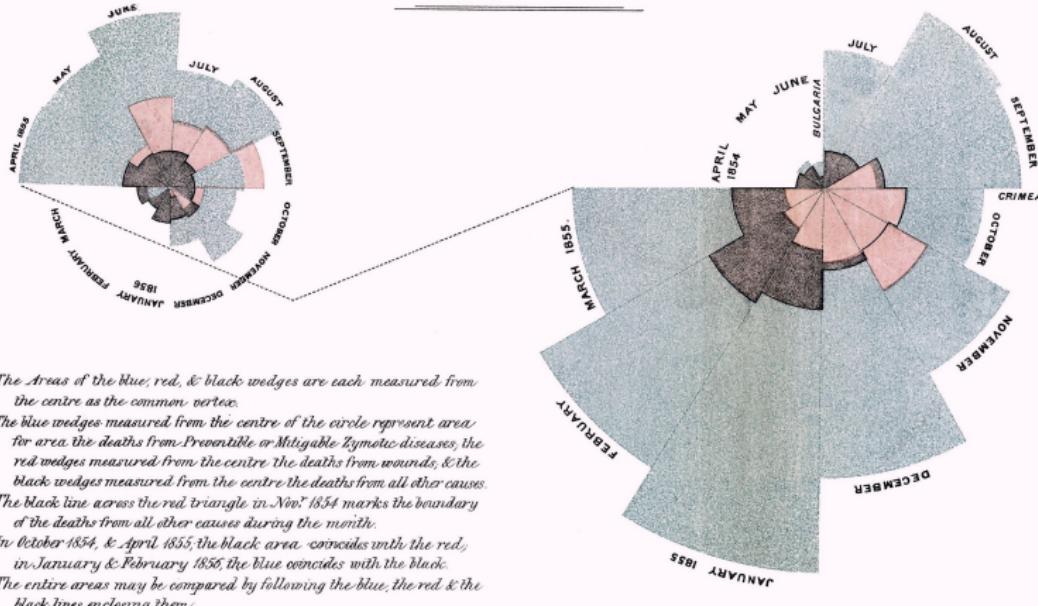
	1		2		3		4		
	X	Y	X	Y	X	Y	X	Y	
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58	
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76	
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71	
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84	
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47	
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04	
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25	
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50	
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56	
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91	
Statistics	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89	
	Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
	Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816		

Aceptar razonamiento



Transmitir información a otras personas

2. APRIL 1855 to MARCH 1856.
1. APRIL 1854 to MARCH 1855.



Contenidos

¿Qué es?

¿Por qué?

¿Cómo?

¿Cómo?

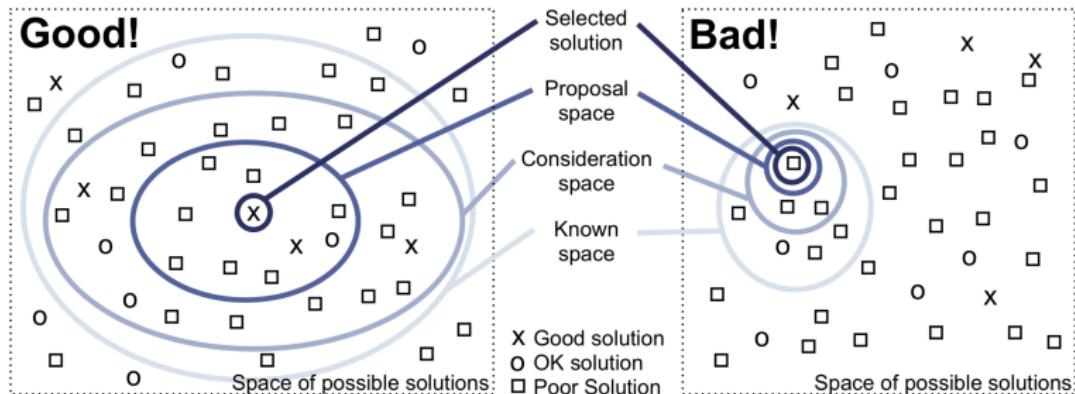
Cuando diseñamos visualizaciones, nos enfocamos en su **efectividad** de transmitir información hacia un objetivo.

