
IIC1005 - Computación: Ciencia y Tecnología del Mundo Digital

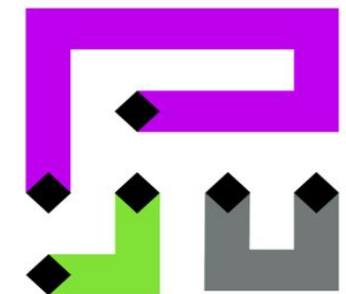
Visualización de información

[Hernán Felipe Valdivieso López](#)

hfvaldivieso@uc.cl

Hernán Valdivieso

- Actualmente profesor diplomados “Big Data” y “Big Data y Ciencias de datos”.
- Fuí ayudante en Diplomado “Big Data” (Visualización y Sistemas Recomendadores).
- Hice ayudantía en Pregrado y Postgrado: 21 ayudantías entre 9 cursos distintos.
- Soy parte de [IALab](#), [HAIVis](#) e [IMFD](#).
- Área de Inteligencia Artificial & Visualización de Información.

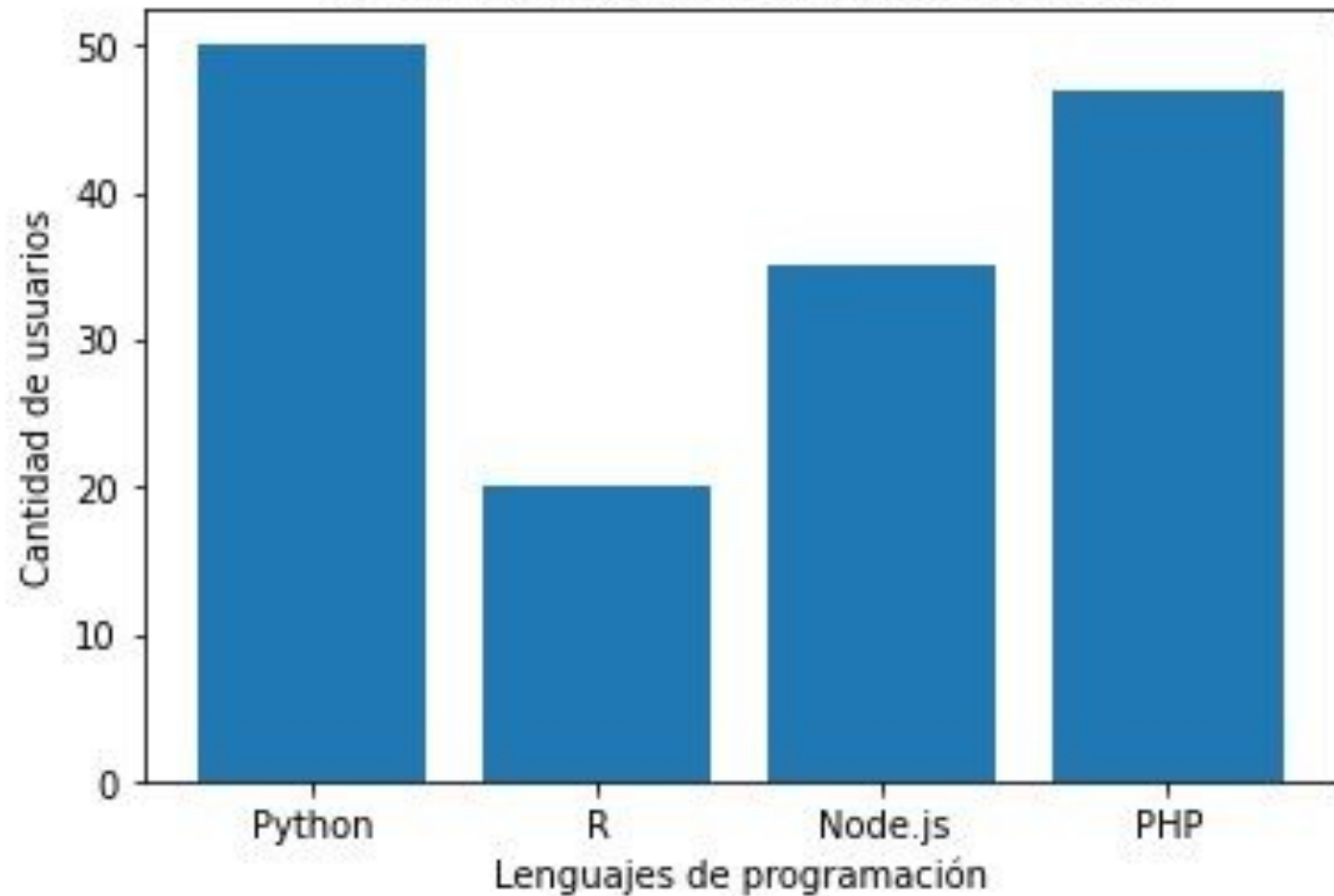


Hernán Valdivieso

- Amante de la cultura y animación japonesa
- Programando visualizaciones en esta área
 - [Proyecto Visualización IIC2026 \(2017\) - Anime](#)
 - [Anime Visualización 2021](#) (En proceso)
 - Otros: [Proyecto Sustentabilidad](#)
- Creando datasets (conjunto de datos) en esta área
 - [Anime Recommendation Database 2020 \(kaggle\).](#)
 - [Anime-Planet Recommendation Database 2020 \(kaggle\).](#)
 - [Anime-Planet Character \(waifus/husbando\) Database 2020 \(kaggle\).](#)

Antes de empezar...

Usuarios de lenguajes de programación



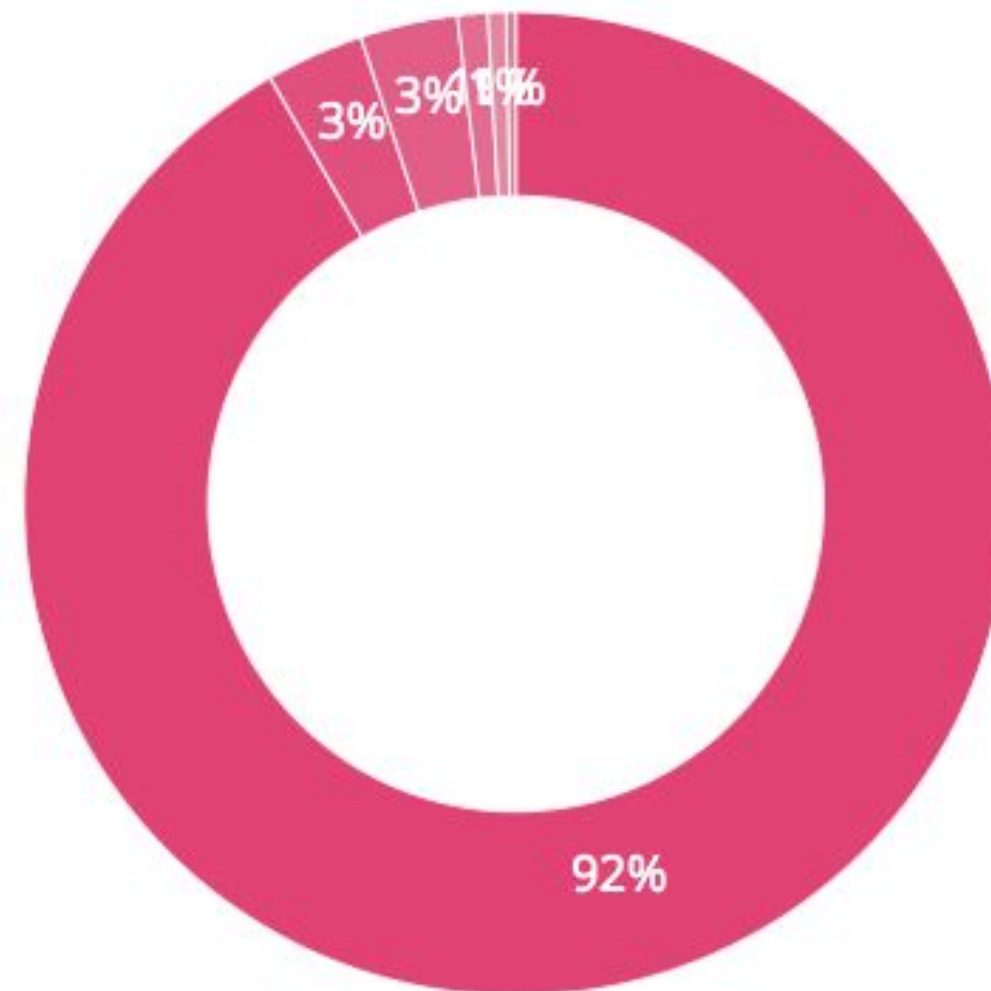
Las vistas al ejercicio en casa aumentaron en Chile

Vistas diarias de videos de ejercicio con 'en casa' o 'sin equipo' en el título



Fuente: Perú Vistas de YouTube, 2/1/2020 – 5/1/2020

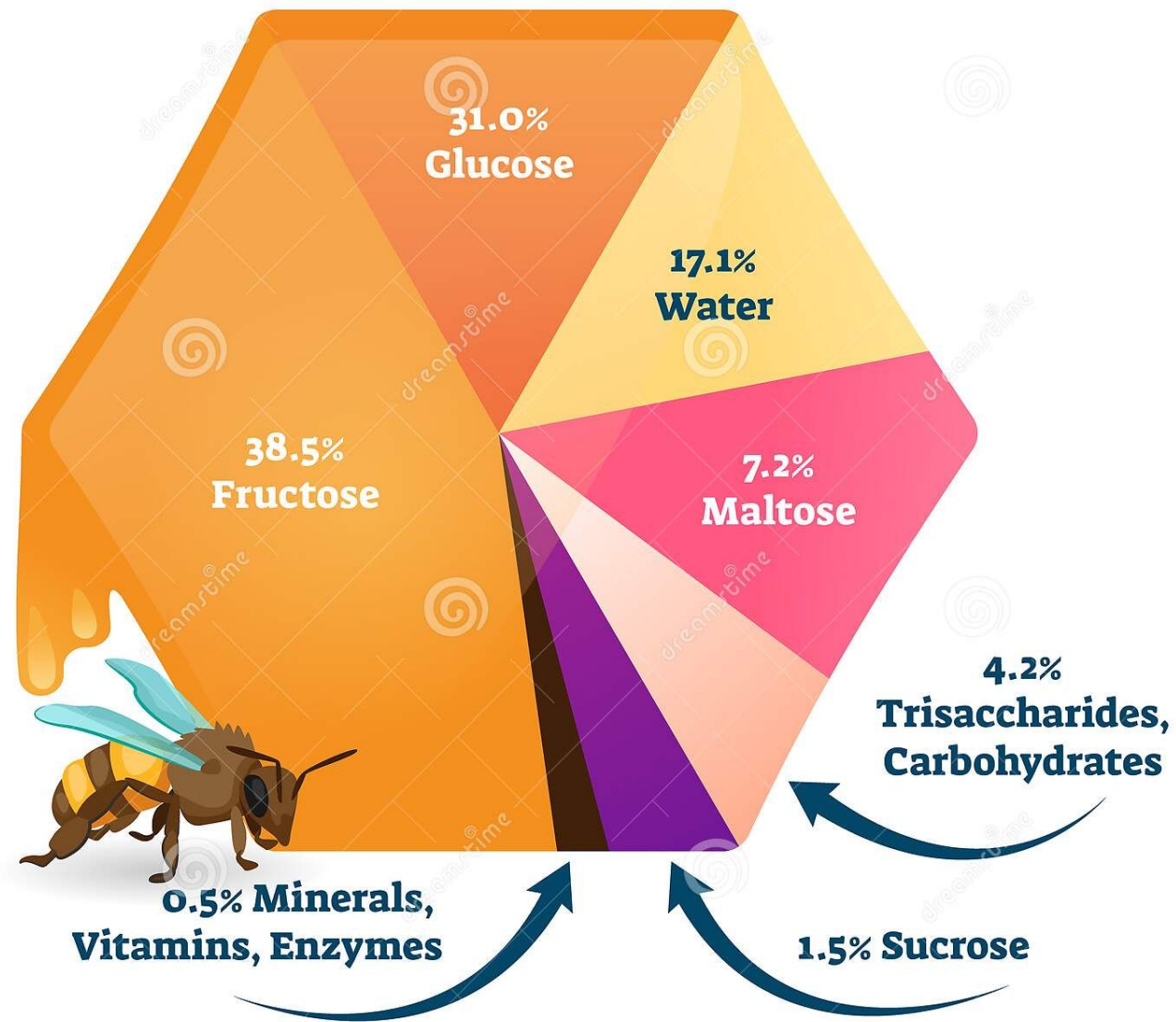
Antes de empezar...



● Contribuições ● Outros Serviços de Terceiros - Pessoa Jurídica ● Despesas de Exercícios Anteriores ● Material de Consumo ● Locação de Mão-de-Obra ● Outros

Antes de empezar...

COMPOSITION OF HONEY



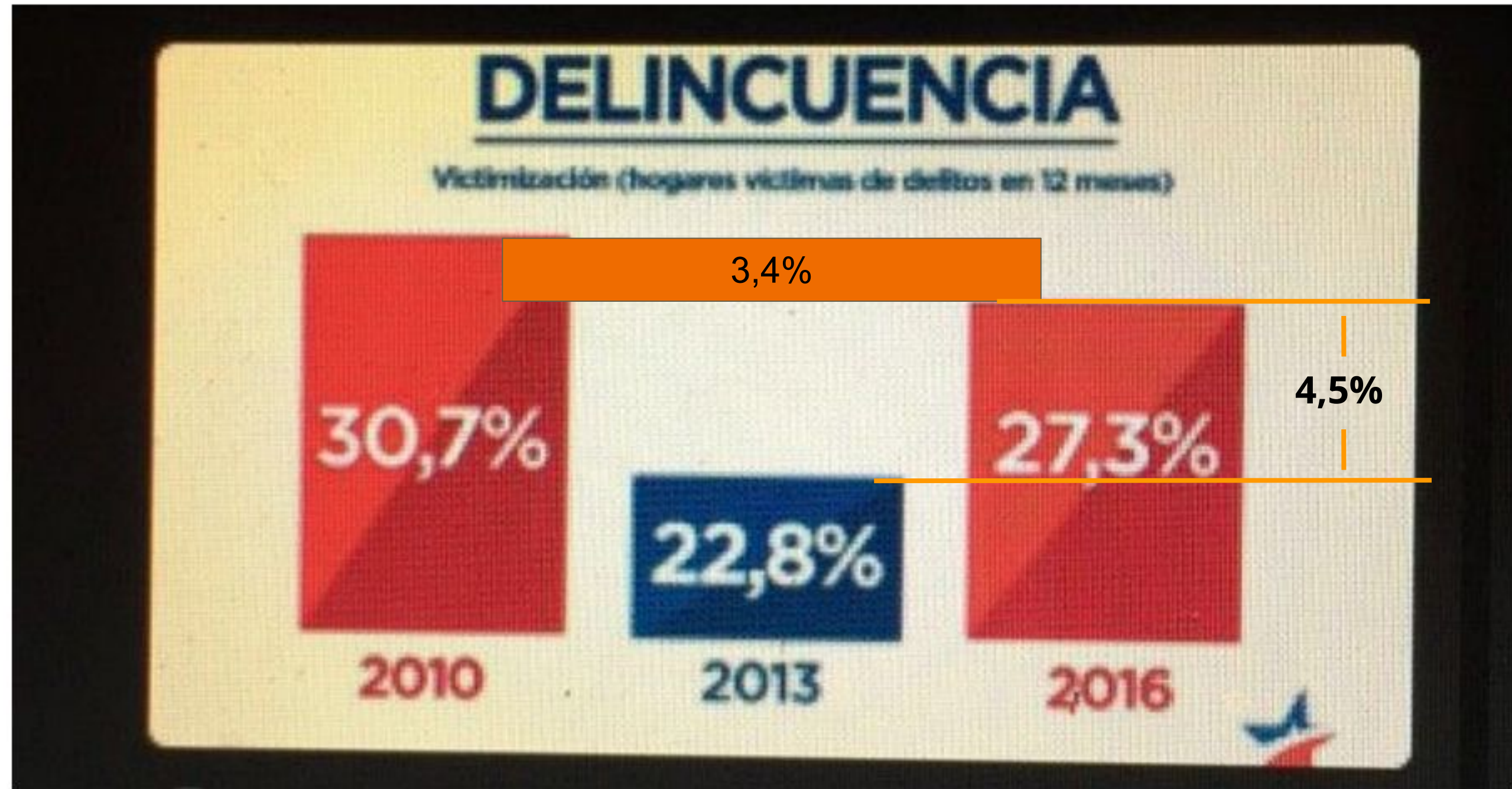
Antes de empezar...



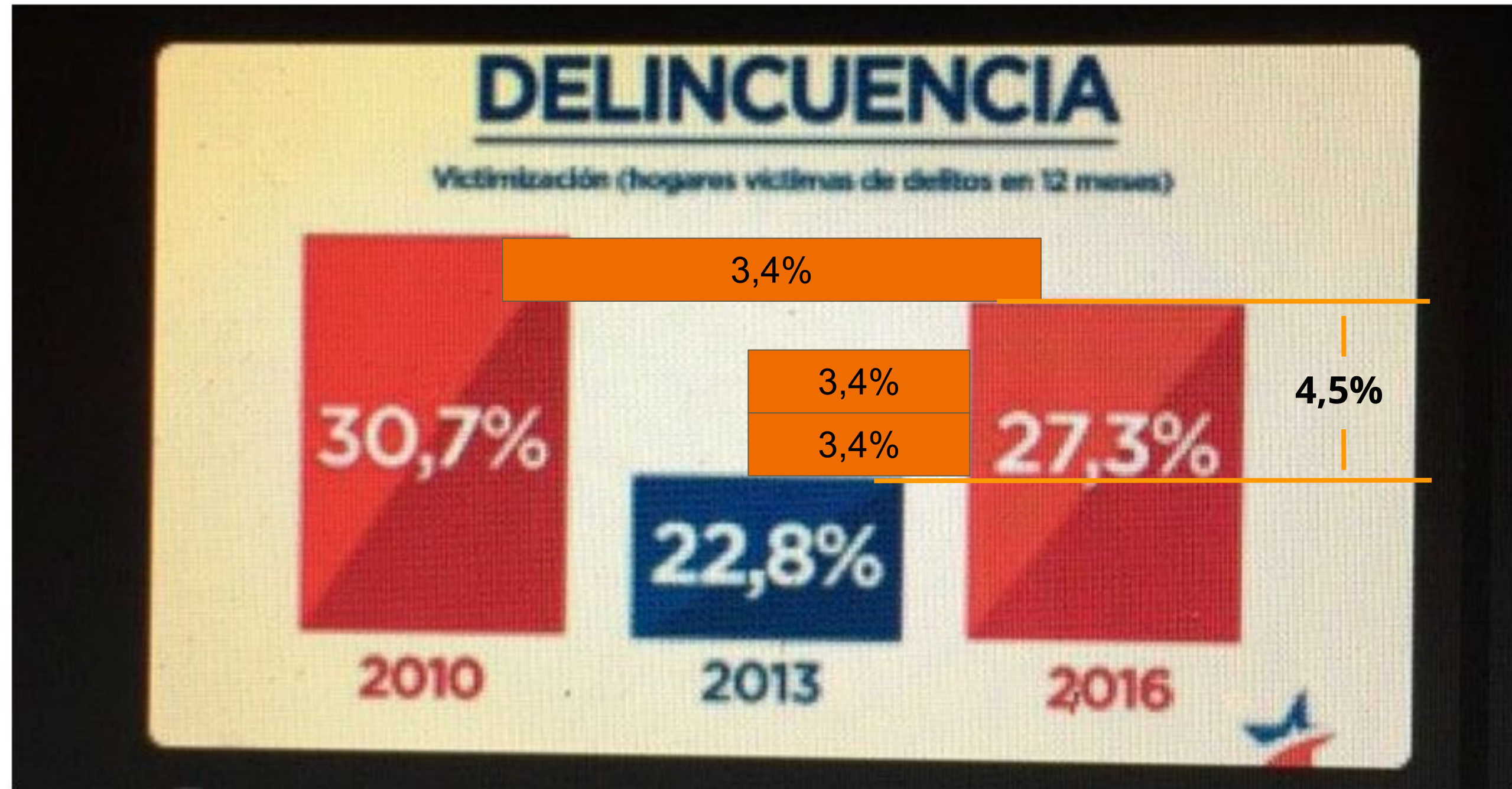
Antes de empezar...



Antes de empezar...



