



Teoría de la Visualización de datos

Dra. Mariana Esther Martínez Sánchez



¿Qué es la visualización de datos?

Mapear valores a visuales o volver números en imágenes e historias.

La información visual es más fácil de procesar.

Es un arte y una ciencia. La ciencia está en el proceso de generar la visualización.

Kerzner, E., Goodwin, S., Dykes, J., Jones, S., & Meyer, M. (2018). A framework for creative visualization-opportunities workshops. IEEE transactions on visualization and computer graphics, 25(1), 748-758. DOI: 10.1109/TVCG.2018.2865241

**Your greatest
insight is only as
good as your ability
to communicate it**

Proceso de la ciencia de datos



Fry, B. J. (2004). *Computational information design* (Doctoral dissertation, Massachusetts Institute of Technology).

Tipos de visualización en ciencia de datos

Exploratoria

- Relación entre analista y los datos
- Entender los datos
- Entender contexto de los datos
- Conectar y relacionar
- Hipótesis alternativas
- Obtener conocimiento
- Evitar sesgos

Explicativa

- Relación entre analista y el público
- Comunicar un mensaje
- Entender el contexto de la audiencia
- Seleccionar herramientas visuales
- Eliminar desorden visual
- Dirigir atención al conocimiento obtenido
- Crear historias

Cuarteto de Anscombe

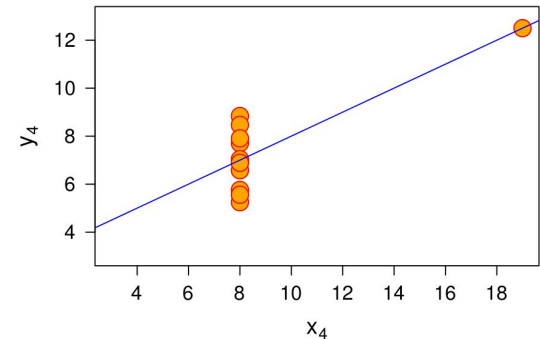
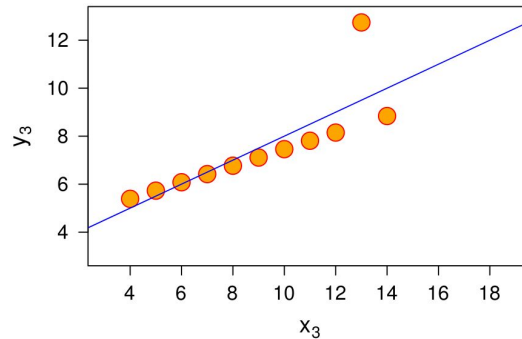
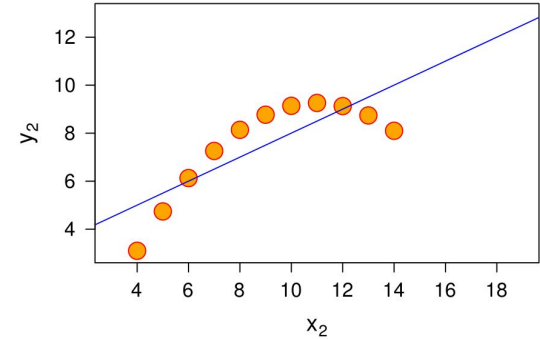
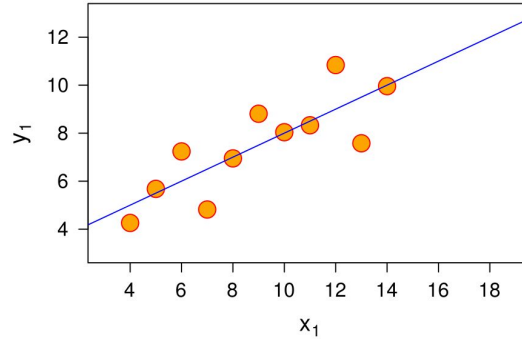
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	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

Los cuatro conjuntos de datos tienen las mismas medidas estadísticas

Propiedad	Valor
Media x	9.0
Varianza x	11.0
Media y	7.5
Varianza y	4.12
Correlación x e y	0.816
Recta de regresión	$y = 3 + 0.5 x$

Cuarteto de Anscombe

Edward Tufte usó el cuarteto en la primera página del primer capítulo de su libro *The Visual Display of Quantitative Information*, para enfatizar la importancia de mirar los datos antes de analizarlos.