A diferencia de otras disciplinas que utilizan imágenes; como las artes, películas, *marketing*; en Visualización de Información no se busca hacer algo **bonito**, si no algo **efectivo**.

¿Cómo?

La mayoría de los diseños de visualización son **inefectivos**.

El espacio de posibles soluciones es **enorme**. No hay un claro método para optimizar, pero si existen **guidelines** que uno puede seguir. Existen pocas **verdades** en esta disciplina.



¿Cómo?

Validar un diseño de visualización es un proceso sumamente **difícil**. Existen interminables preguntas sobre la validez de un diseño:

- ¿Cómo sabes si es que funciona?
- ¿Cómo argumentas que este diseño es mejor que otro para los usuarios?
- Espera, ¿qué significa mejor?
- ¿Cómo decides que tipo de tareas debe hacer un usuario al momento de probar el sistema?
- ¿Y quién es el usuario? ¿Un experto? ¿Un novato?
- ¿Cuán complejo es dibujar esta visualización?
- ¿Dependiendo de la cantidad de datos o de la cantidad de pixeles?
- ¿Existe un *trade-off* entre la velocidad del computador y el uso de memoria al computar esto?
- ...

Rules of thumb

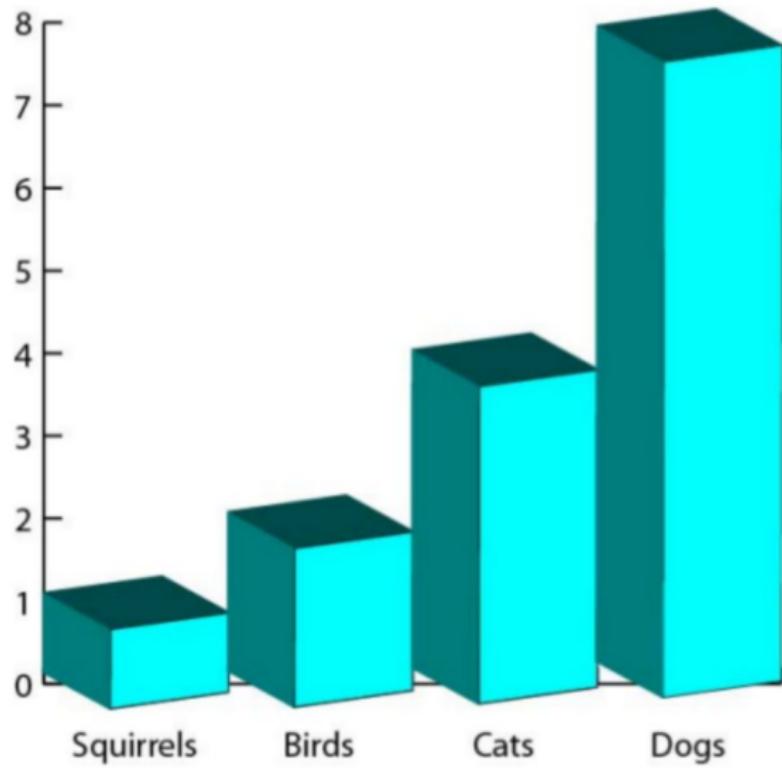
Un **rule of thumb** es un **principio** o una **guía** basado en **experiencia** y/o en **práctica** más que en teoría.

Data ink ratio

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

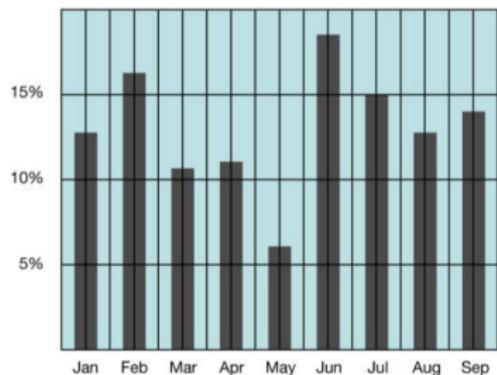
En nuestras visualizaciones, buscamos **maximizar** esta proporción. Todos los **elementos gráficos** que utilicemos deben tener una **razón de existir**. Si se llega a 1, entonces cada **pixel** está justificado.