Antes de empezar...



Contenidos

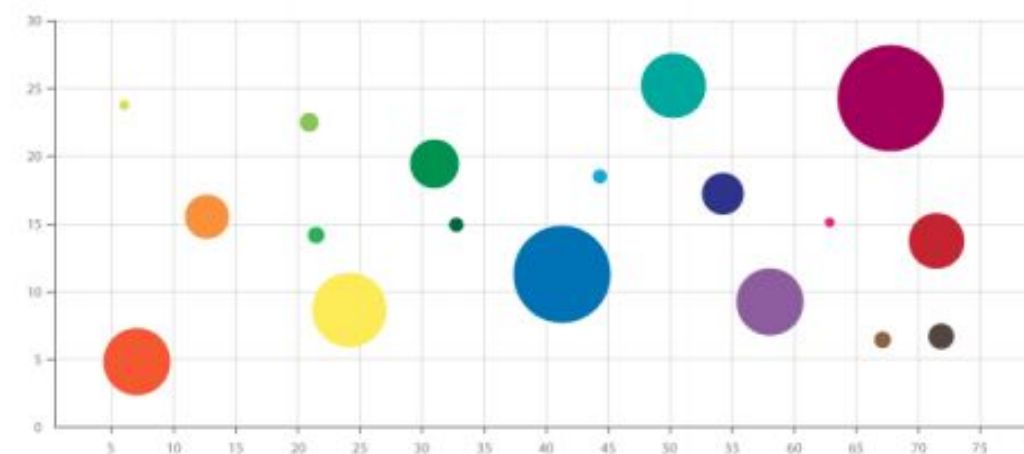
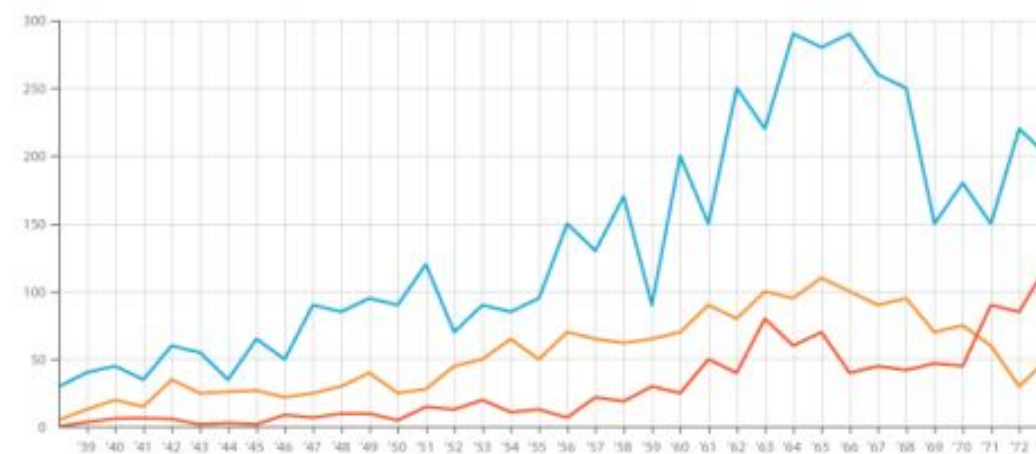
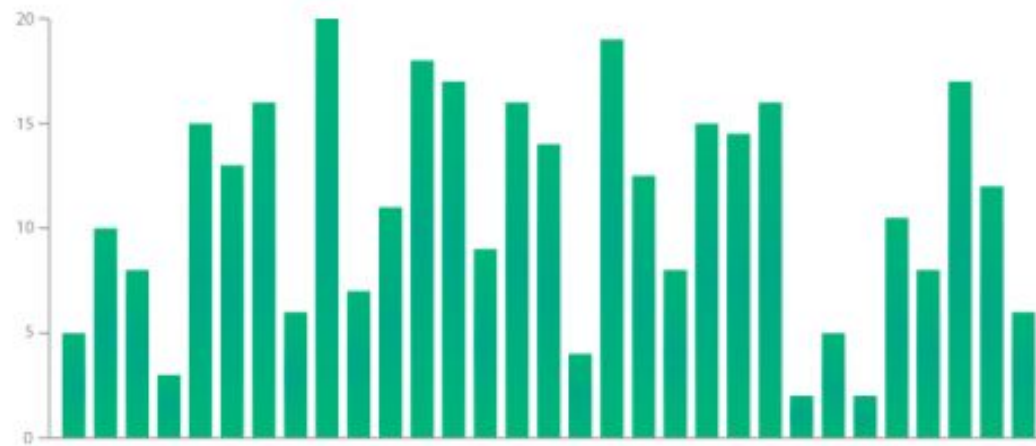
- ¿Qué?
- ¿Por qué?
- ¿Cómo diseñar una visualización? - *Rules of Thumb*
- *Framework* de visualización
 - Etapas que componen el *framework*
 - Marcas y canales
 - Efectividad de canales
- IIC2026 - Visualización de Información

Contenidos

- ¿Qué?
- ¿Por qué?
- ¿Cómo diseñar una visualización? - *Rules of Thumb*
- *Framework* de visualización
 - Etapas que componen el *framework*
 - Marcas y canales
 - Efectividad de canales
- IIC2026 - Visualización de Información

¿Qué es la visualización?

¿Qué es la visualización?



Catálogo de Visualización de Datos

¿Qué es la visualización?

Según los investigadores del área:

1. "Transformación de lo simbólico a lo **geométrico**". [McCormick et al. 1987]
2. "[...] encontrar la **memoria artificial** que mejor apoya a nuestros medios naturales de **percepción**." [Bertin 1967]
3. "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?

“Sistemas de visualización computarizado que brindan una **representación visual de los *datasets*** (conjunto de datos) que están diseñados para ayudar a las personas a **realizar tareas más eficazmente**”.

“La visualización es adecuada cuando existe la necesidad de **aumentar las capacidades humanas** en lugar de reemplazar a las personas con métodos computacionales para la toma de decisiones”

Definiciones provistas por de [Tamara Munzner](#)
Autora del libro “[Visualization Analysis and Design](#)” (2015)

¿Qué es la visualización de información?

Se identifican 3 temas claves:

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

Contenidos

- ¿Qué?
- ¿Por qué?
- ¿Cómo diseñar una visualización? - *Rules of Thumb*
- *Framework* de visualización
 - Etapas que componen el *framework*
 - Marcas y canales
 - Efectividad de canales
- IIC2026 - Visualización de Información

¿Por qué?

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

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

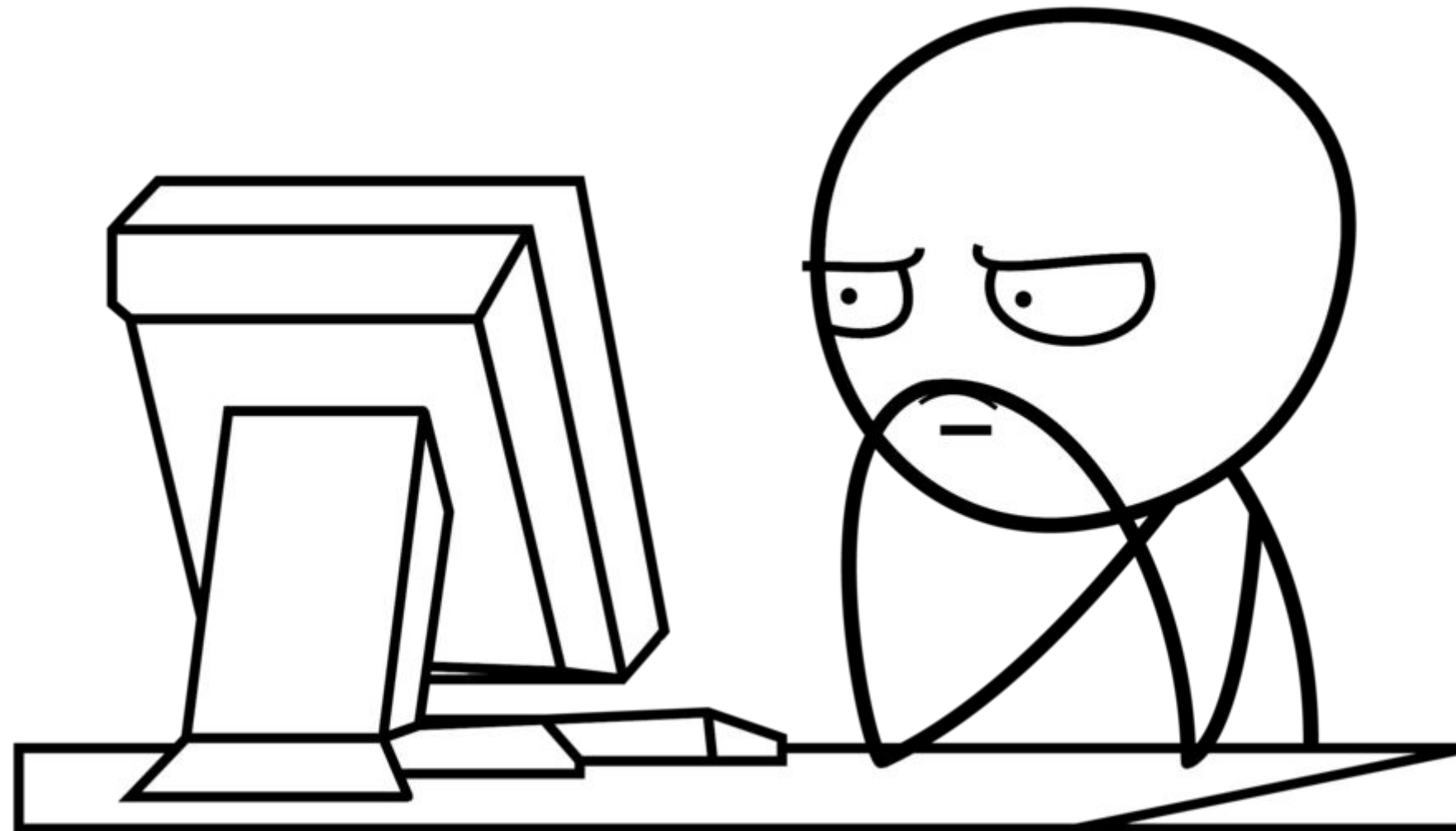
$$1 \text{ *zettabyte*} = 10^{21} \text{ *bytes*} = 10^9 \text{ *terabytes*}$$

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

¿Por qué?

175 zetabytes...

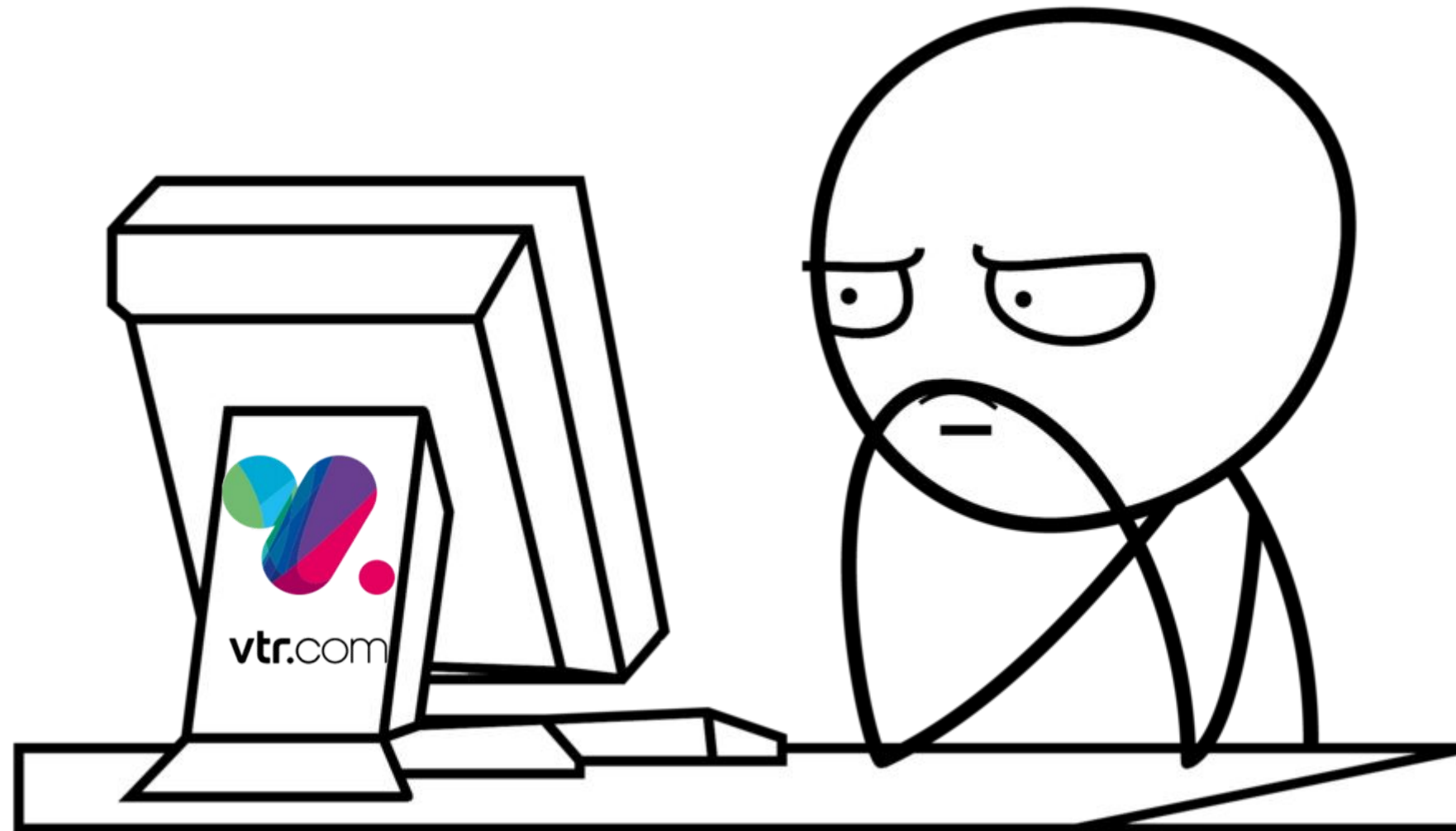
Descargando a 25MB/s toma **1,8 billones** de años



¿Por qué?

175 zetabytes...

Descargando a 5MB/s toma **9 billones** de años



¿De dónde proviene toda esta información?

Sensores físicos: datos en vivo de vuelos.



¿De dónde proviene toda esta información?

Nosotros mismos generamos nuevos 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* (datos atípicos/anómalos)
- Descubrir datos faltantes
- Comunicar información
- Reducir carga cognitiva para procesar información
- Los gráficos son necesarios para explicar y verificar los datos

Un poco de historia - John Snow

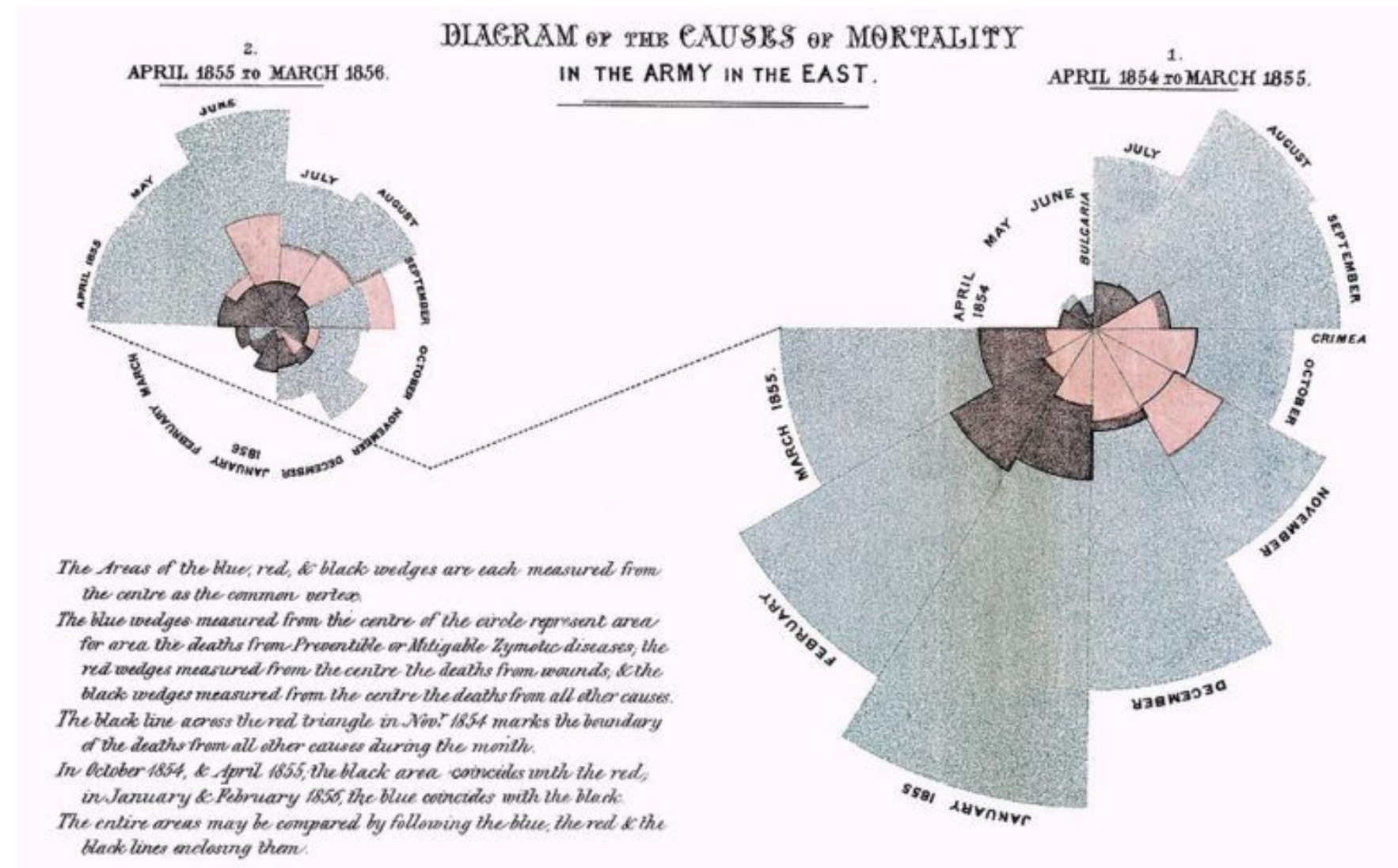
En 1854, durante una epidemia de Cólera en Londres, el Dr. John Snow usa un análisis espacial para **apoyar su hipótesis**.

Encontró que las muertes eran principalmente por una bomba de agua contaminada.



Un poco de historia - Florence Nightingale

En 1858, Florence Nightingale, dibujó este gráfico para fin de ilustrar las **causas de la mortalidad** de los soldados en el hospital militar que dirigía durante la guerra de Crimea. Se dio cuenta de que el 80% de los muertos eran víctimas de los deficientes tratamientos sanitarios.