Patrones, patrones, patrones

El cerebro es
excepcional en
reconocer
patrones...

...para bien y
para mal.





Gráficas



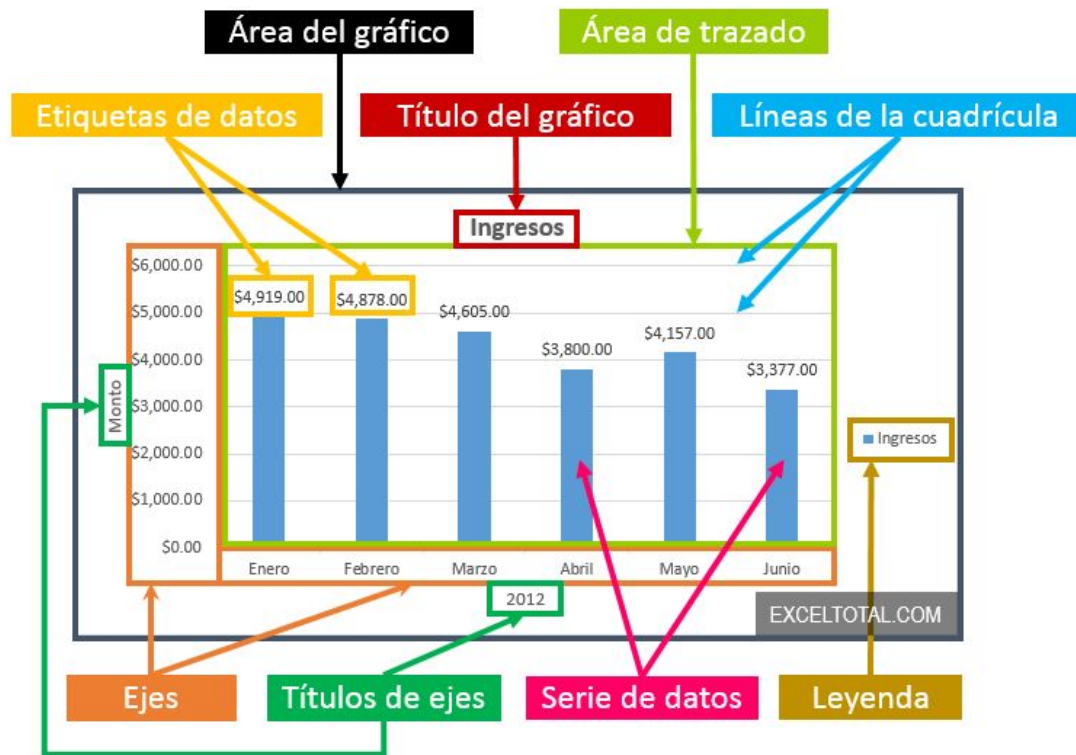
Gráficas

Es un patrón común o convención para combinar codificaciones visuales o símbolos

- + codificación visual
- + tipos de datos
- + relaciones

Considerar:

- tipo de análisis
- mensaje a transmitir
- facilidad de generación
- facilidad de interpretación
- familiaridad del público con el gráfico