Data ink ratio

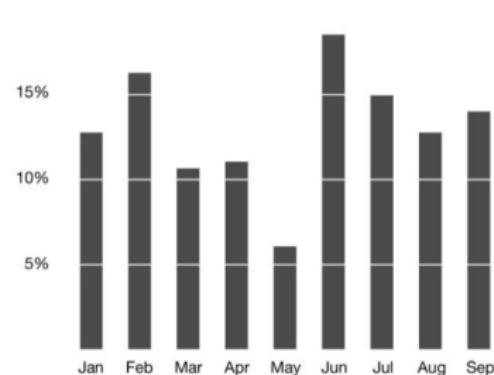


Data ink ratio

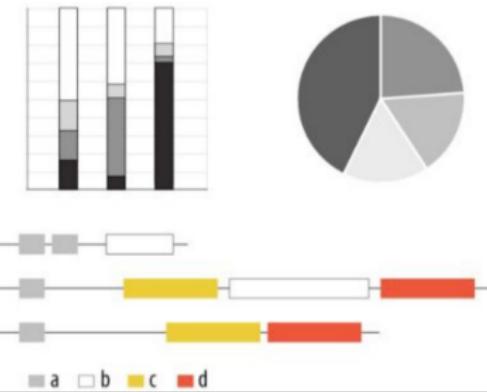
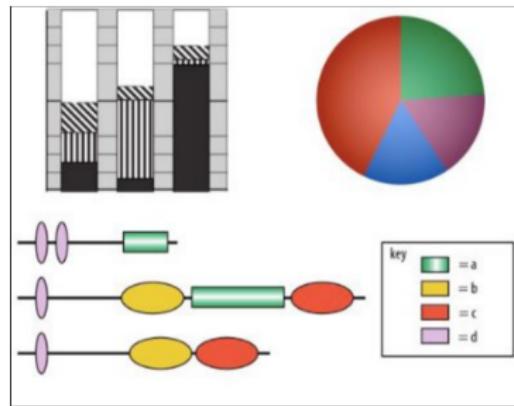
LOW INK-DATA RATIO



HIGH INK-DATA RATIO



Data ink ratio



Remove
to improve
(the **line graph** edition)