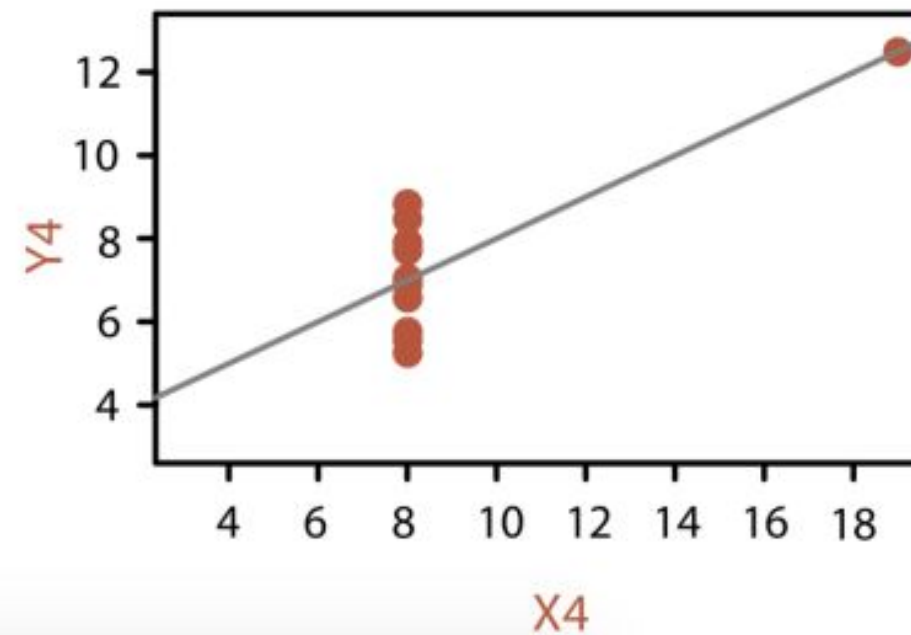
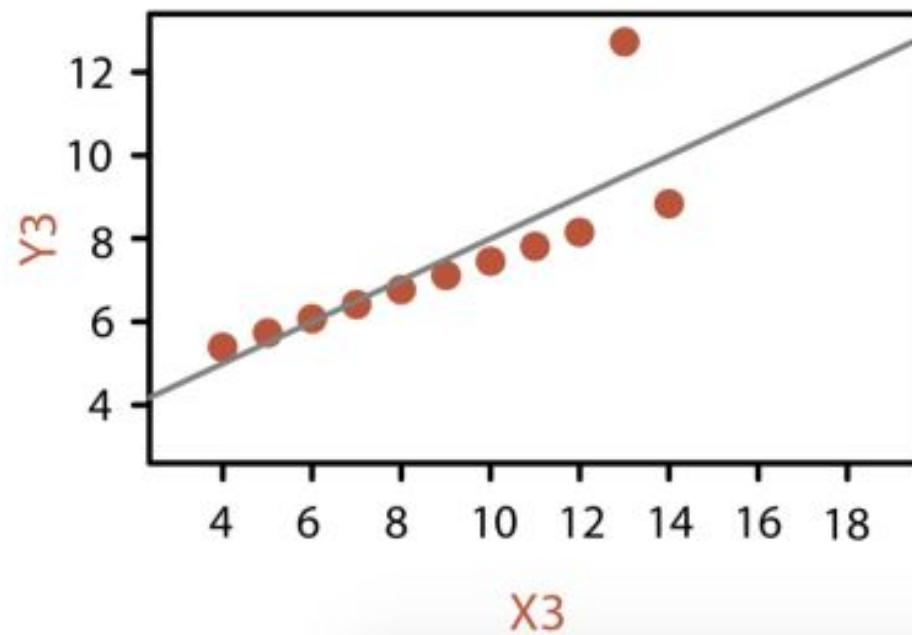
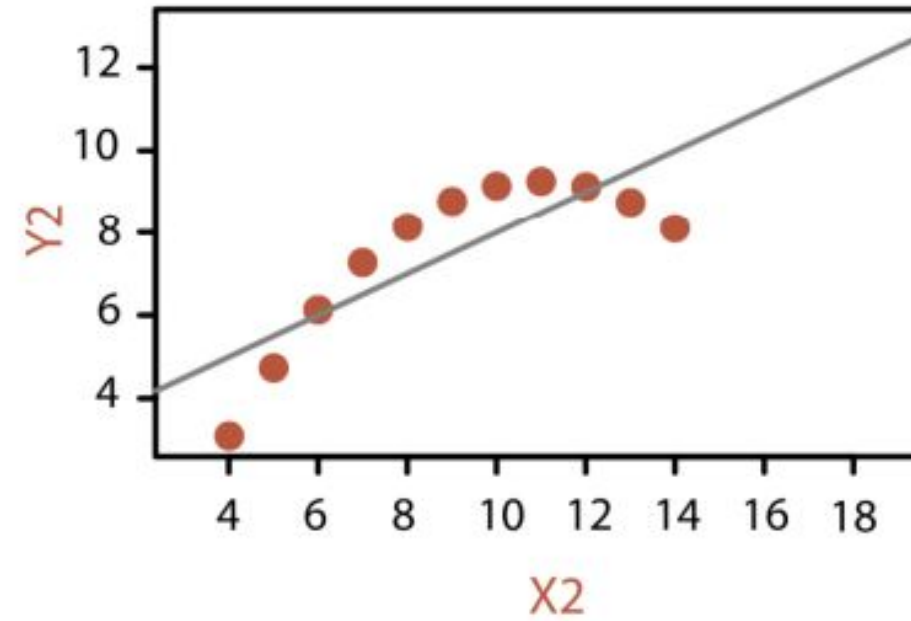
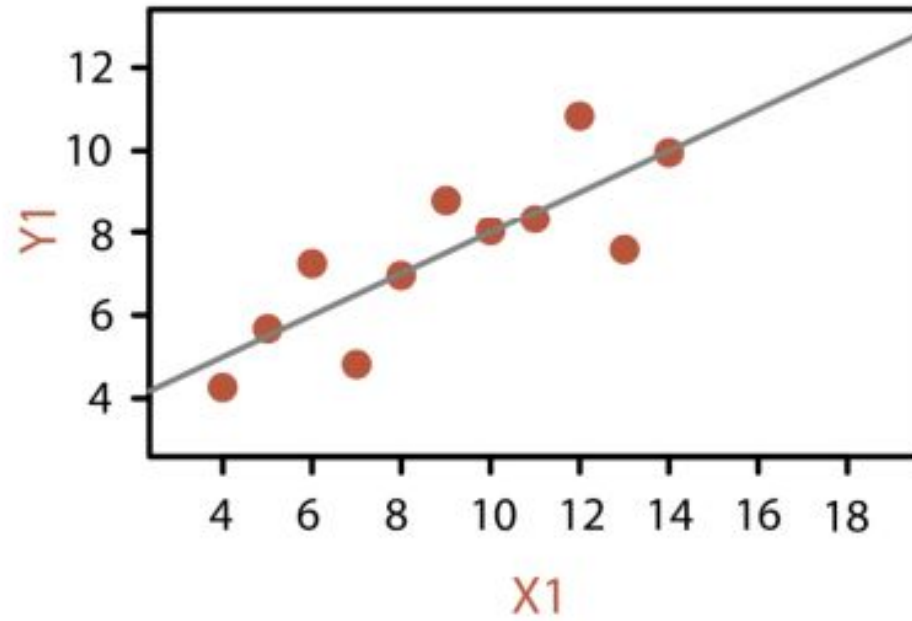
Un poco de historia - Anscombe's Quartet (1973)

Análisis de Regresión

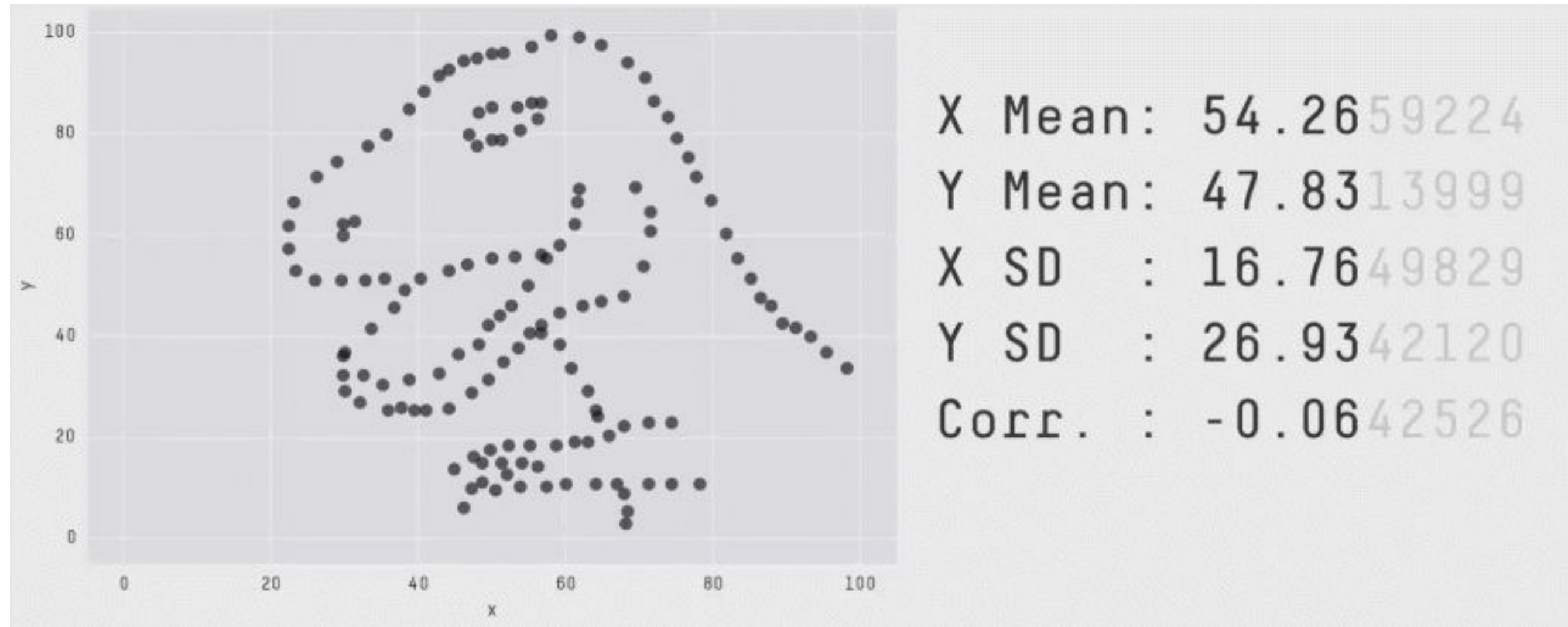
	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	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
	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	

Un poco de historia - Anscombe's Quartet (1973)

Análisis de Regresión



Anscombe's Quartet Versión 2017



[Fuente](#)

Contenidos

- ¿Qué?
- ¿Por qué?
- ¿Cómo diseñar una visualización? - *Rules of Thumb*
- *Framework* de visualización
 - Etapas que componen el *framework*
 - Marcas y canales
 - Efectividad de canales
- IIC2026 - Visualización de Información

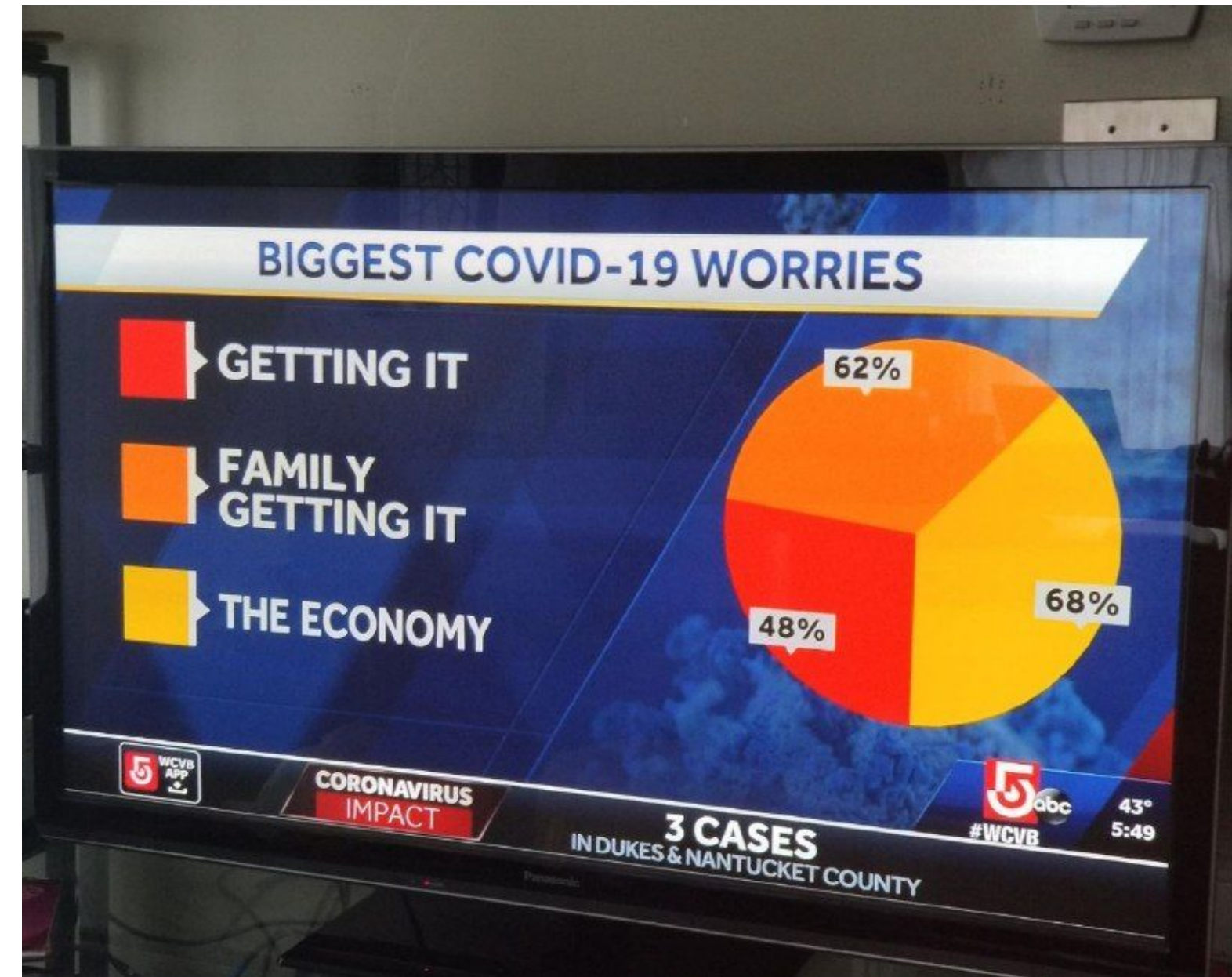
¿Cómo diseñar una visualización?

- 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. Primero debe ser efectivo** (y luego que sea bonito).
- La mayoría de los diseños de visualización son **inefectivos**.

¿Cómo diseñar una visualización?

- Existen **pocas verdades** en esta disciplina.
- No hay un claro método para optimizar, pero si existen **guidelines** que uno puede seguir.
- **Validar** un diseño de visualización es un proceso **sumamente difícil**.

Malos gráficos hasta el día de hoy...



Malos gráficos hasta el día de hoy...

EXPLORATION OF INTERPRETABILITY TECHNIQUES FOR DEEP COVID-19 CLASSIFICATION USING CHEST X-RAY IMAGES

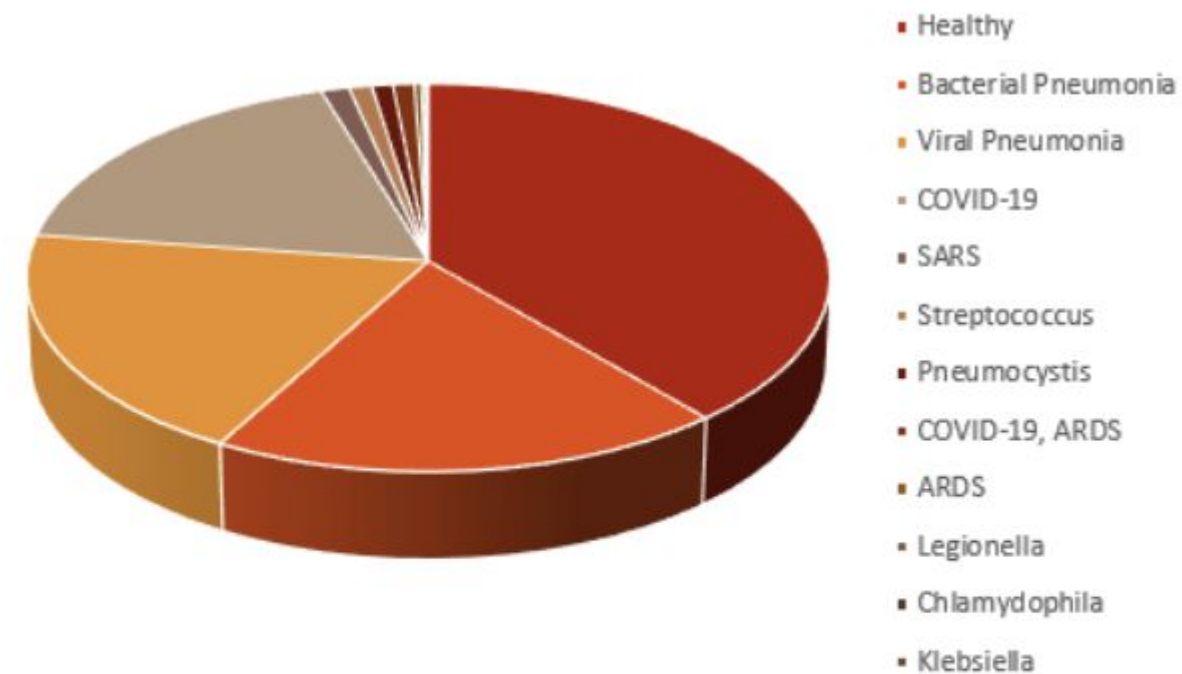


Fig. 1. CXR images distribution for each infection type in the dataset

Fuente

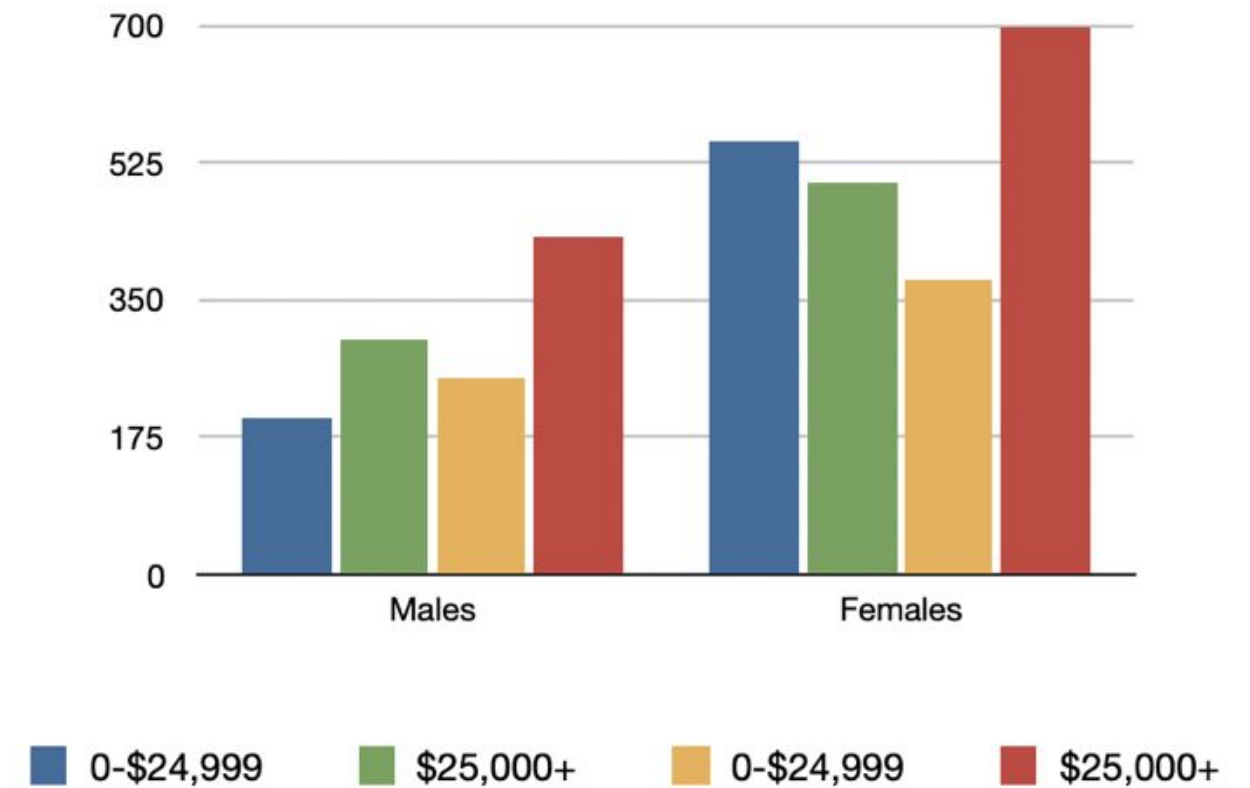
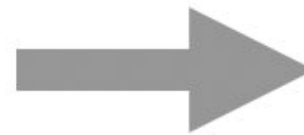
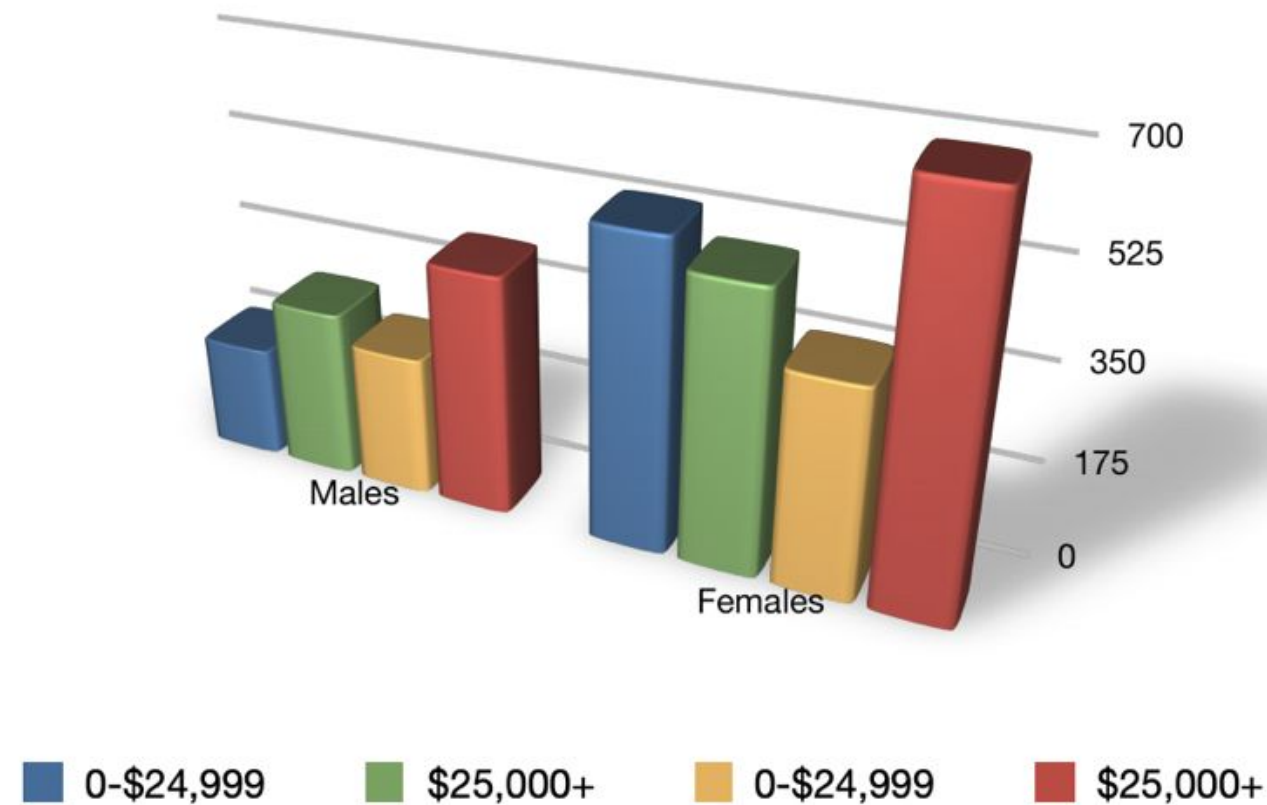
Rules of Thumbs