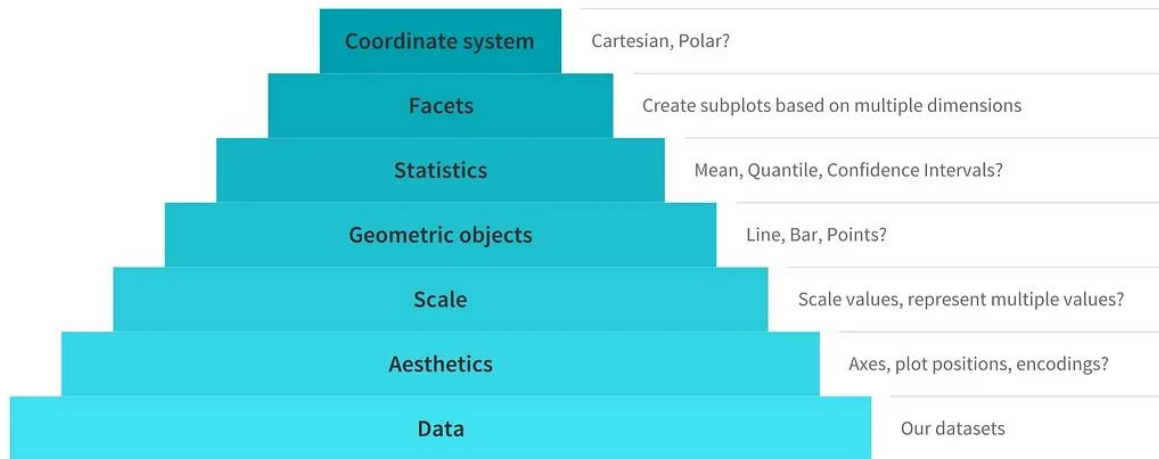
Gramática de gráficas

Un marco conceptual para crear y visualizar gráficos de datos, desarrollado por Leland Wilkinson en 1999.

Principios clave:

- Separación de datos de estética
- Definición de elementos comunes de gráficos
- Composición de estos elementos comunes.

Major Components of the Grammar of Graphics



Gráficas

Gráfica: codificación visual + tipos de datos + relaciones

Es un patrón común o convención para combinar codificaciones visuales o símbolos

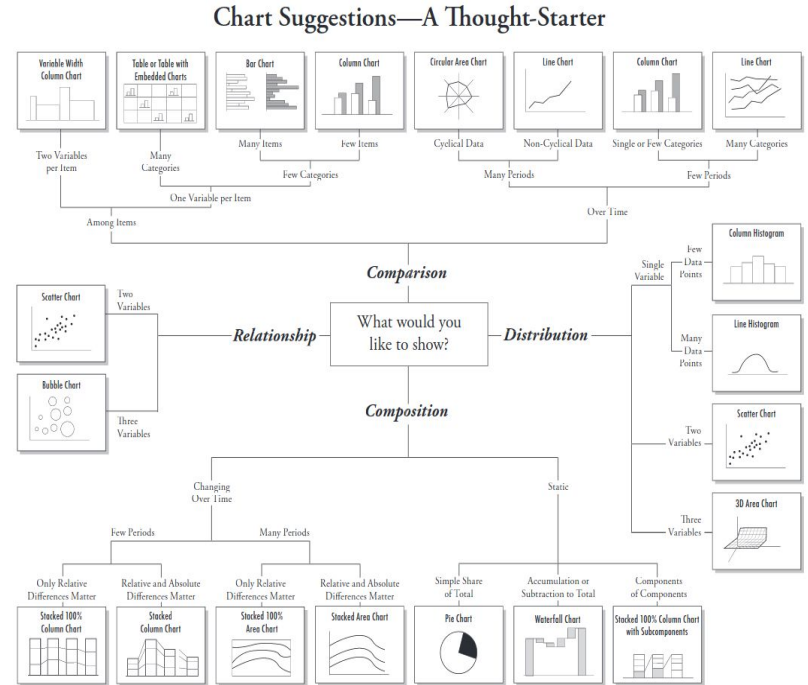
Considerar:

- tipo de análisis
- mensaje a transmitir
- facilidad de generación
- facilidad de interpretación
- familiaridad del público con el gráfico

Interactive chart chooser:

<https://www.juiceanalytics.com/chartchooser>

Few, S. (2004). Eenie, meenie, minie, moe: selecting the right graph for your message. Intelligent Expertise. Perceptual Edge.



<https://extremepresentation.typepad.com/files/choosing-a-good-chart-09.pdf>



Codificación y elementos visuales



Codificación visual

Se representa la información de manera gráfica mediante elementos visuales

Datos → Elementos visuales

Elementos visuales: colores, formas, tamaños y posiciones, etc.

La codificación depende de los datos, sus tipos y la jerarquía interna.

Generar codificación

1. Determina variables y tipo de dato
2. Jerarquiza variables
3. Asigna mapeo

Tipos de datos

Cuantitativos

- Se les puede asignar un número
- Ej: enteros, decimales, porcentajes, proporciones

Cualitativos

- Ej: categóricos, ordenados

Especiales

- Tienen una estructura particular
- Ej: coordenadas, series de tiempo, secuencias genéticas