<https://www.youtube.com/watch?v=bDbJBWvonVI>

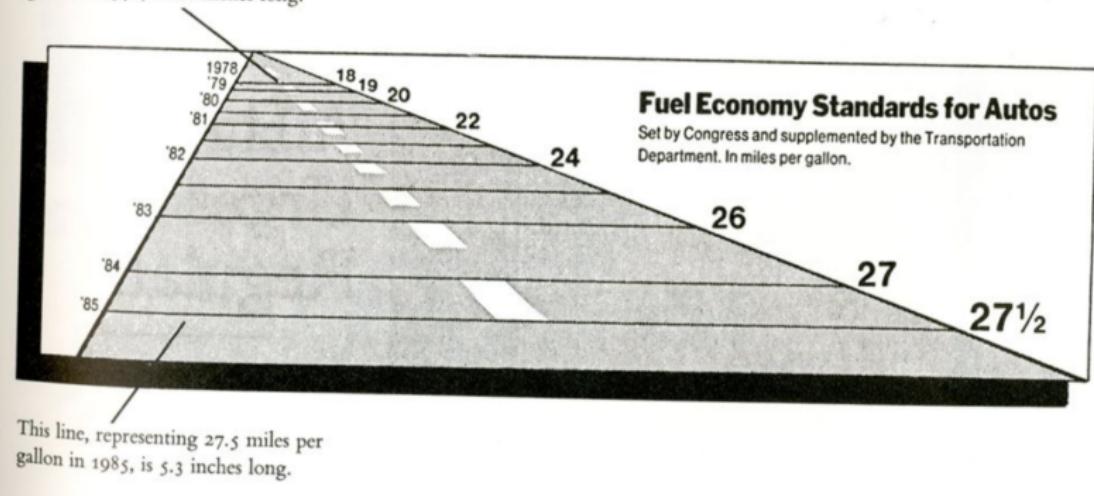
Lie factor

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

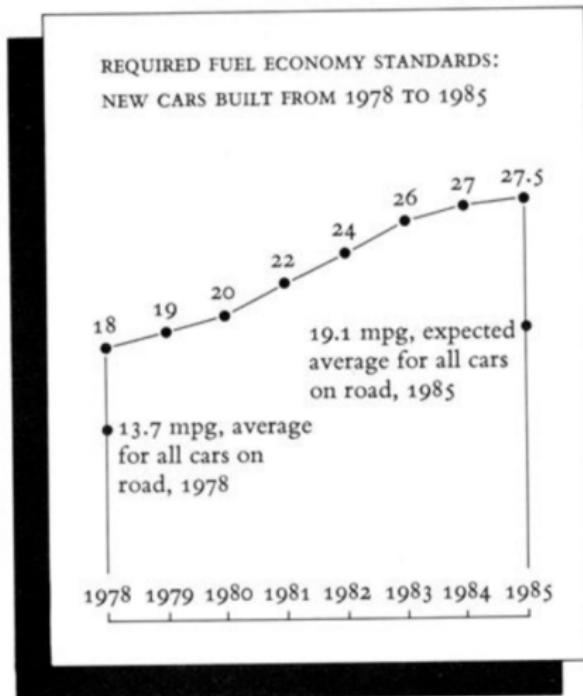
En nuestras visualizaciones, buscamos llegar que esta proporción se vuelva 1. Todos los **efectos de tamaño** que utilicemos gráficamente deben **fielmente reflejar** los efectos de los datos.