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

Rules of Thumbs

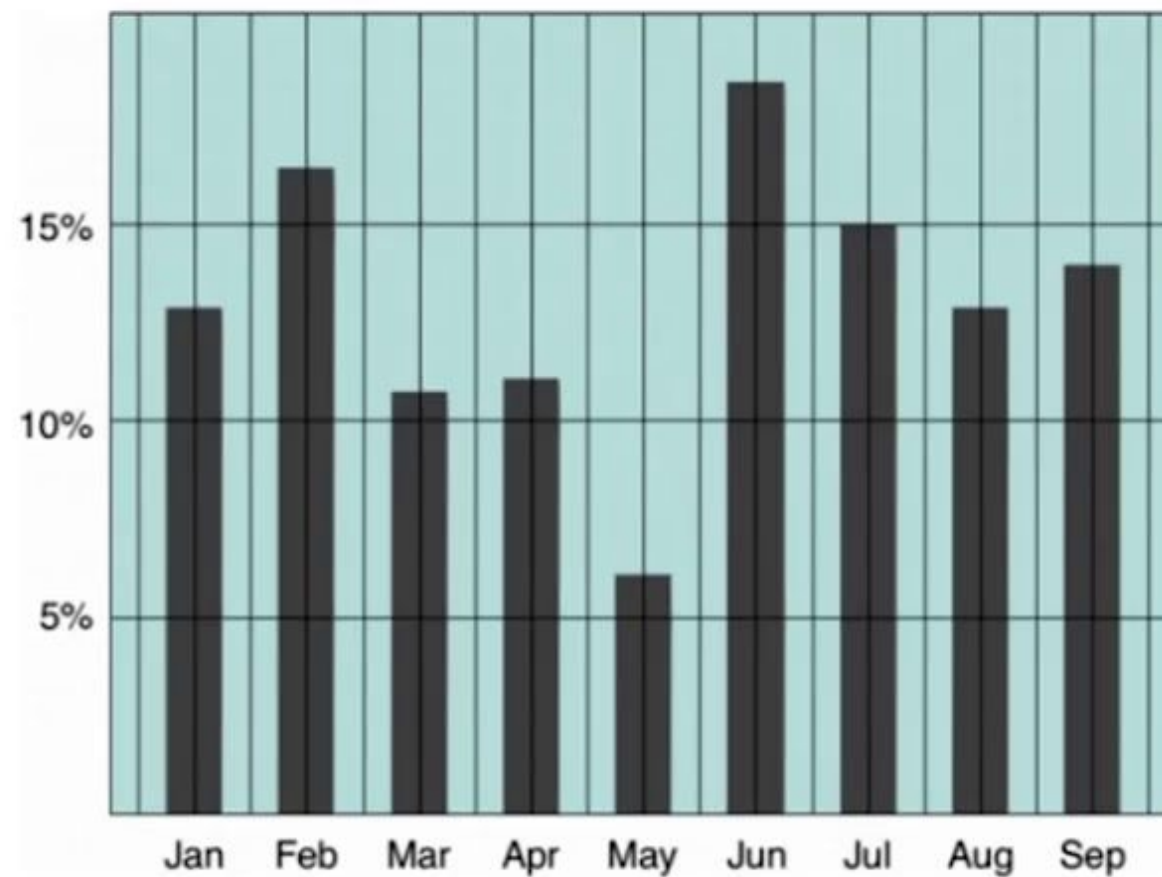
Data ink ratio (Tufte)

$$\text{Data-Ink Ratio} = \frac{\text{Data ink}}{\text{Total ink used in graphic}}$$

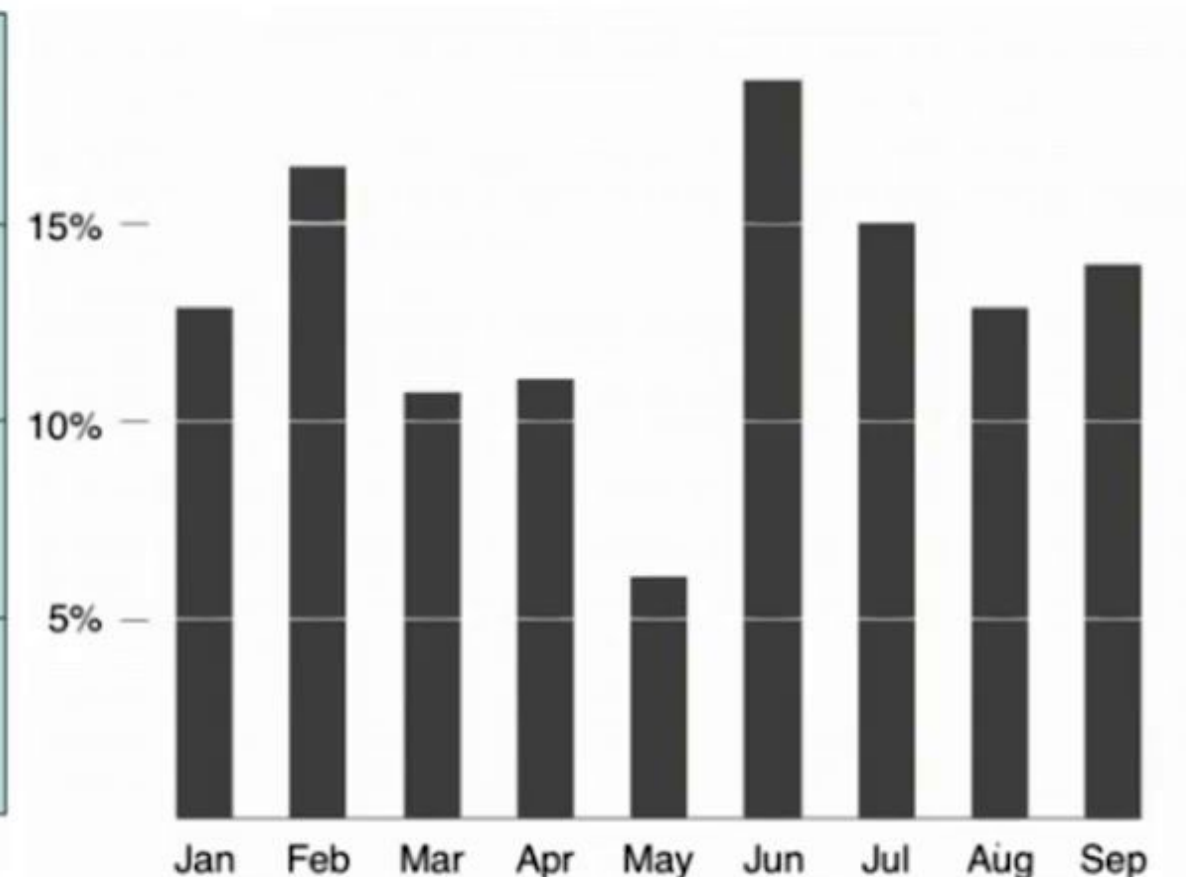


Rules of Thumbs

Data ink ratio (Tufte)



low data-ink ratio



high data-ink ratio

Rules of Thumbs

Data ink ratio (Tufte)

GOLD PRICE PER OUNCE IN \$US

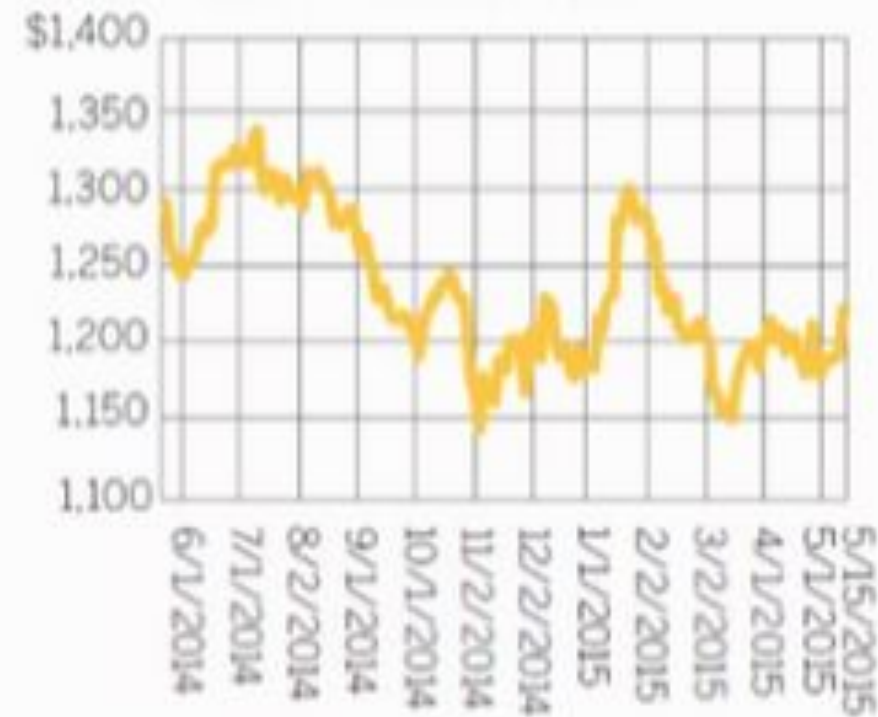


Context: Prototype

Use: Research, individual, informal

Media: Personal screen, paper

GOLD PRICE PER OUNCE IN \$US

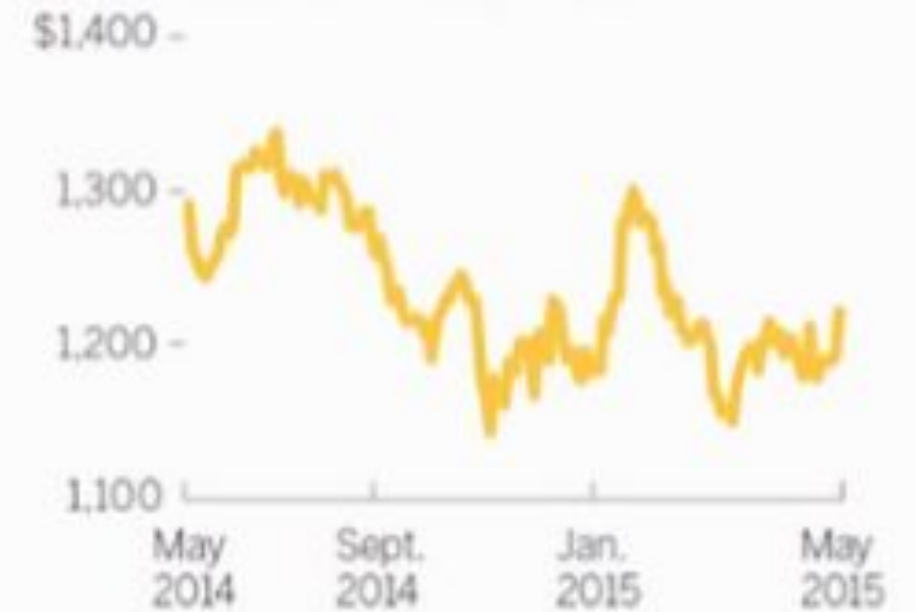


Context: "Let's talk about gold prices"

Use: Analysis, informal or formal,
one-on-one, small group

Media: Paper, personal screen, public screen

GOLD PRICE PER OUNCE IN \$US



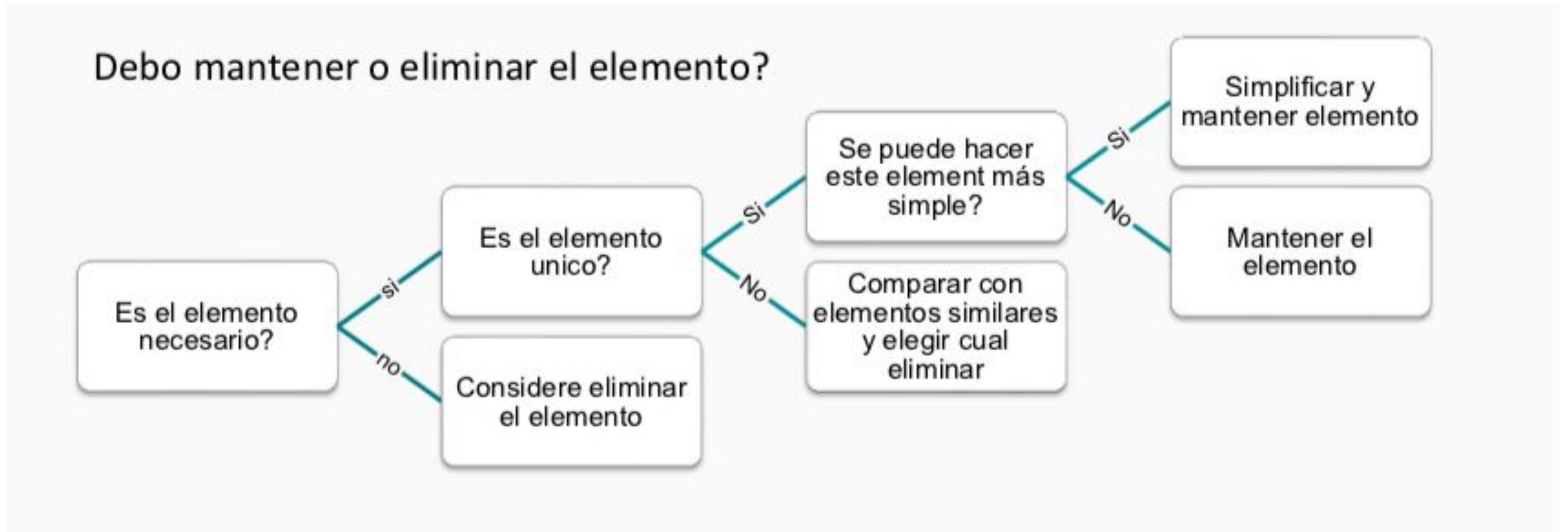
Context: "Gold prices are dropping this year"

Use: Presentation, formal, small or large group

Media: Paper, small screen or large screen

Rules of Thumbs

Data ink ratio (Tufte)



Rules of Thumbs

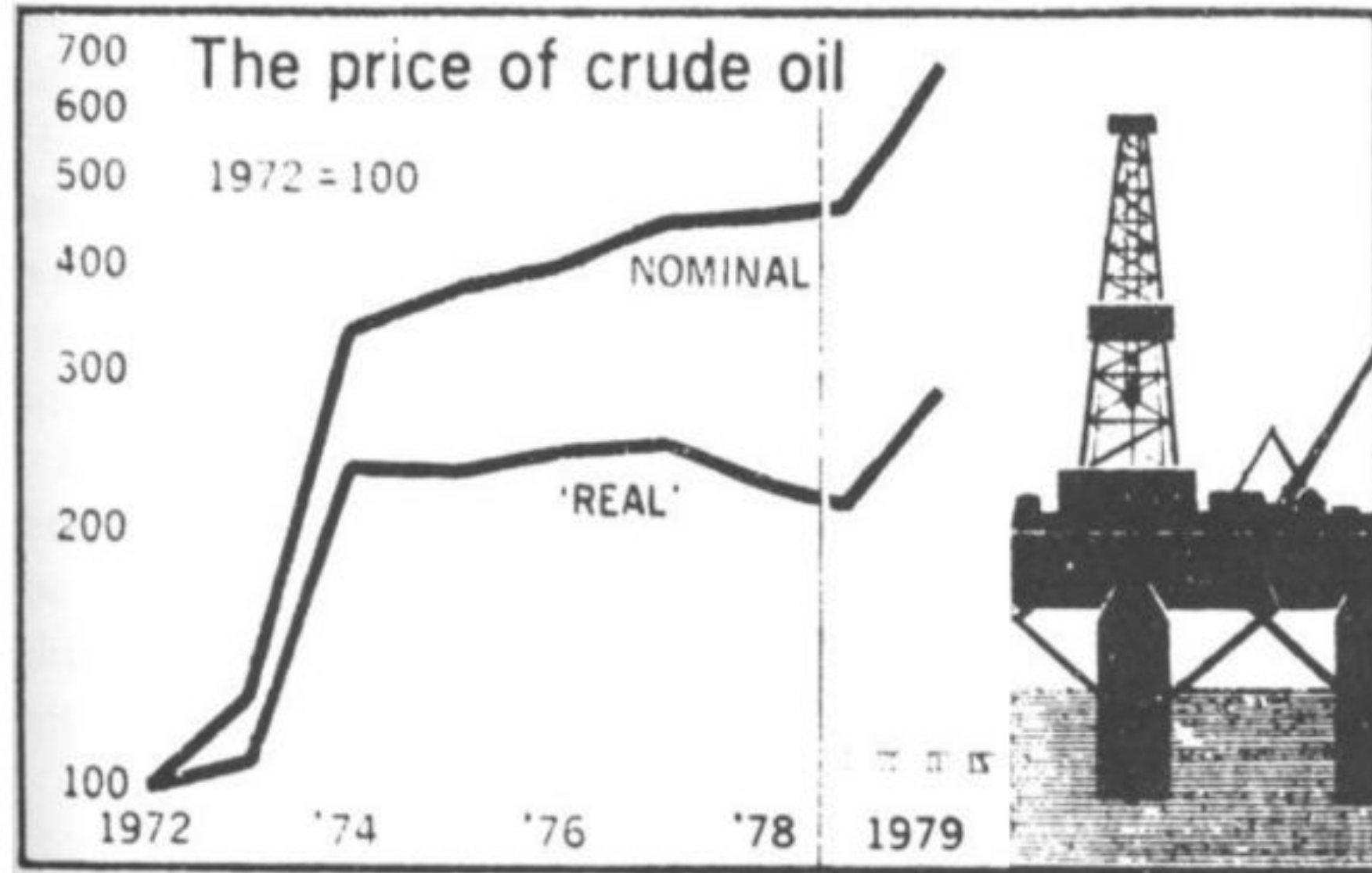
Lie factor

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

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**.