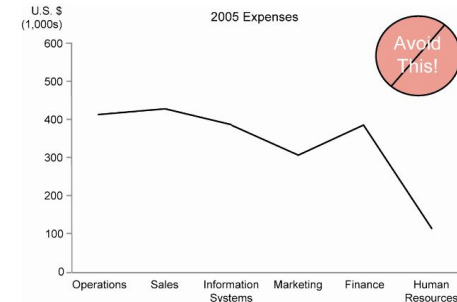
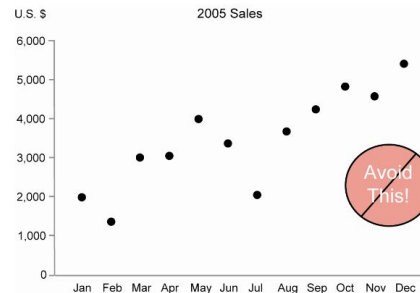
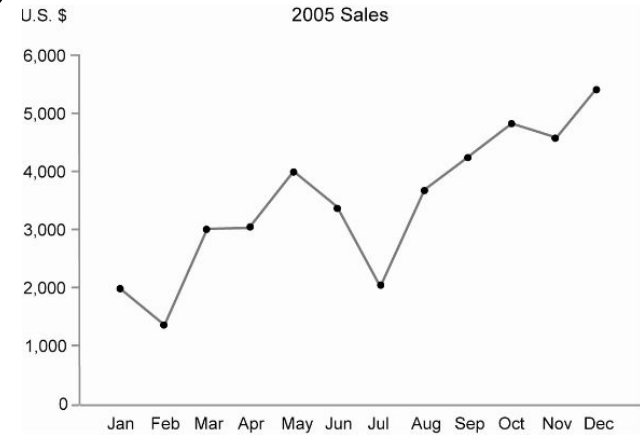
Elementos visuales posicionales

Puntos:

- dos escalas cuantitativas
- mediciones independientes

Líneas:

- una de las escalas es secuencial, por ejemplo tiempo
- resaltar forma o cambios
- líneas suaves y continuas indican proceso continuo
- usa puntos y líneas para resaltar mediciones
- NO usar para categóricos

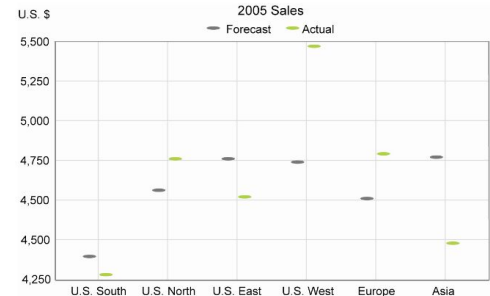
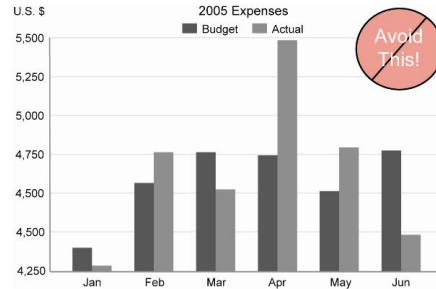
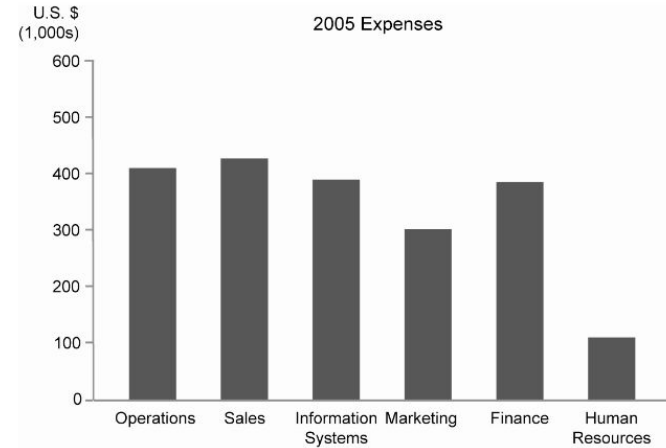