Lie factor

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



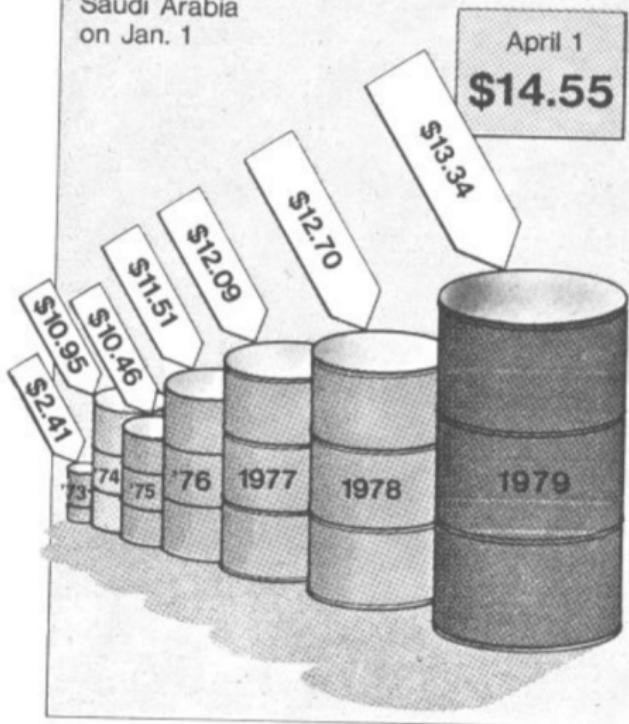
Lie factor



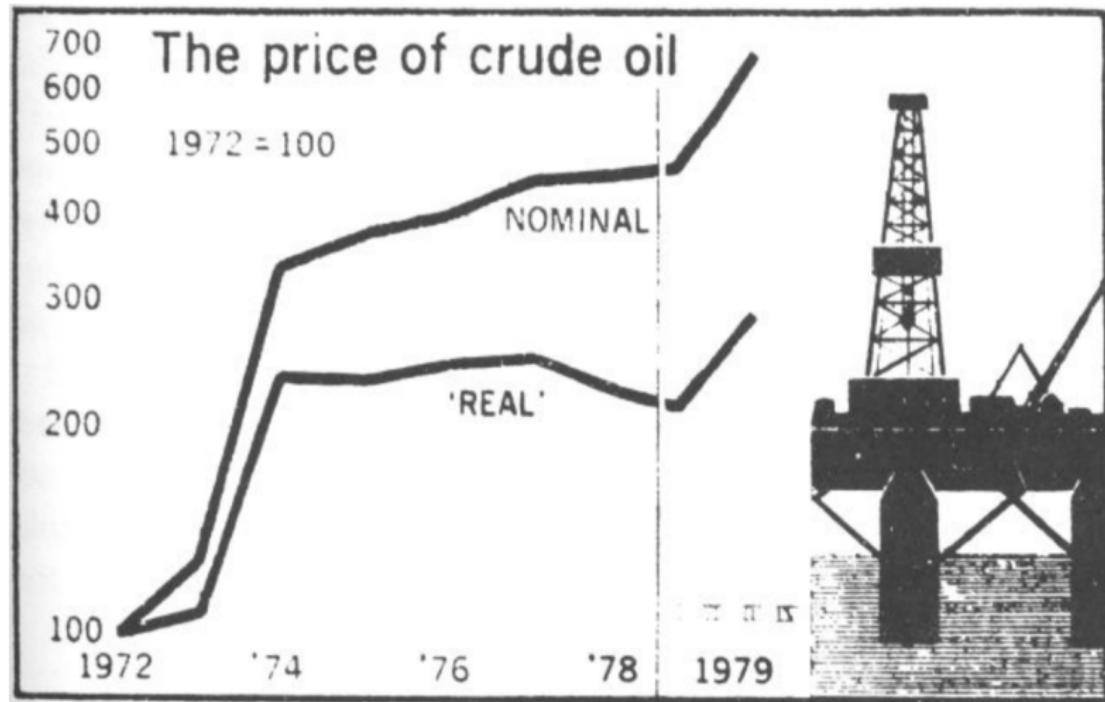
Lie factor

IN THE BARREL...

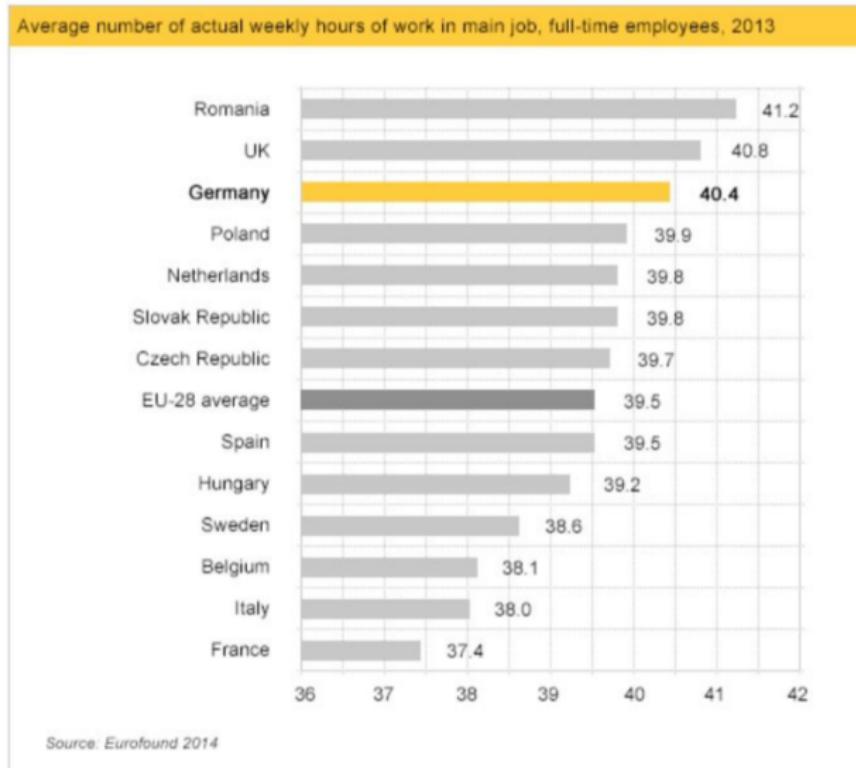
Price per bbl. of
light crude, leaving
Saudi Arabia
on Jan. 1