Rules of Thumbs

Lie factor



Bonus - Percepción

A raíz de lo anterior, hay que tener cuidado con la visualización a utilizar. Hay canales que se perciben diferente. Por ejemplo, el cambio de largo de una barra se percibe distinto al cambio de área de una circunferencia.

Esto repercute totalmente en el *lie factor*.

Bonus - Percepción

Estimación de magnitud

¿Cuántas veces es más grande el círculo de la derecha? (En términos del área)



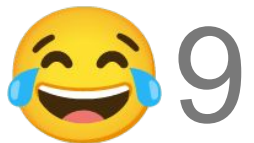
2



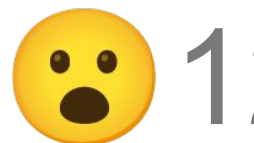
3



6



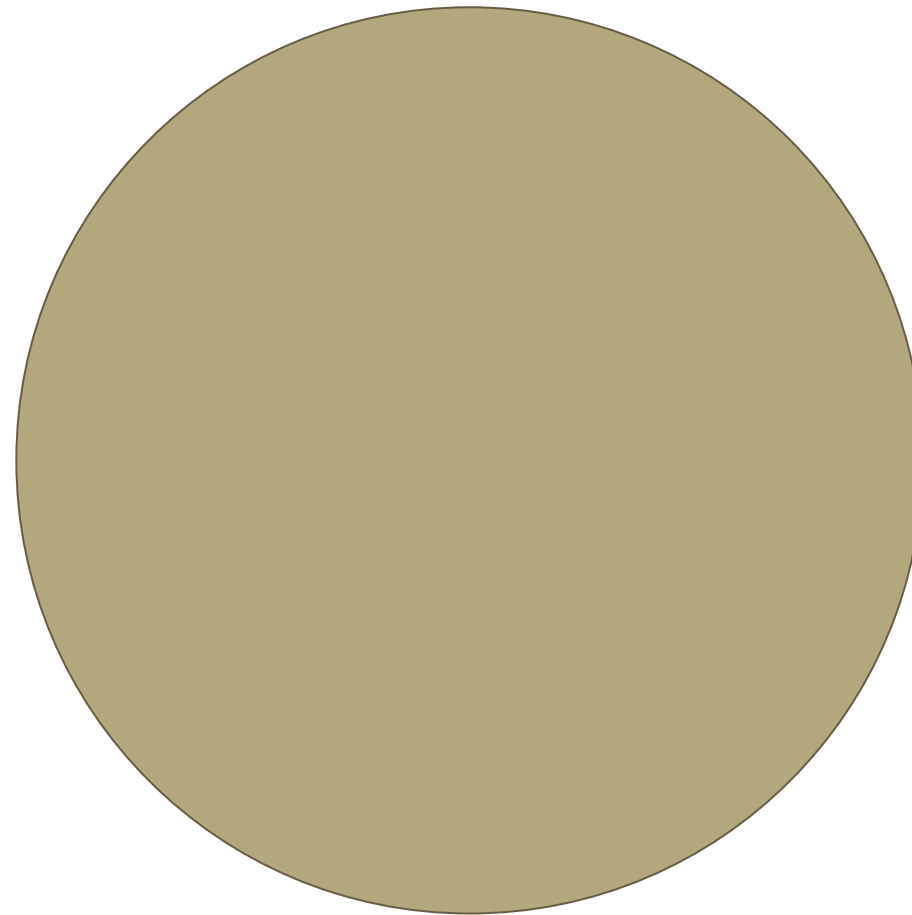
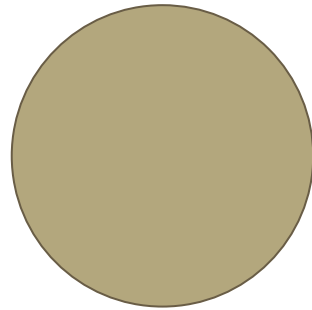
9



12



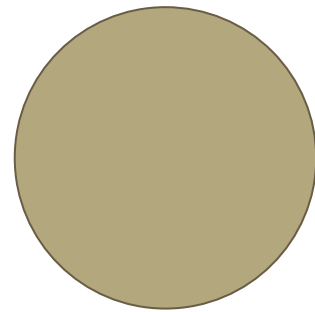
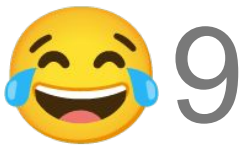
15



Bonus - Percepción

Estimación de magnitud

¿Cuántas veces es más grande el círculo de la derecha? (En términos del área)

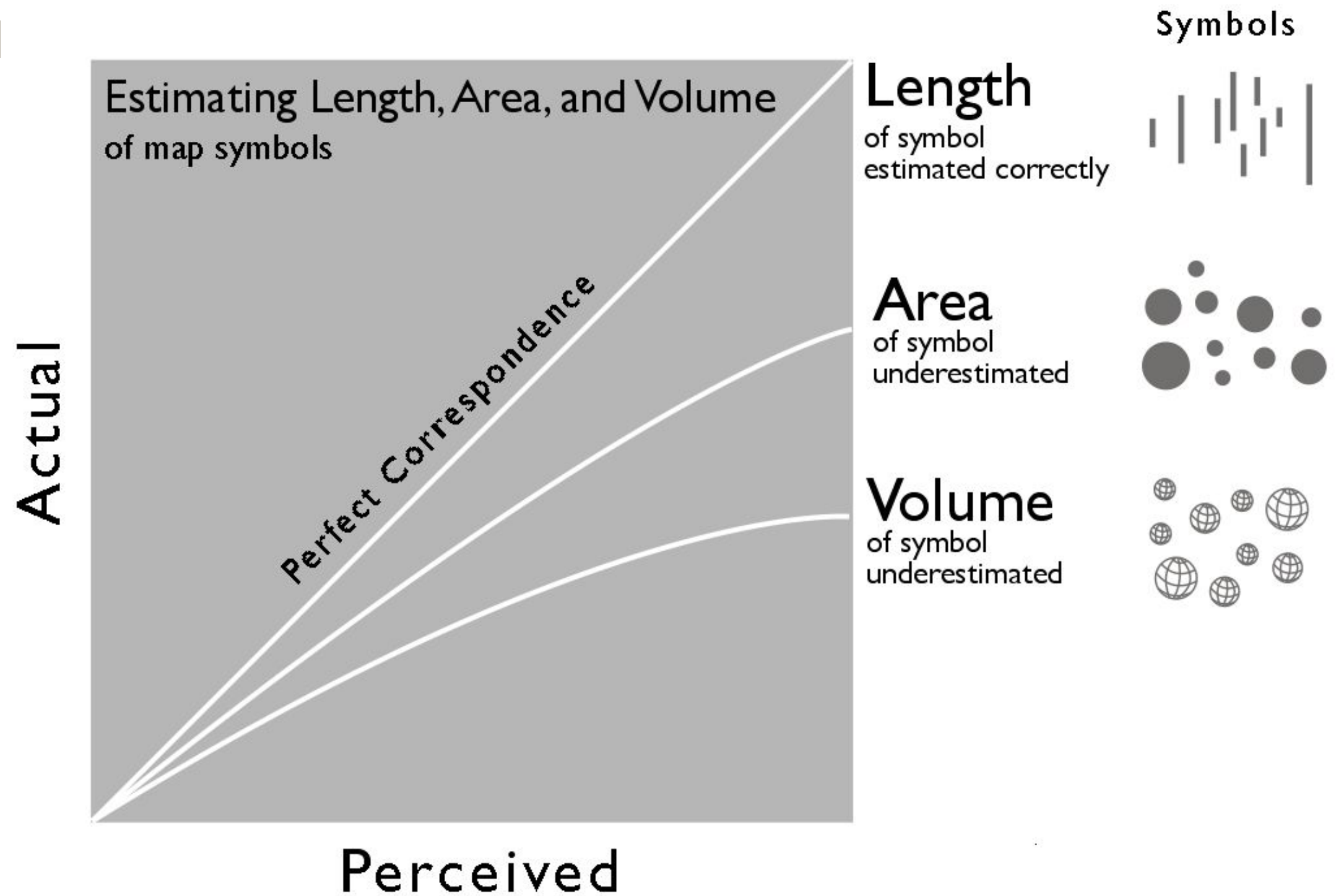


Radio = 2 cm
Area = $2 \cdot 2 \cdot \pi$



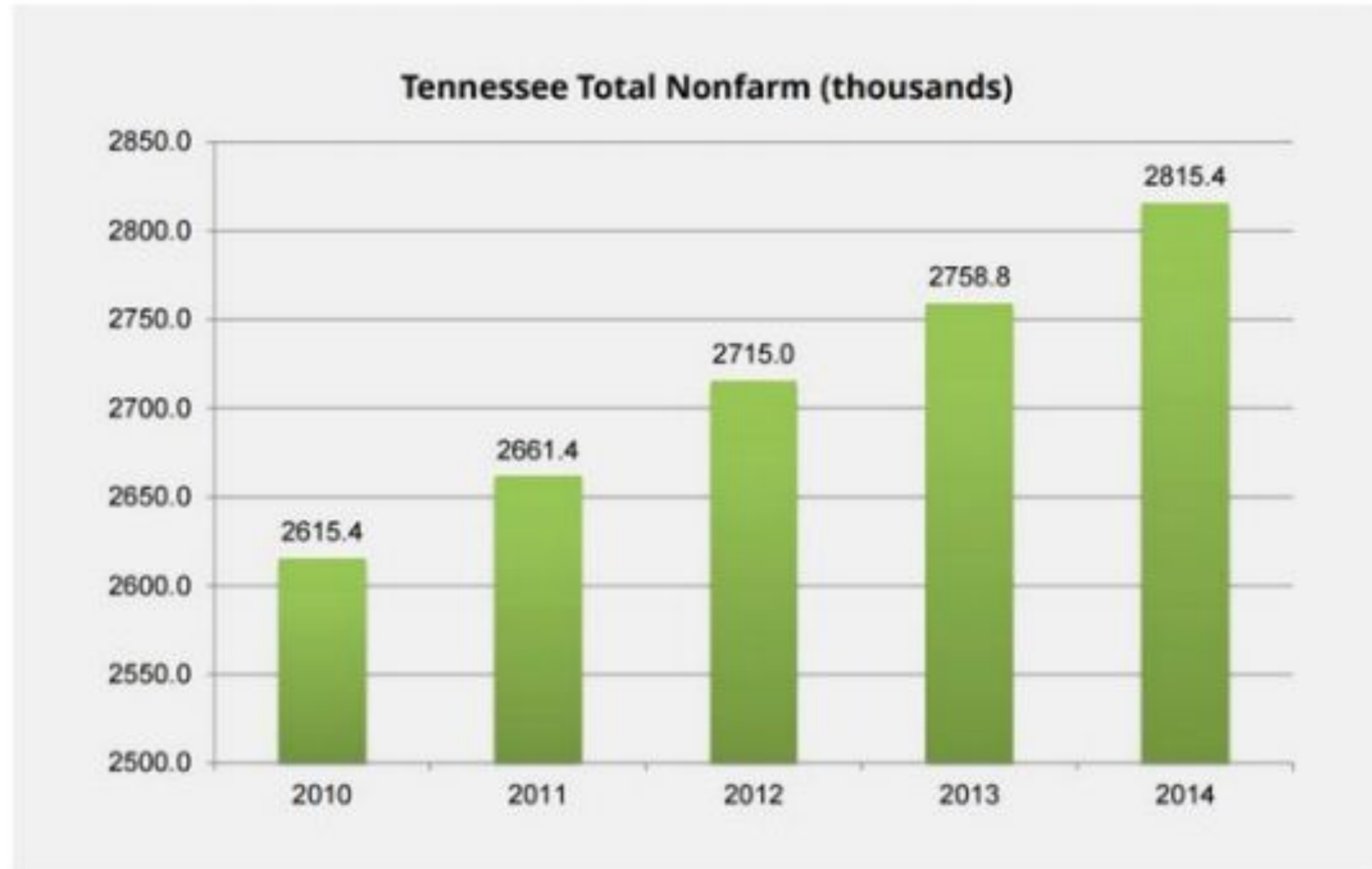
Bonus - Percepción

Estimación de magnitud



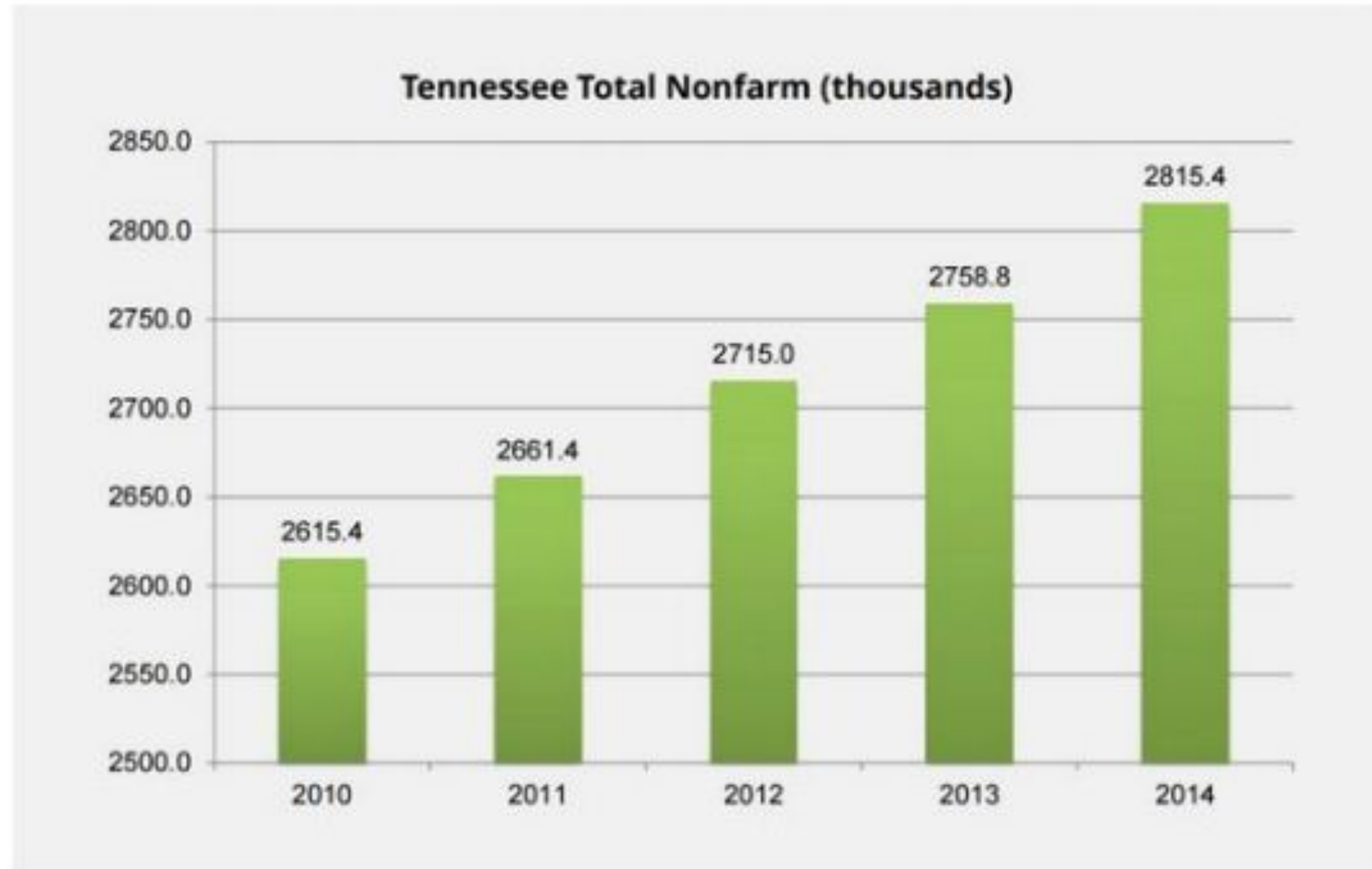
Rules of Thumbs

Lie factor → Ejes engañosos



Rules of Thumbs

Lie factor → Ejes engañosos

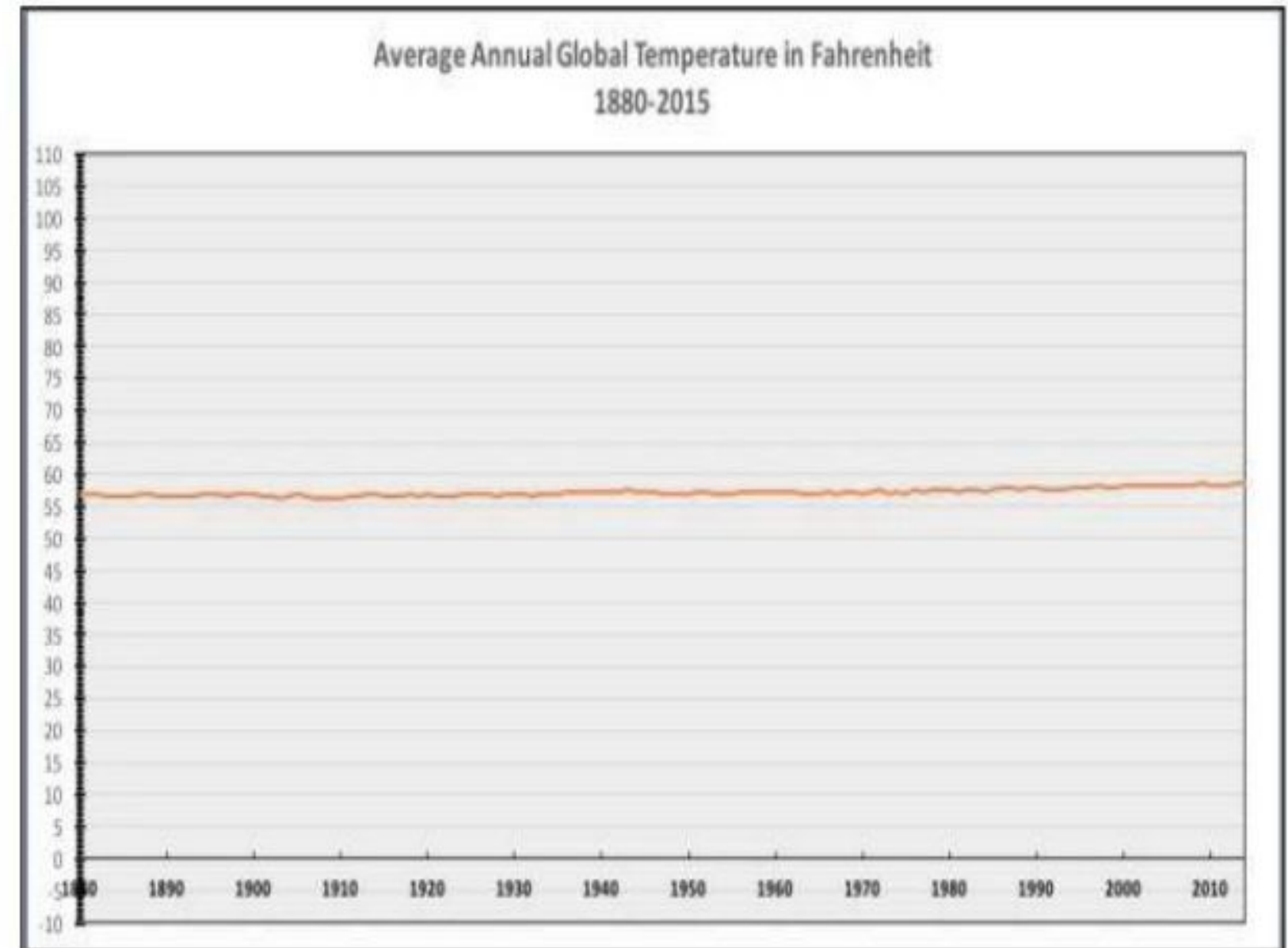


Rules of Thumbs

Lie factor → Ejes engañosos

Incluso, mostrar el cero en el eje puede ser engañoso, ya que se intenta ocultar la tasa con la que ocurren los cambios

En el gráfico, donde se quiere mostrar los cambios de temperatura por año, a simple vista no se alcanzan a apreciar los cambios