Elementos visuales posicionales

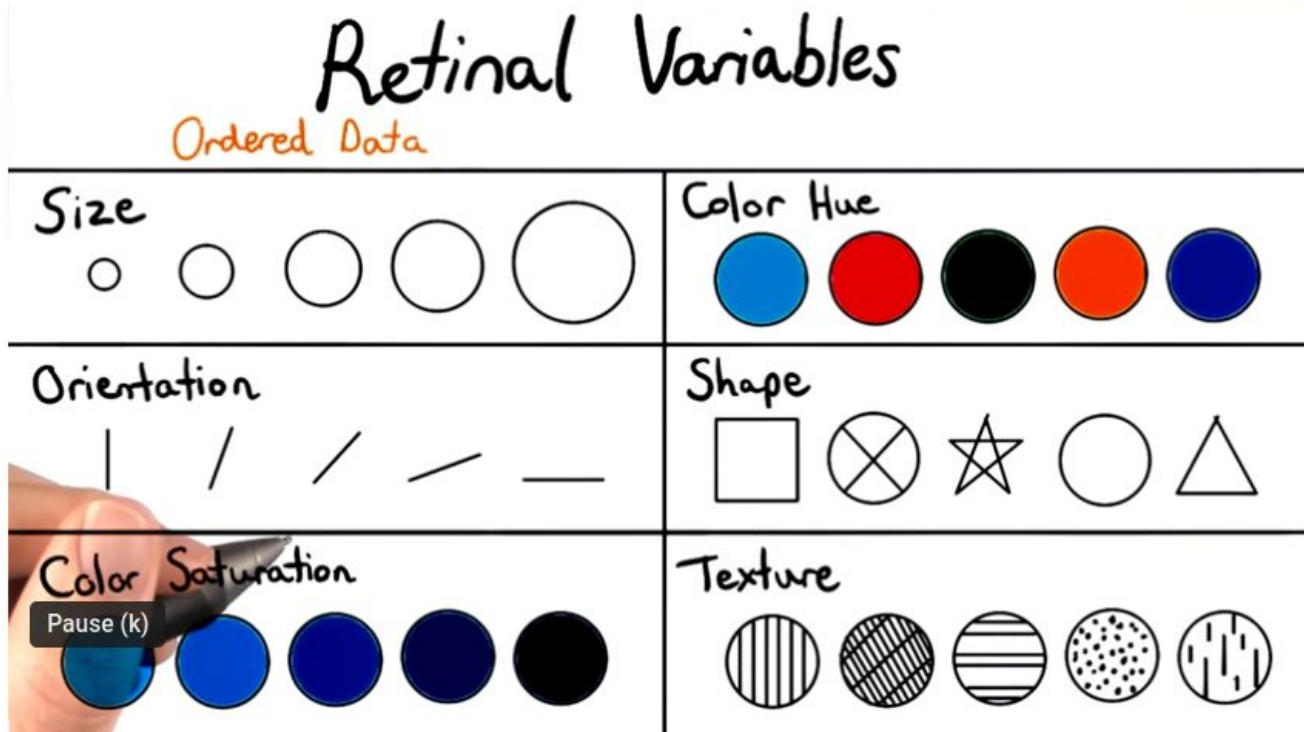
Barras:

- Representar y comparar valores en categorías
- Series de tiempo si interesa comparación entre puntos
- Siempre empieza en cero
- Si eje no empieza en cero usar puntos

Few, S. (2006). Data visualization: Rules for encoding values in graph.



Elementos visuales retinales

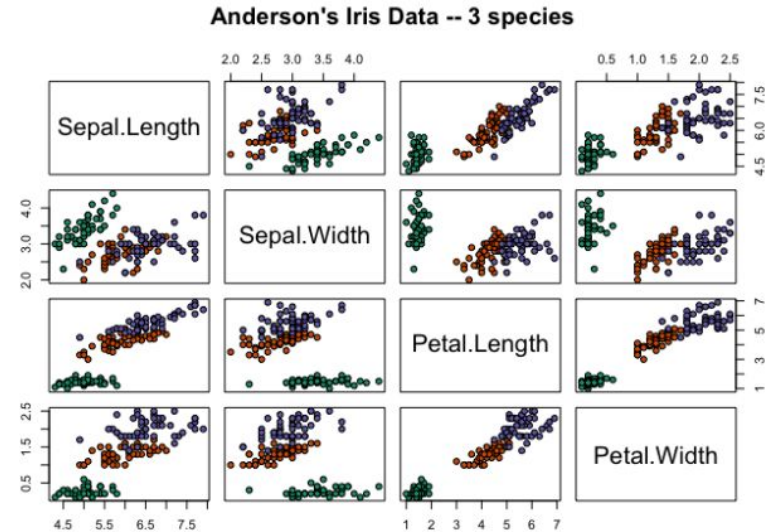


Jerarquía de la percepción de elementos visuales

Preciso

- Posición a lo largo de una escala común (gráfico de dispersión)
- Posición en escalas idénticas pero no alineadas, (múltiples gráficos de dispersión)
- Longitud (gráfico de barras)
- Ángulo y pendiente (gráfico circular)
- Área (burbujas)
- Volumen, densidad/opacidad y saturación de color (mapa de calor)
- Tono de color (mapa grises)