Lie factor



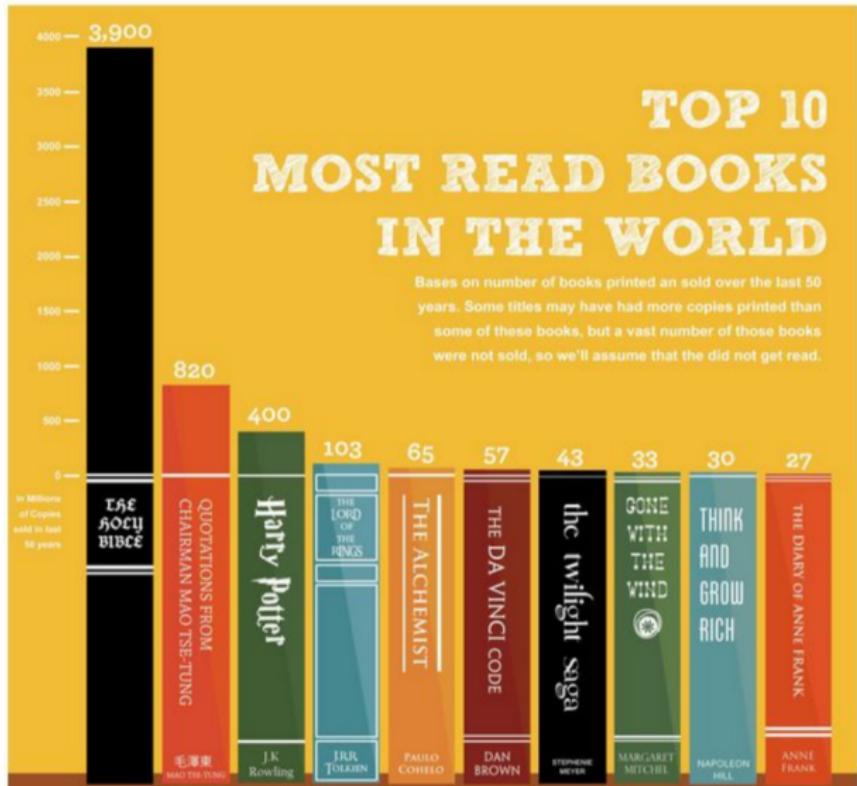
Lie factor



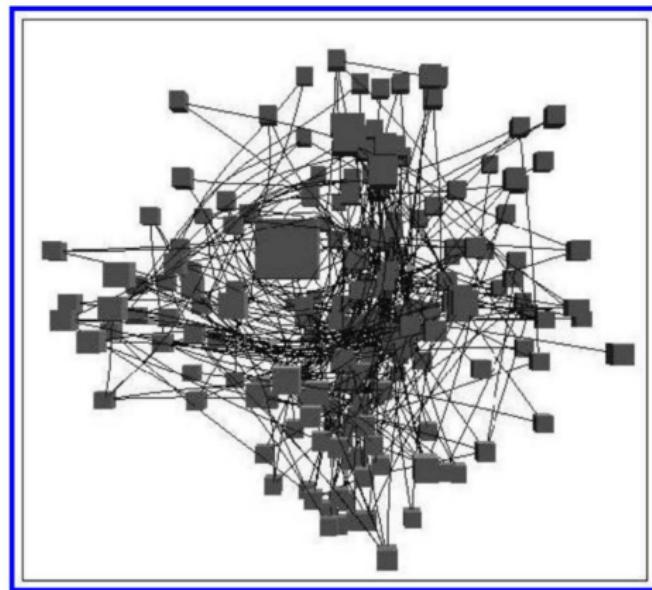
Lie factor



Lie factor



No al 3D injustificado

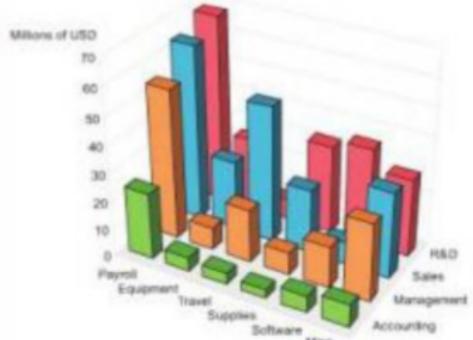


No al 3D injustificado

Graph Design I.Q. Test

Question 7: Which graph makes it easier to determine R&D's travel expense?

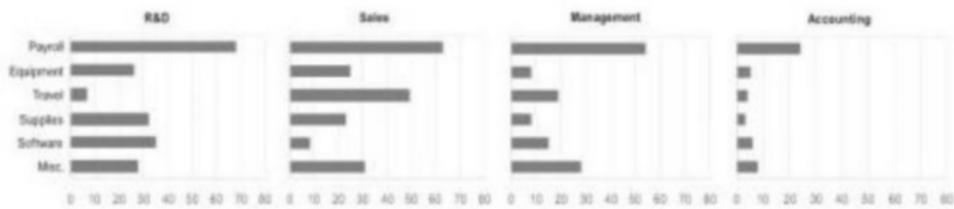
2006 Expenses by Department



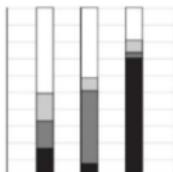
3-D Bar Graph (left)