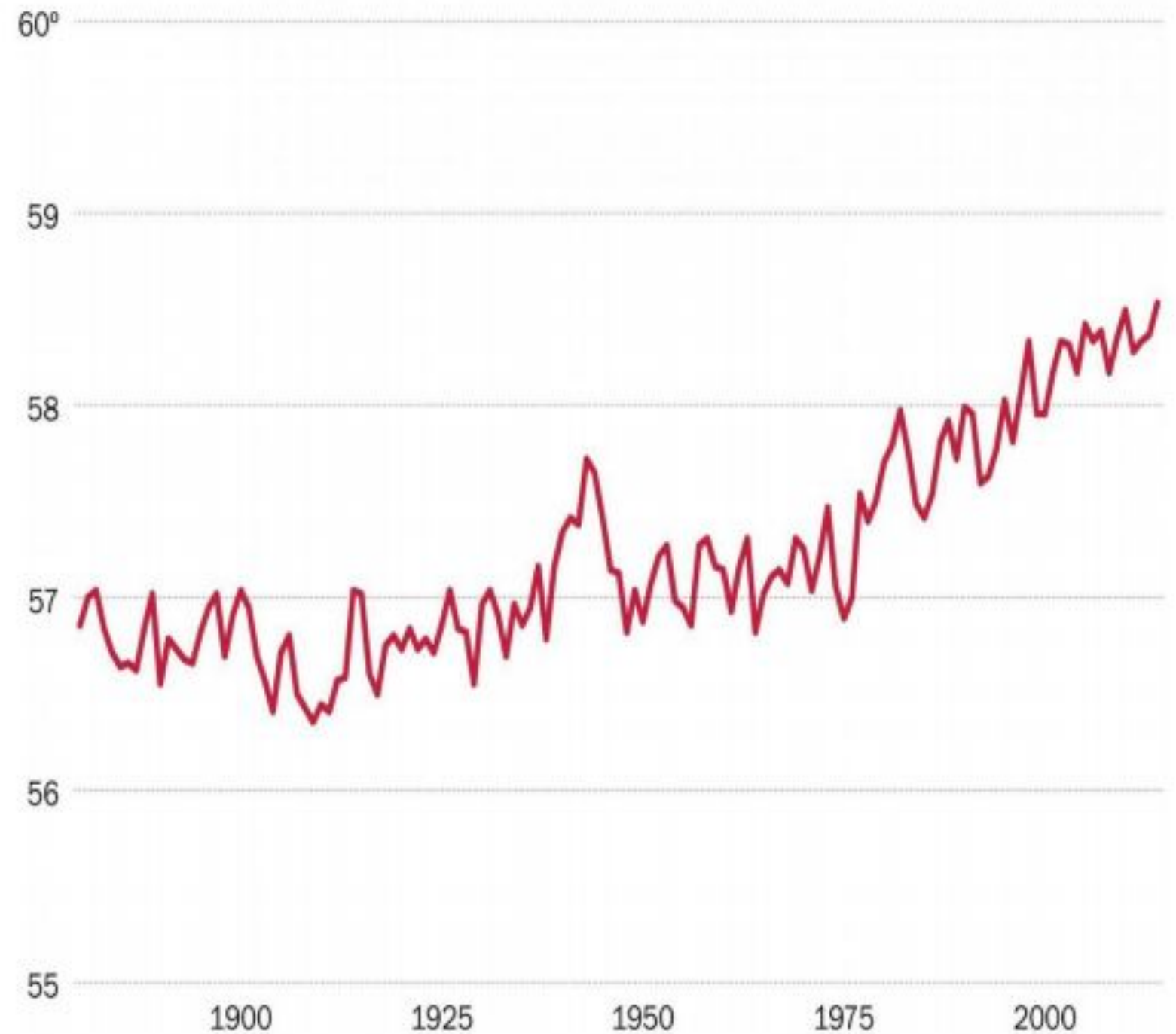
Rules of Thumbs

Lie factor → Ejes engañosos

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 que el cambio

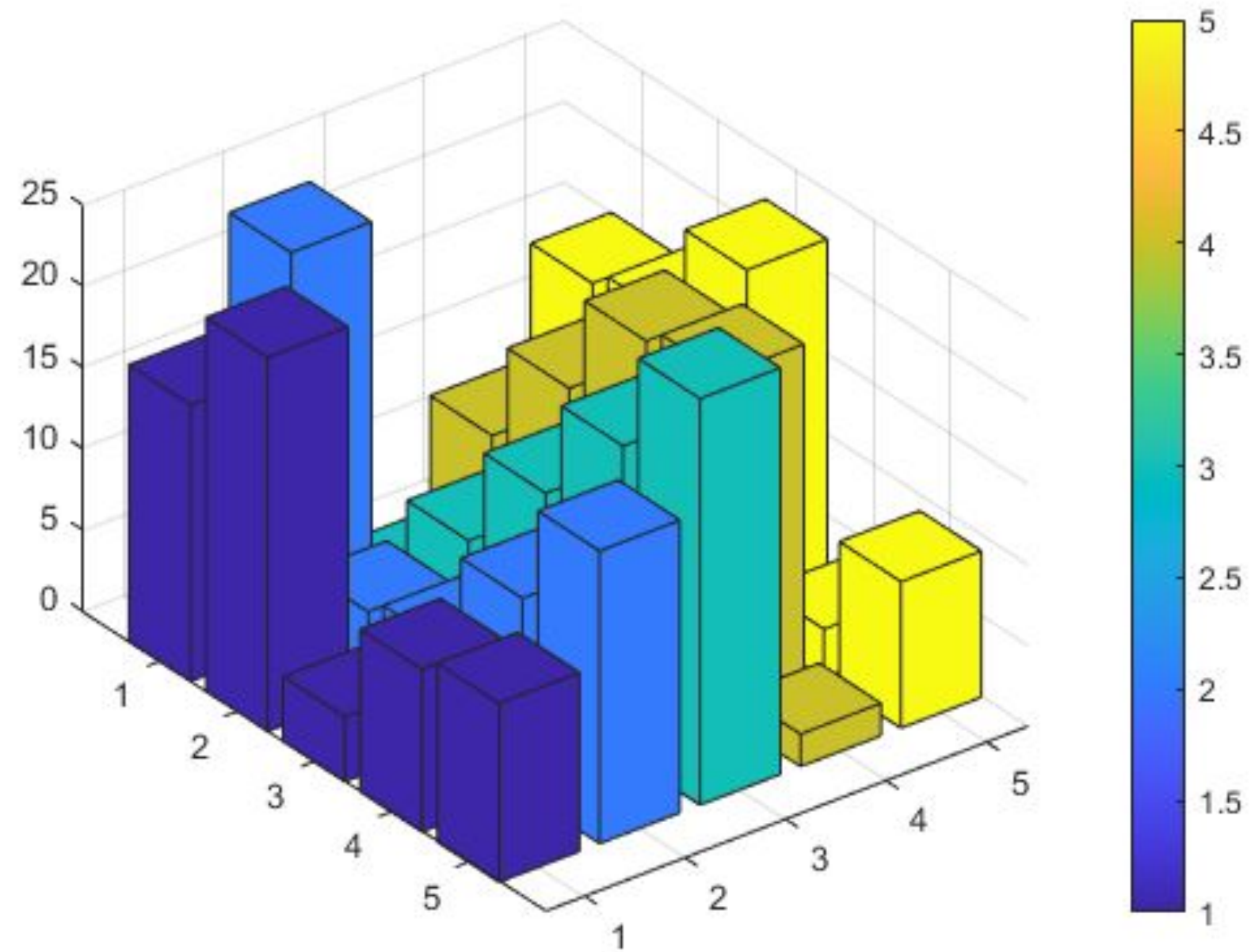
Average global temperature by year

Data from NASA/GISS.



Rules of Thumbs

No al 3D injustificado - La oclusión es un gran problema.



Rules of Thumbs

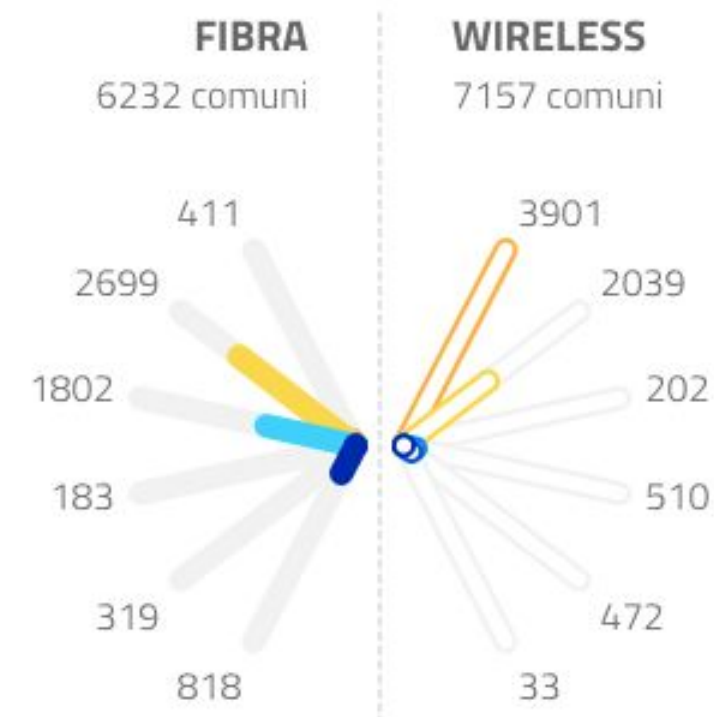
La oclusión igual puede ocurrir en 2D

Progreso del trabajo en Italia

Municipios de la región divididos según el avance de las obras:

- En diseño final
 - En planificación ejecutiva
 - ✓ Corriendo
 - Obras cerradas
 - En prueba
 - Finalizado
-
- Conexión de fibra
 - Conexión inalámbrica

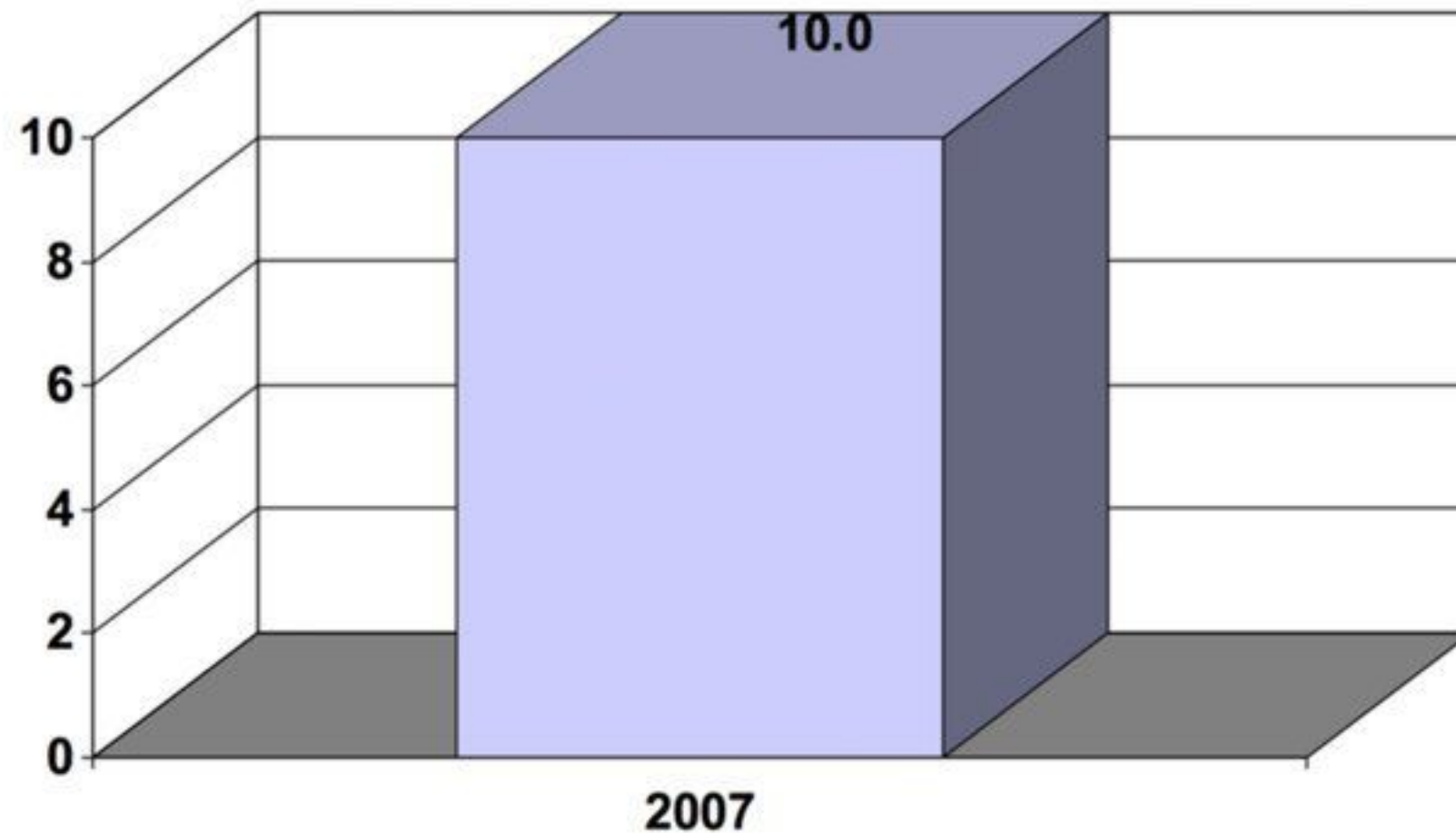
Estado actual del trabajo en los municipios



Rules of Thumbs

No al 2D injustificado (y tambien 3D)

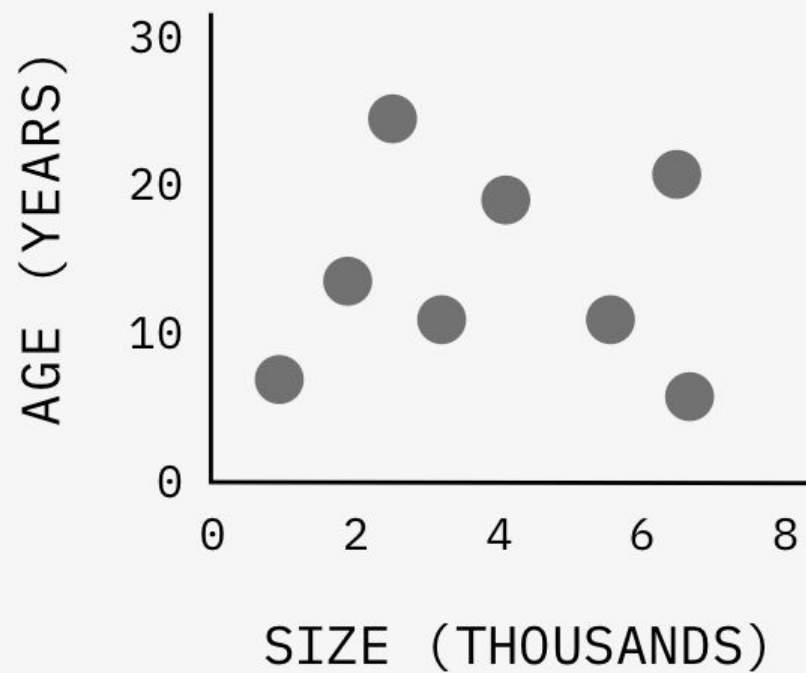
**Average Number of New Automatic Guided Vehicles
Retrofit Within Existing System – Year 2007**



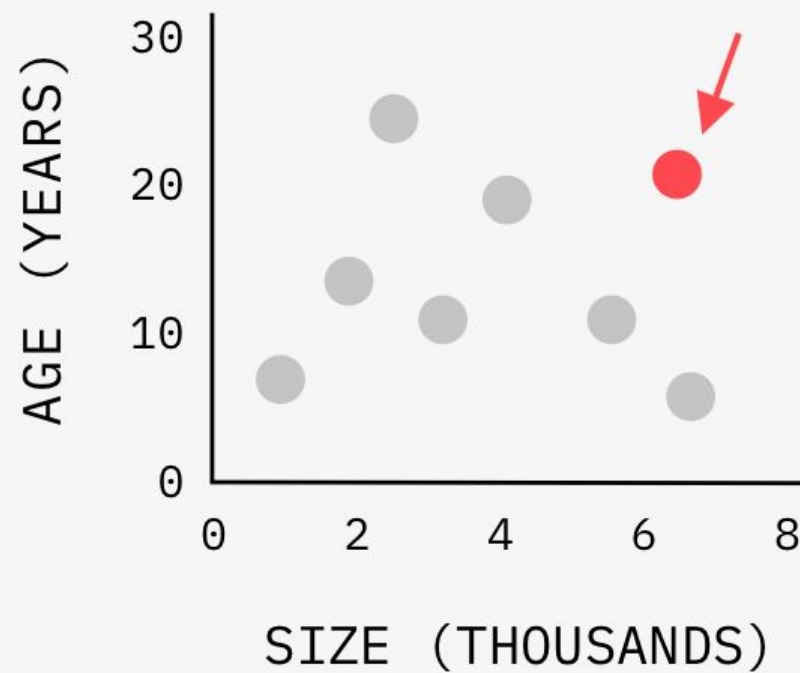
Rules of Thumbs

“Overview first, details on demand” (Shneiderman's mantra)

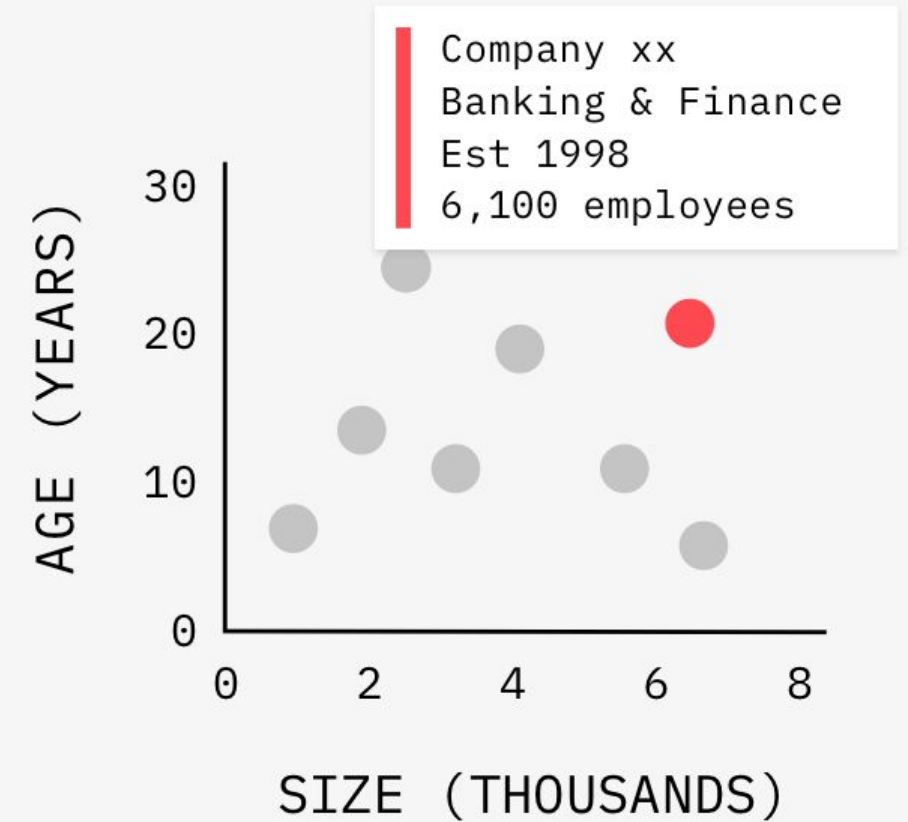
Example: companies by age and size



1. OVERVIEW FIRST



2. ZOOM AND FILTER

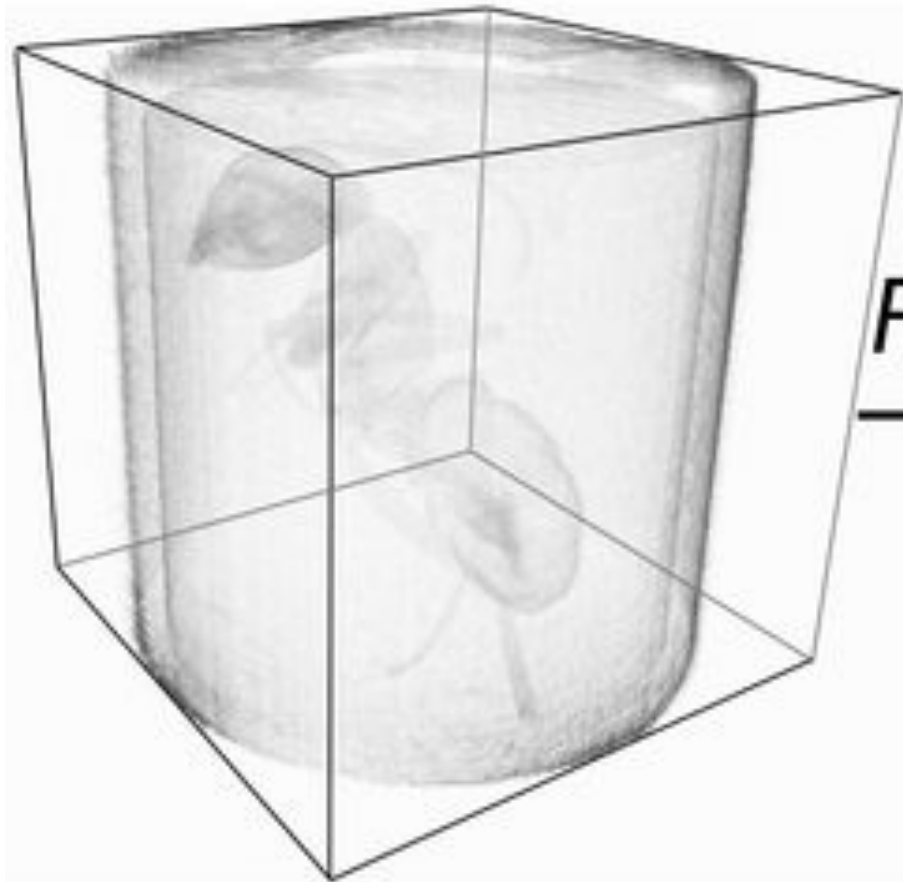


3. DETAILS ON DEMAND

Rules of Thumbs

“Overview first, details on demand” (Shneiderman's mantra)