Inexacto



Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387), 531-554.
https://web.cs.dal.ca/~sbrooks/csci4166-6406/seminars/readings/Cleveland_GraphicalPerception_Science85.pdf

Ejercicio:

The Facebook Offering: How It Compares

[< Prev](#)[Next >](#)[1](#) **[2](#)** [3](#) [4](#) [5](#)

Find a company

Determina:

1. Variables
2. Tipo de dato
3. Importancia
4. Codificación
5. Claridad

Detalles importantes:

- Título efectivo y leyenda con mensaje principal
- Codificación doble: año y valor
- Uso de saturación

Company
value
in billions of
today's dollars

100 —

80 —

60 —

40 —

20 —

Facebook

Facebook's offer price was \$38 a share, giving the company a valuation of \$104 billion, nearly four times larger than Google in 2004.

Facebook



10 —

Year of I.P.O.

1980 1985 1990 1995 2000 2005 2010

1980 1985 1990 1995 2000 2005 2010

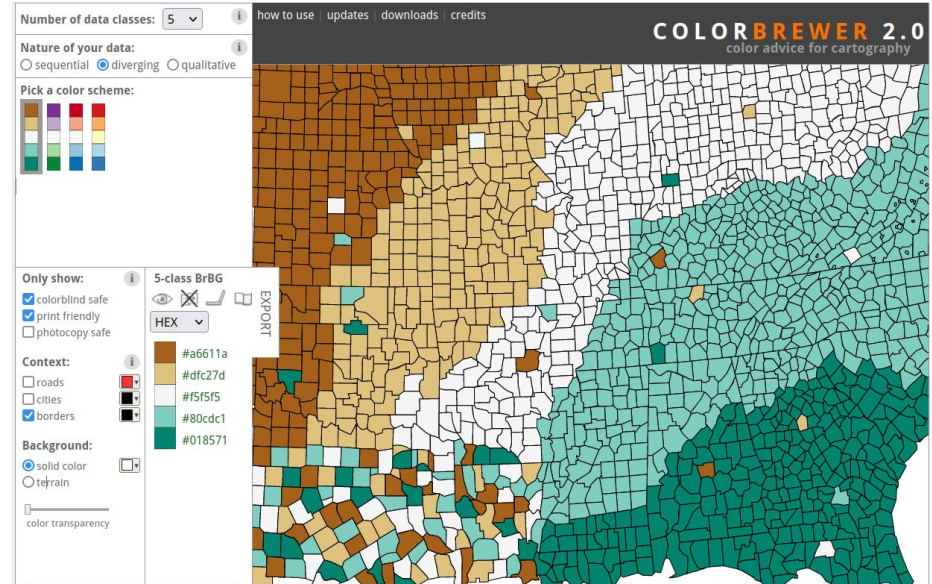


Color



Color

- Diseña en blanco y negro
- Usa colores poco intensos o desaturado
- Evita no colores sólidos #ffffff
- Usa colores vibrantes para resaltar
- Selecciona color según el tipo de variable (secuencial, divergente, cíclico, cualitativa)
- Considera el impacto emocional y asociaciones culturales
- Considera a las personas daltónicas
- Usa solo uno o dos colores (excepto que sean codificación de variable)