2-D Bar Graphs (below)

2006 Expenses by Department in Millions of USD

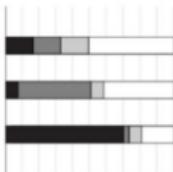


Consistencia



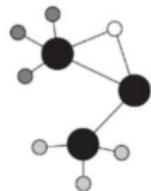
□ A
■ B
■ C
■ D

■ A
■ B
□ C
□ D



■ A
■ B
■ C
□ D

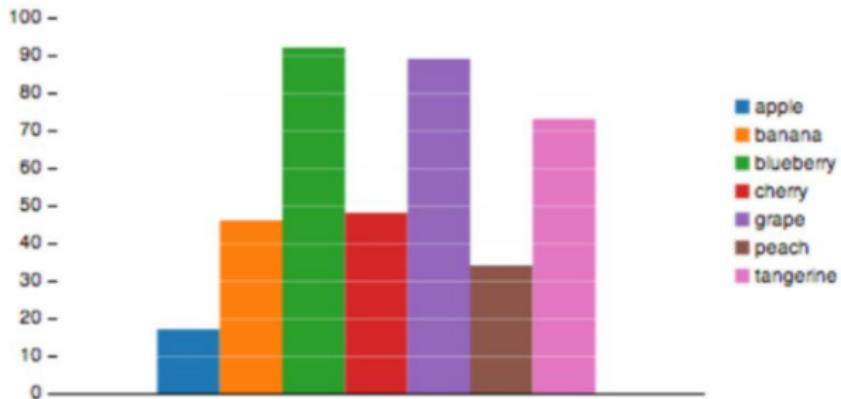
□ A
■ B
■ C
■ D



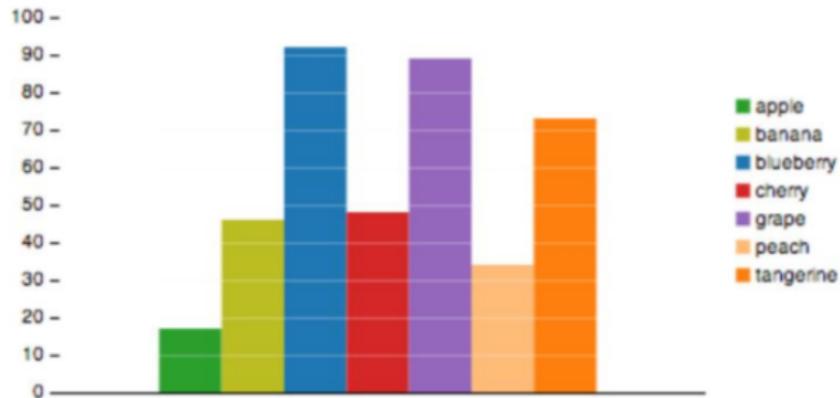
○ A
○ B
● C
● D

□ A
■ B
■ C
■ D

Consistencia



Consistencia



Más rules of thumb

- No al **2D injustificado**.
- Las unidades de medida deben ser **estandardizadas**.
- Los datos deben ser mostrados en su **contexto**.
- Elección de colores debe ser a prueba de **daltonismo**.
- Elección de **tipografía**.
- Gráficos **autoexplicativos**.

Framework

Modelo conceptual creado para guiar el diseño y elaboración de visualizaciones de información. Se concentra en tres preguntas alrededor de una visualización:

- ¿Qué?
- ¿Por qué?
- ¿Cómo?

Un poco sobre el ¿qué?

Semántica versus tipo de datos:

14, 2.8, 30, 30, 15, 1001

Santiago, 3, N, Nacimiento

Tipos de datos y *datasets*

Tipos de datos:

- Ítem
- Atributo
- Enlaces
- Posición
- ...

Tipos de *datasets*:

- Tablas
- Redes y árboles
- Geométricos
- ...