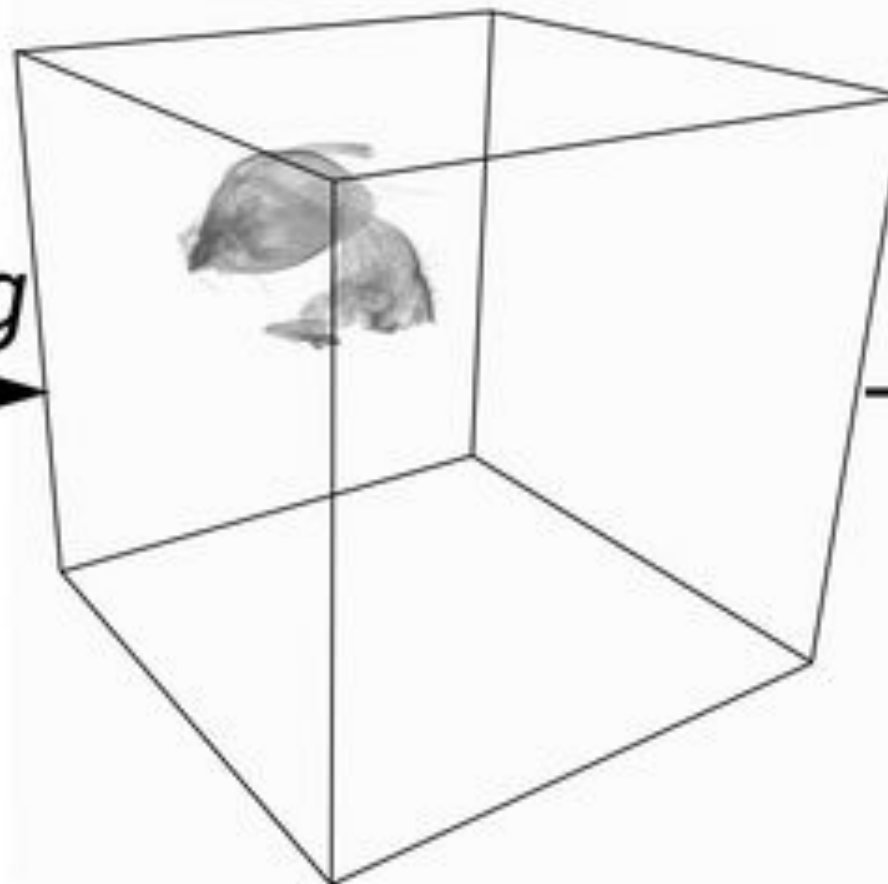
Overview



Filtering



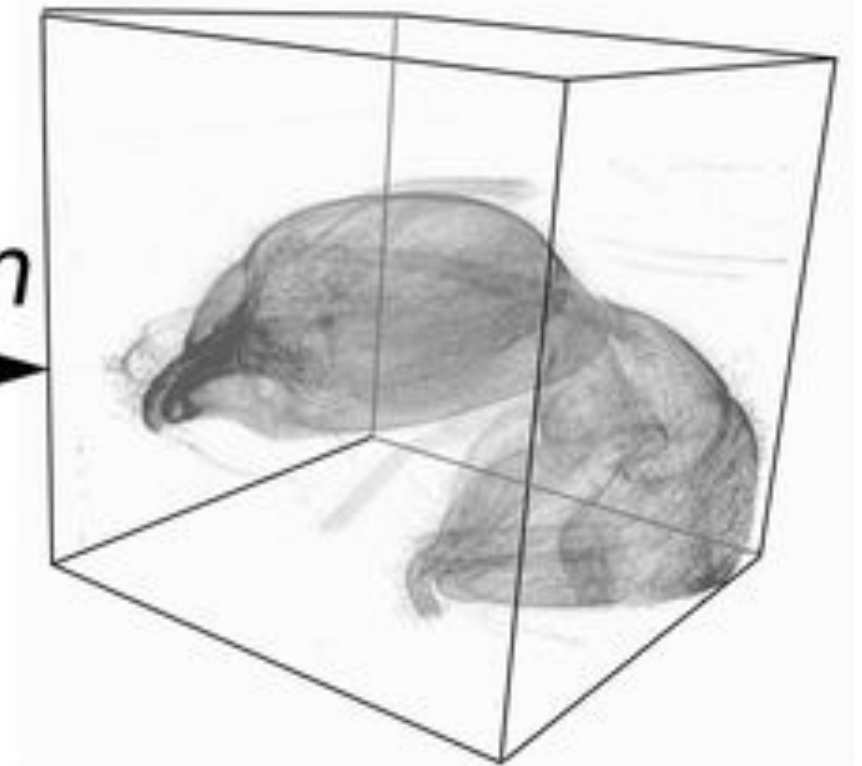
Selection



Zoom



Details



Muchas más Rules of Thumbs

- Primero el **fondo**, luego la forma.
- *Responsiveness is required.*
- Ser consistente.
- Datos mostrados en un **contexto**.
- Elección de colores a prueba de **daltonicos**.
- Elección de **tipografía**.

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

[Fuente](#)

Otras referencias

- *"The Visual Display of Quantitative Information"* — Edward Tufte
- *"Visualization, Analysis and Design"* — Tamara Munzner
- [WTFViz](#)

Contenidos

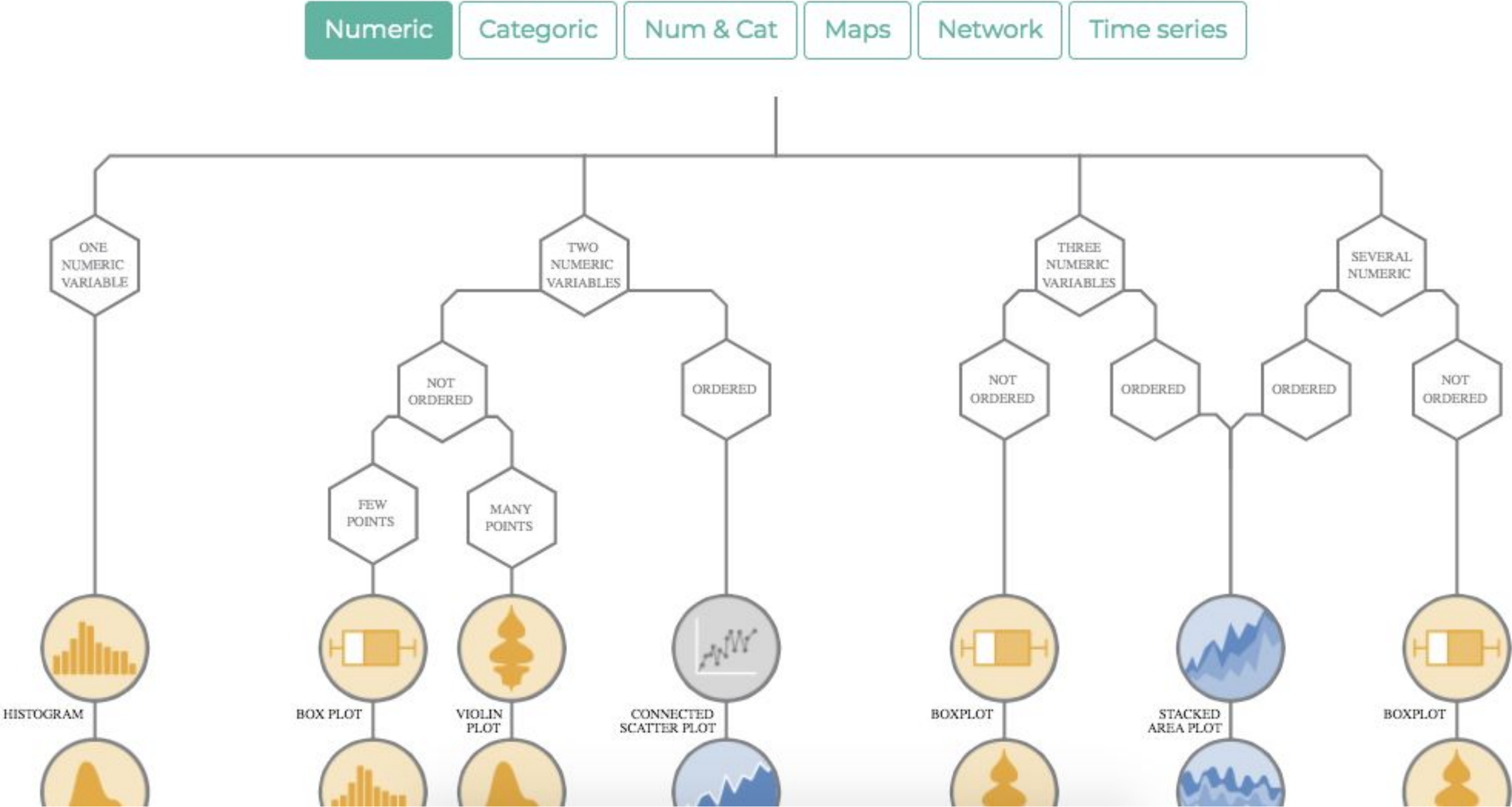
- ¿Qué?
- ¿Por qué?
- ¿Cómo diseñar una visualización? - *Rules of Thumb*
- **Framework de visualización**
 - **Etapas que componen el *framework***
 - Marcas y canales
 - Efectividad de canales
- IIC2026 - Visualización de Información

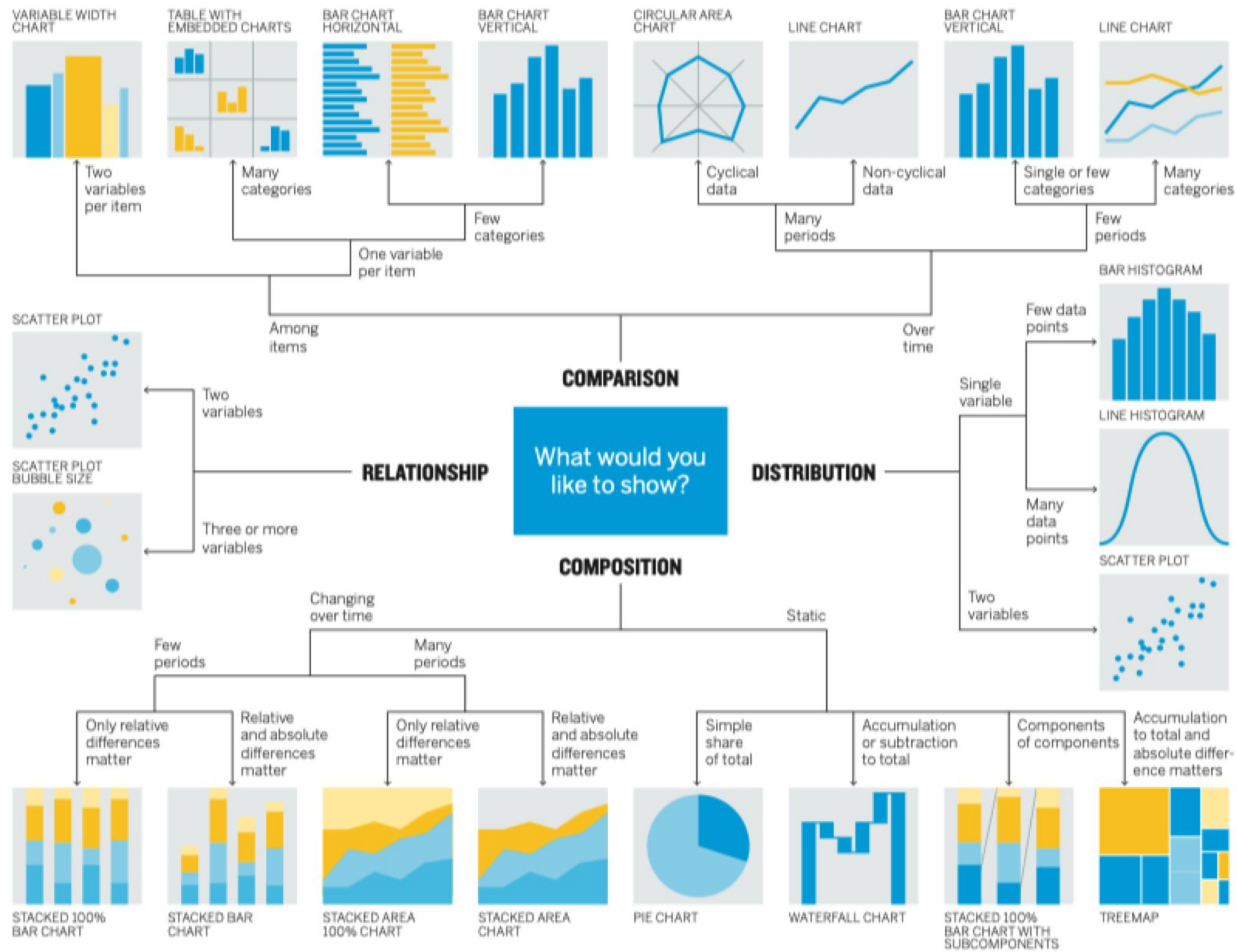
Etapas que componen el framework

- **What** → ¿Qué datos se van a visualizar?
- **Why** → ¿Por qué/para qué usar la visualización?
- **How** → ¿Cómo se verán las visualizaciones?

Imagen con los elementos por etapa

https://drive.google.com/file/d/1XxY3mmg2_LvXvkT-2GTtY5UHU9l3079j/view?usp=sharing





Etapas que componen el framework



Domain situation

Observe target users using existing tools

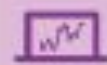


Data/task abstraction



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

Contenidos

- ¿Qué?
- ¿Por qué?
- ¿Cómo diseñar una visualización? - *Rules of Thumb*
- **Framework de visualización**
 - Etapas que componen el *framework*
 - **Marcas y canales**
 - Efectividad de canales
- IIC2026 - Visualización de Información

Marcas y canales

Marcas

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

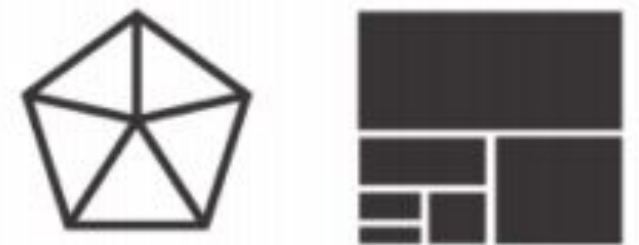
➔ Points



➔ Lines



➔ Areas



Marcas y canales

Canal

- Permite controlar la apariencia de las marcas, independientemente de la dimensionalidad de este elemento primitivo

➞ Position

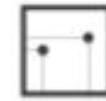
➞ Horizontal



➞ Vertical



➞ Both



➞ Color



➞ Shape



➞ Tilt



➞ Size

➞ Length



➞ Area



➞ Volume

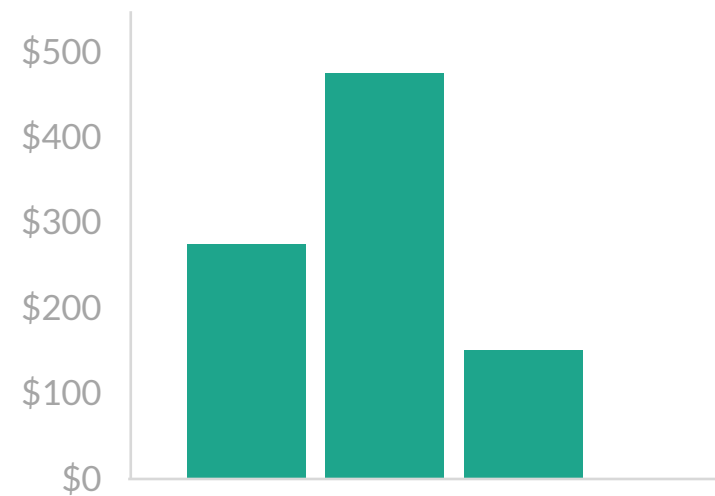


Marcas y canales

Ejemplos

Marca: línea

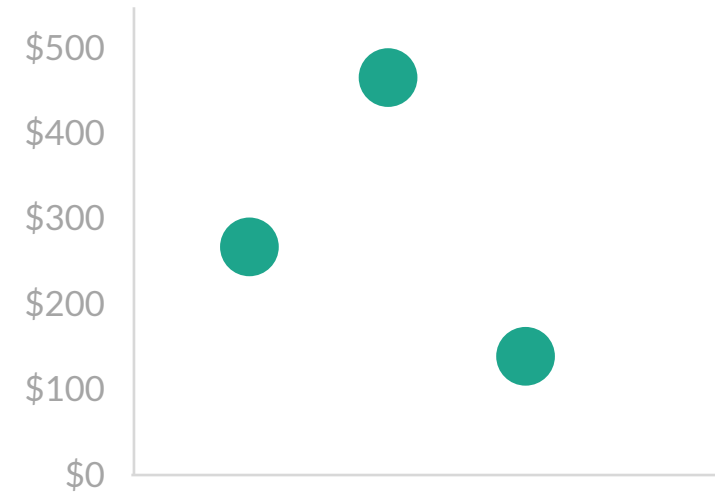
Canal: posición y largo



A

Marca: punto

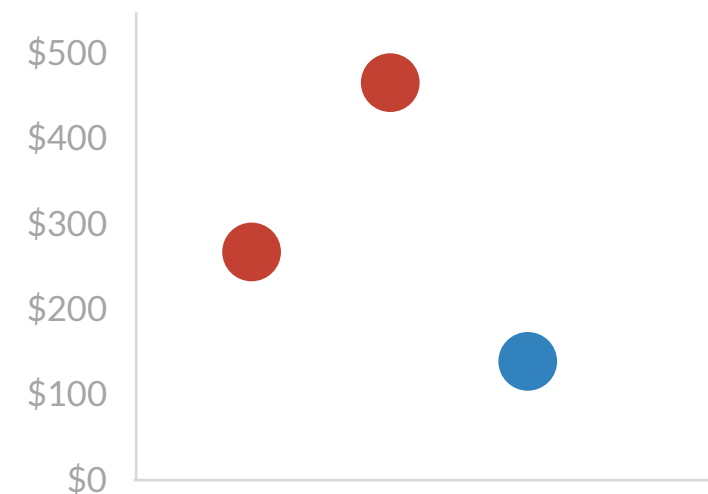
Canal: posición



B

Marca: punto

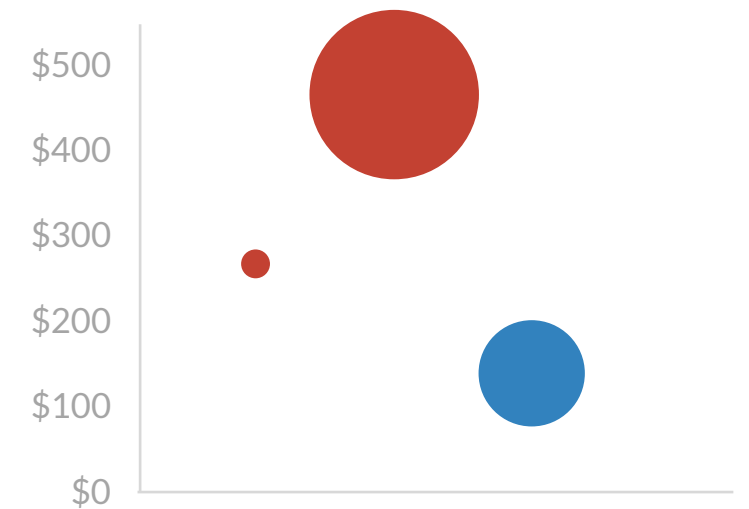
Canal: posición y color (hue)