<https://colorbrewer2.org/#type=sequential&scheme=GnBu&n=5>

Reglas de color

Practical Rules for Using Color in Charts

Stephen Few, Perceptual Edge

Visual Business Intelligence Newsletter

February 2008

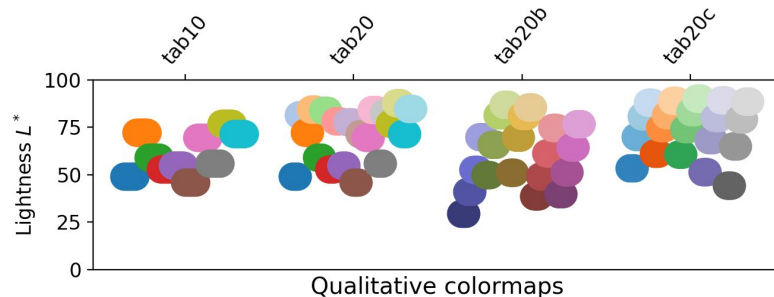
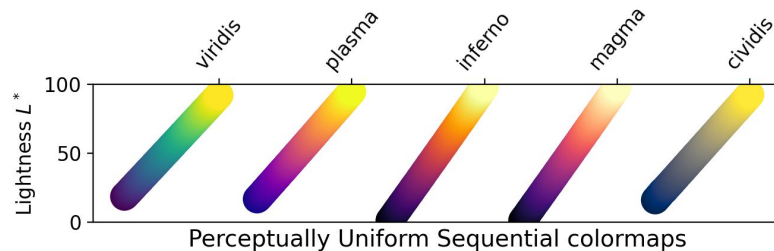
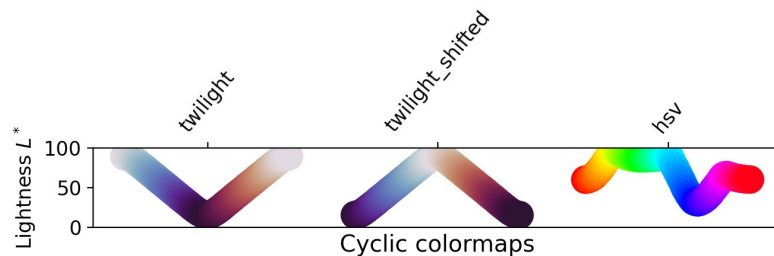
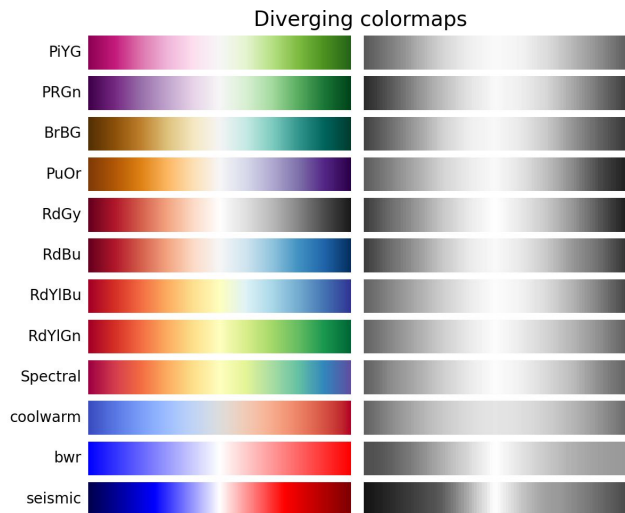
http://www.perceptualedge.com/articles/visual_business_intelligence/rules_for_using_color.pdf

Component	Default Color
Axis lines	Use thin gray lines of medium intensity.
Borders	Whether around the graph as a whole, its plot area, or a legend, when borders are needed (and usually they are not), use thin gray lines of medium intensity.
Background	Use white (or in Excel select “None” for color).
Bars	Use a distinct hue of medium intensity for each data series.
Lines	For thin lines, use a distinct hue of fairly high intensity for each; otherwise, use distinct hues of medium intensity.
Data Points	For small points, use a distinct hue of fairly high intensity for each; otherwise, use distinct hues of medium intensity.

Escalas de colores

- Colores perceptualmente uniformes
- Conversión a grises
- Daltonismo

<https://matplotlib.org/stable/tutorials/colors/colormaps.html>





Principios de diseño



Simplicidad

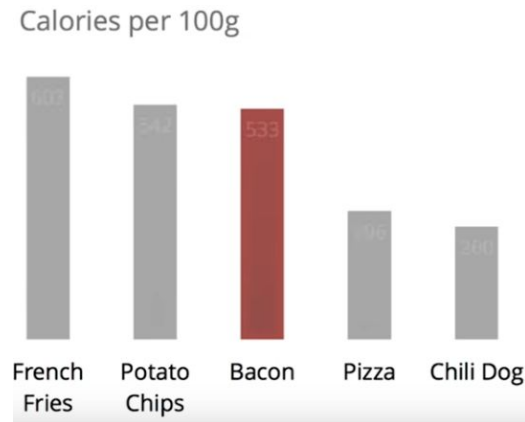
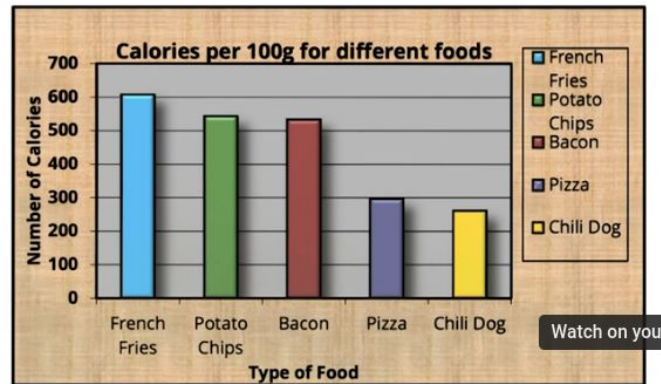
Simplicidad (KISS)