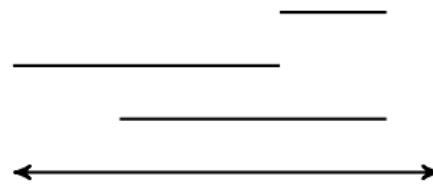
Tipos de atributos

- Categóricos



- Ordenados

- Cuantitativo



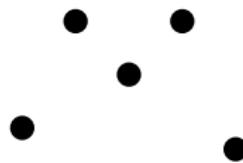
- Ordinal

S - M - L

1er - 2ndo - 3er - ...

Marcas y canales

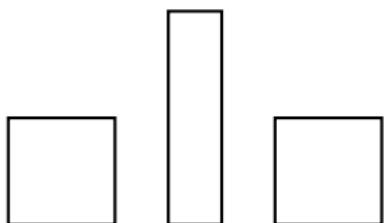
Una **marca** es un **elemento geométrico básico** utilizado para representar un dato.



Puntos



Líneas



Áreas

Marcas y canales

Mientras que un **canal** visual es una característica de la marca que me permite alterar su apariencia para representar un atributo del dato.



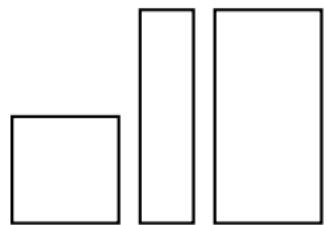
Posición



Color

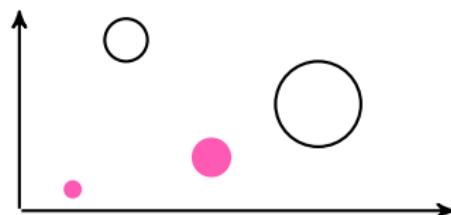
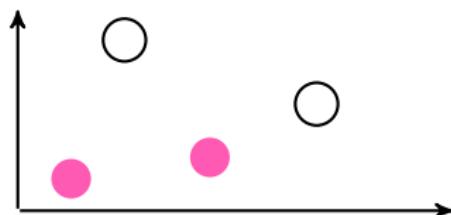
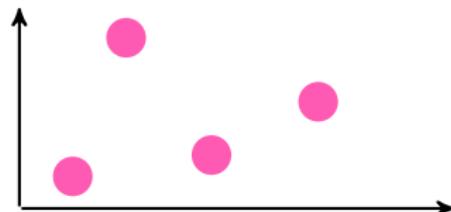
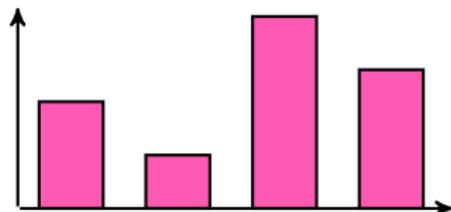


Largo

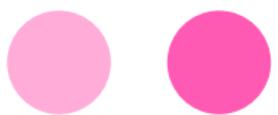
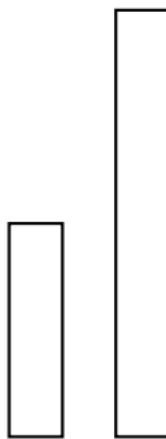
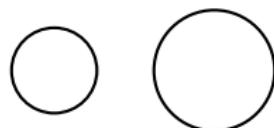


Área

Marcas y canales



Efectividad de canales



Efectividad de canales

Channels: Expressiveness Types and Effectiveness Ranks

⊕ **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)



⊕ **Identity Channels: Categorical Attributes**

Spatial region



Color hue



Motion



Shape



▲ ↑ Most
Effectiveness
↓ Least
▼ Same

IIC2026 - Visualización de Información

Competencias:

- Aplicar un proceso de diseño para crear visualizaciones efectivas,
- Llevar ideas a prototipos concretos, con la ayuda de bosquejos,
- Utilizar principios de percepción y cognición humana en visualización,
- Evaluar una visualización de forma crítica, pudiendo además sugerir e implementar mejoras,

IIC2026 - Visualización de Información

- Exponerse a distintos dominios de datos (e.g. redes, textos, cartografía),
- Aplicar distintos métodos de visualización para un rango variado de datasets.

Pueden acceder a la organización en **GitHub** del curso en sus tres versiones. Podrán acceder a ayudantías y clases de semestres pasados.

¡Muchas gracias!