C

Marca: punto

Canal: posición, tamaño y color (hue)



D

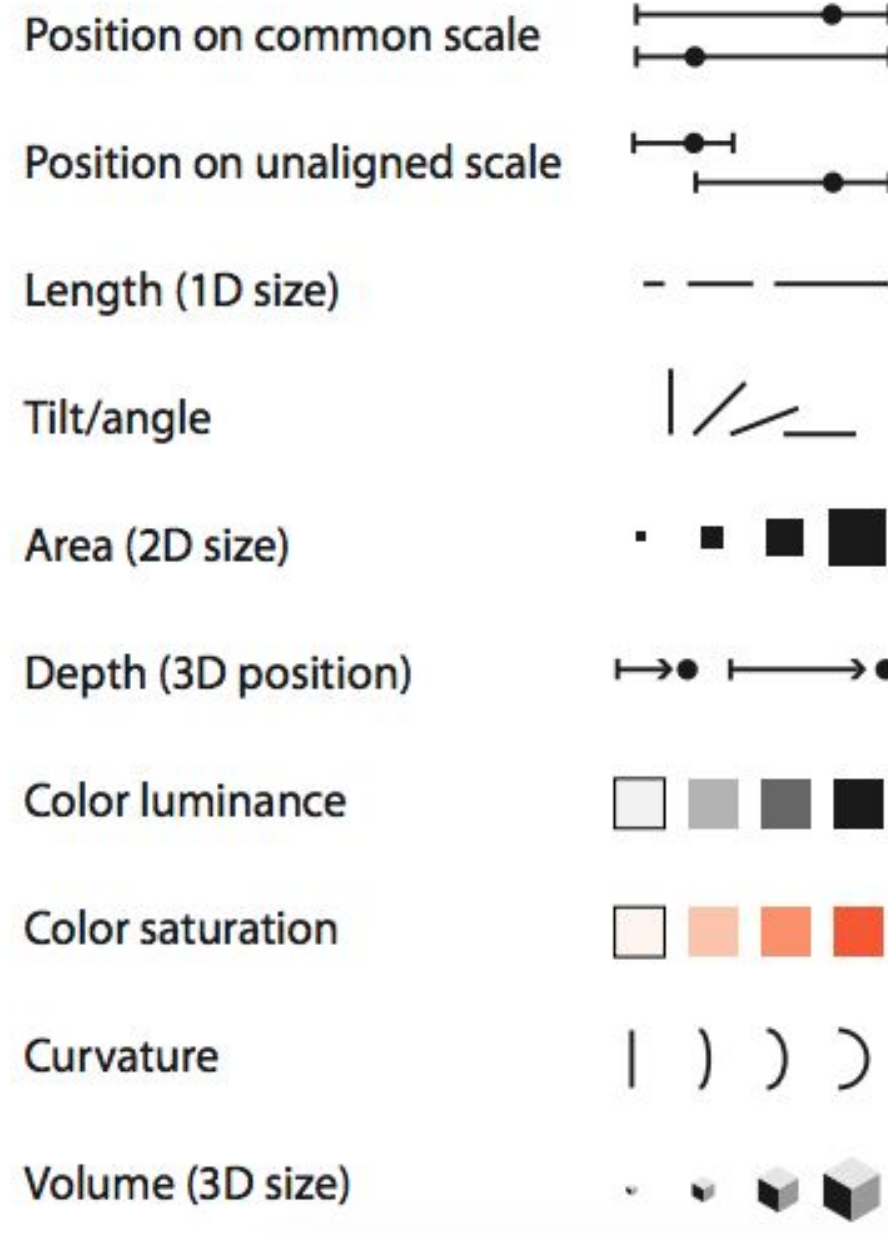
¿Qué es hue? → Un tono distinto (rojo, verde, azul, amarillo).

Contenidos

- ¿Qué?
- ¿Por qué?
- ¿Cómo diseñar una visualización? - *Rules of Thumb*
- **Framework de visualización**
 - Etapas que componen el *framework*
 - Marcas y canales
 - **Efectividad de canales**
- IIC2026 - Visualización de Información

Efectividad de canales

➔ Magnitude Channels: Ordered Attributes



➔ Identity Channels: Categorical Attributes



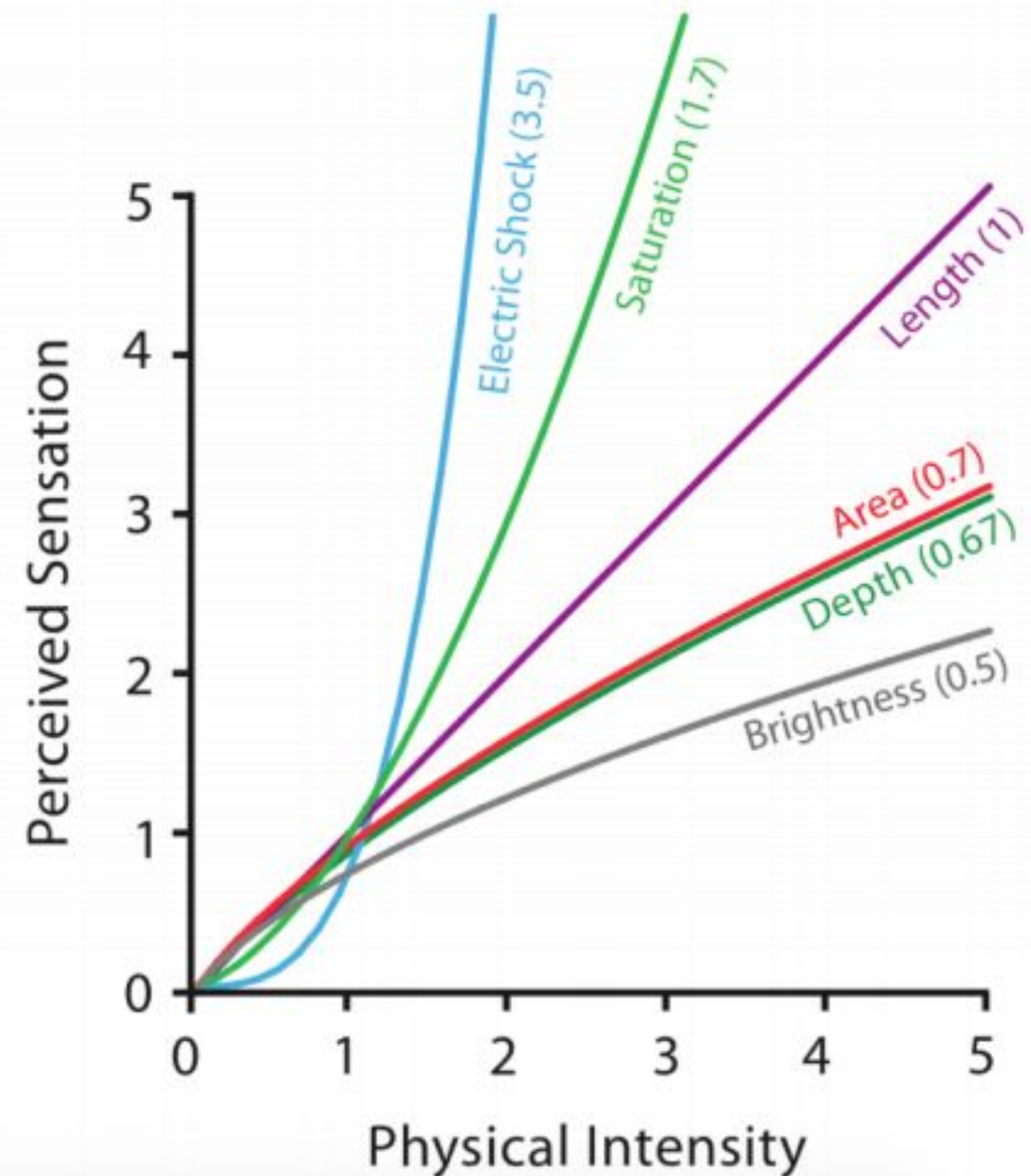
Most
Effectiveness
Least

Efectividad de canales

Accuracy (Stevens's power law)

- Modificación de un canal VS cuánto se percibe el cambio en dicho canal

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



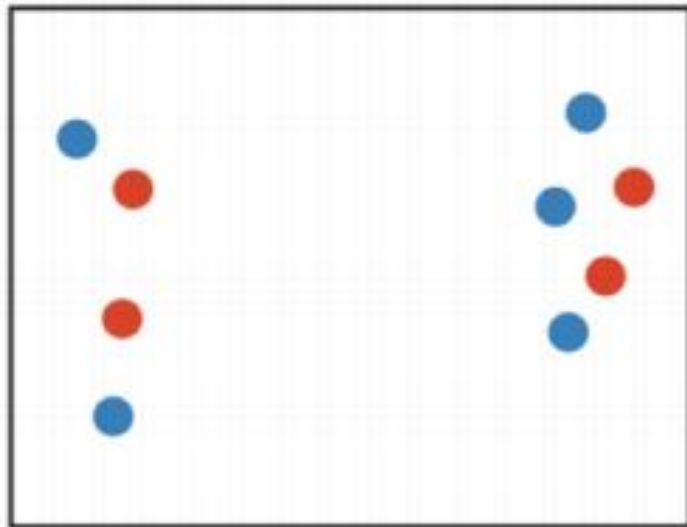


Efectividad de canales

Separability

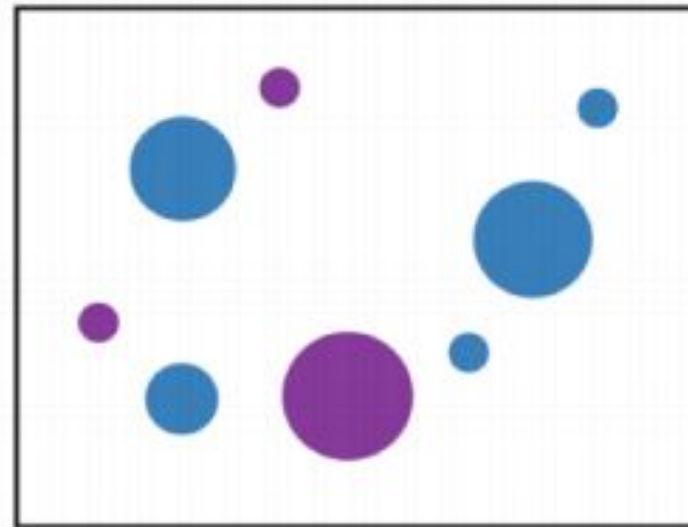
- Hay canales que interfieren en cómo percibir otro canal

Position
+ Hue (Color)



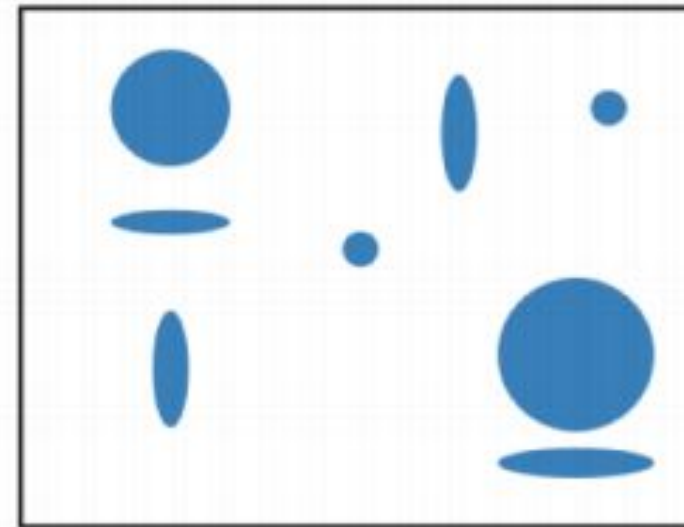
Fully separable

Size
+ Hue (Color)



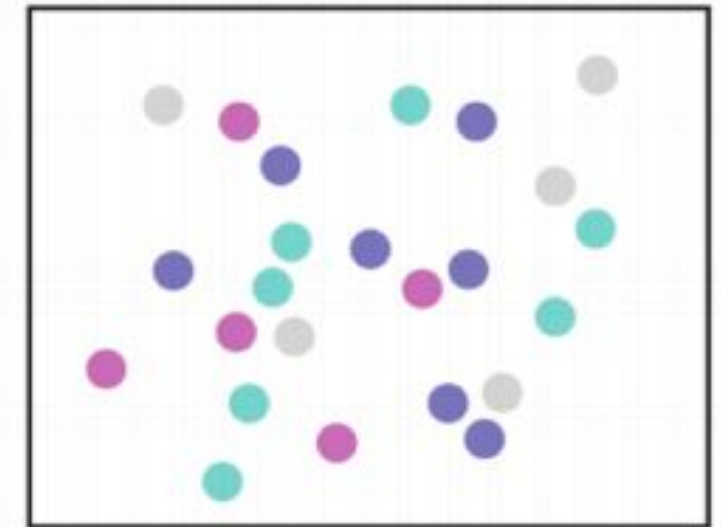
Some interference

Width
+ Height



Some/significant
interference

Red
+ Green



Major interference

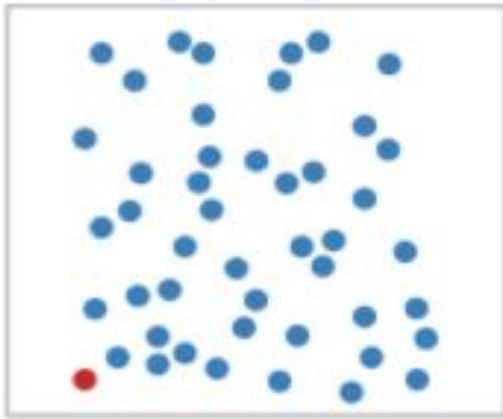
Efectividad de canales

Visual popout

- Tiempo que nos toma encontrar el objeto diferente frente a distractores



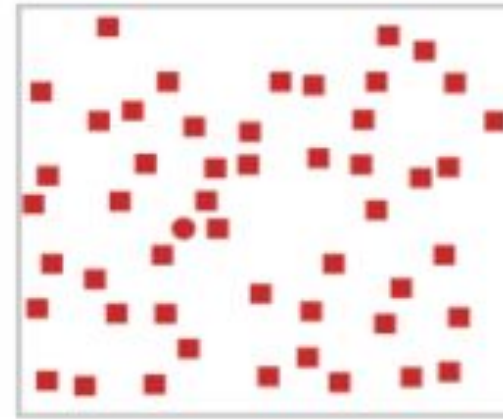
(a)



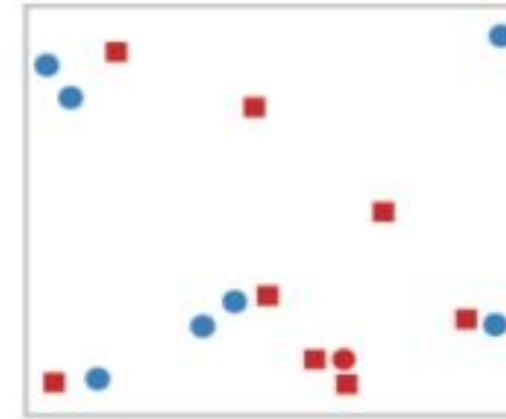
(b)



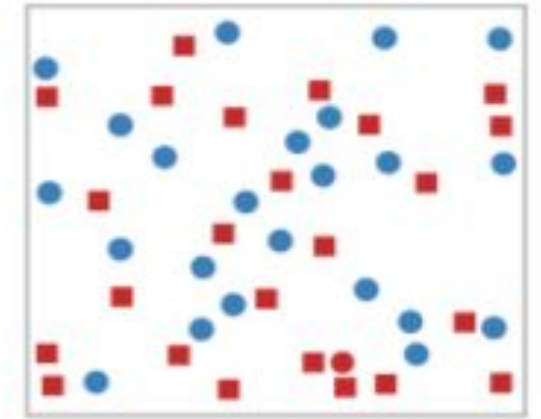
(c)



(d)



(e)

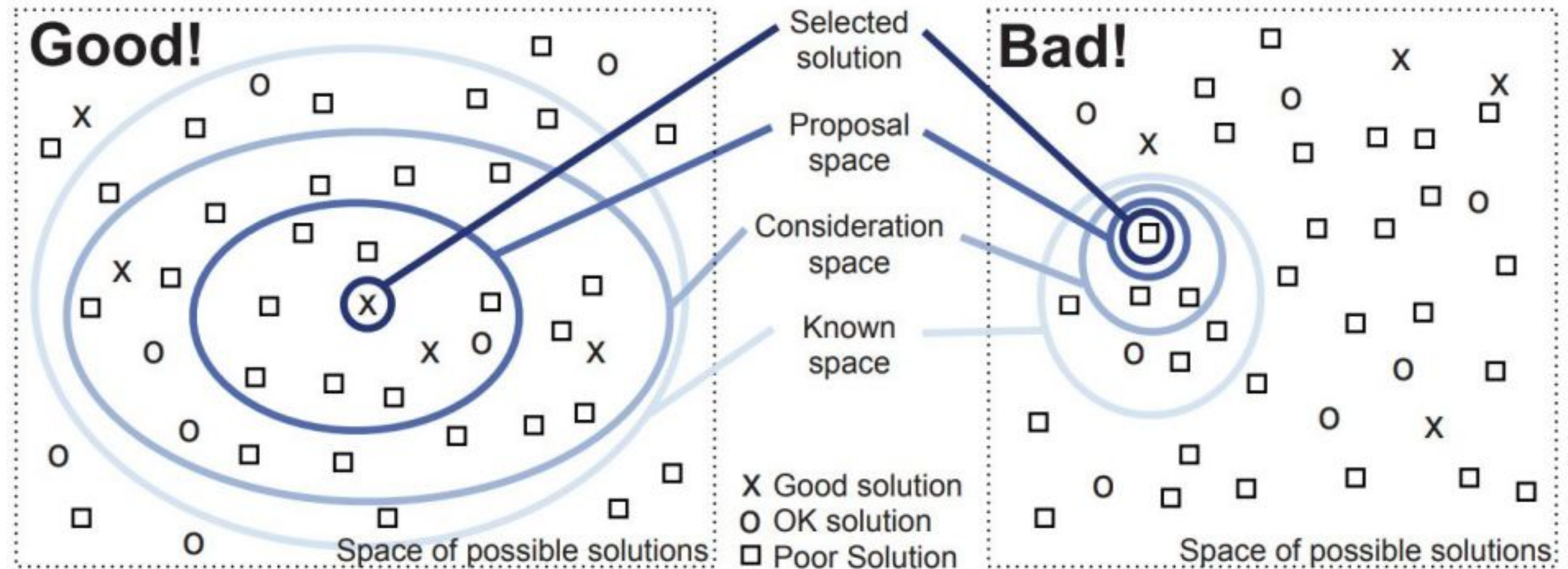


(f)

Contenidos

- ¿Qué?
- ¿Por qué?
- ¿Cómo diseñar una visualización? - *Rules of Thumb*
- *Framework* de visualización
 - Etapas que componen el *framework*
 - Marcas y canales
 - Efectividad de canales
- IIC2026 - Visualización de Información

IIC2026 - Visualización de información



Tamara Munzner

Visualization: Analysis & Design

IIC2026 - Visualización de información

- Es una introducción al aprendizaje de los principios de diseño gráfico y de técnicas interactivas para **visualizar datos**.
- Es un curso multidisciplinario que incorpora subcampos de la **computación**, la **estadística**, del **diseño gráfico** (e.g. teoría de color), 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.
- Página de la última versión: <https://puc-infovis.github.io/version-2021/>

IIC2026 - ¿Qué puedo lograr hacer?

Proyecto Visualización 2017 - Anime

Visualización de toma de cursos Banner UC

Proyecto Sustentabilidad

Mucho más

IIC1005 - Computación: Ciencia y Tecnología del Mundo Digital

Visualización de información

[Hernán Felipe Valdivieso López](#)

hfvaldivieso@uc.cl