- ¿Es el elemento necesario para transmitir el mensaje?
- Ej: quitar leyendas redundantes

$$\text{data-ink ratio} = \frac{\text{tinta usada para describir los **datos**}}{\text{tinta usada para describir **todo lo demás**}}$$

La simplicidad no es una decisión neutra

- “El medio es un mensaje”
- Lo que NO decimos importa
- Una gráfica simple puede implicar un tema simple



Procesamiento pre-atencional

Acumulación subconsciente de información del entorno

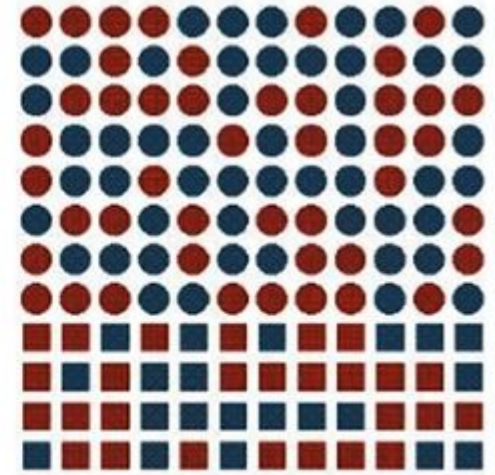
- sirve para resaltar
- sensible a patrones

Detección de :

- color
- forma
- movimiento
- posición espacial



Color + Form ↵

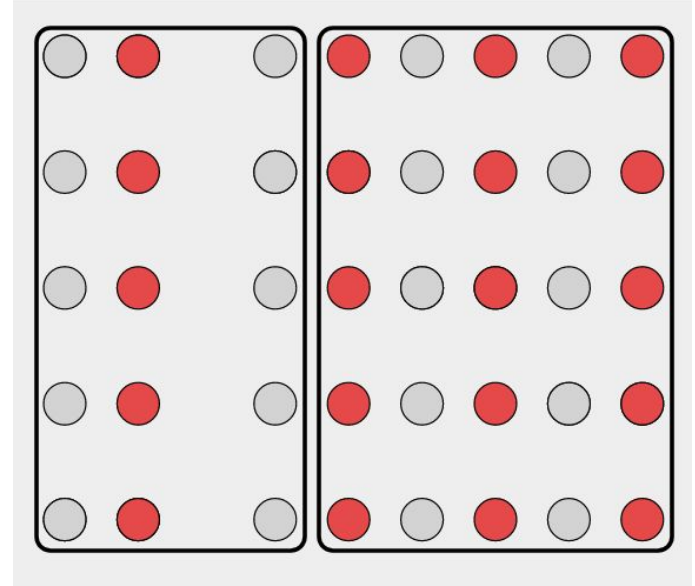


↗ Color + Form

Principios de percepción de Gestalt

Organizamos lo que vemos de maneras particulares para dar sentido a la información visual.

- Proximidad - los objetos cercanos se perciben como un grupo
- Similitud - los objetos similares se ordenan en busca de patrones
- Figura y fondo - percibimos un objeto (figura) y una superficie (fondo)
- Continuidad - los objetos colineales o con la misma dirección se perciben como un grupo
- Cerramiento - tendencia a completar objetos incompletos
- Simplicidad - los objetos se ven como sus elementos simples en lugar de figuras complicadas.



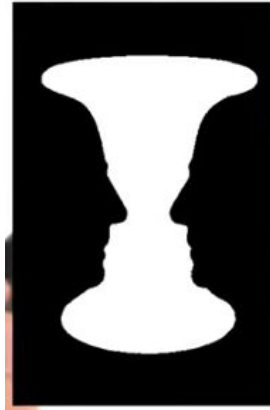
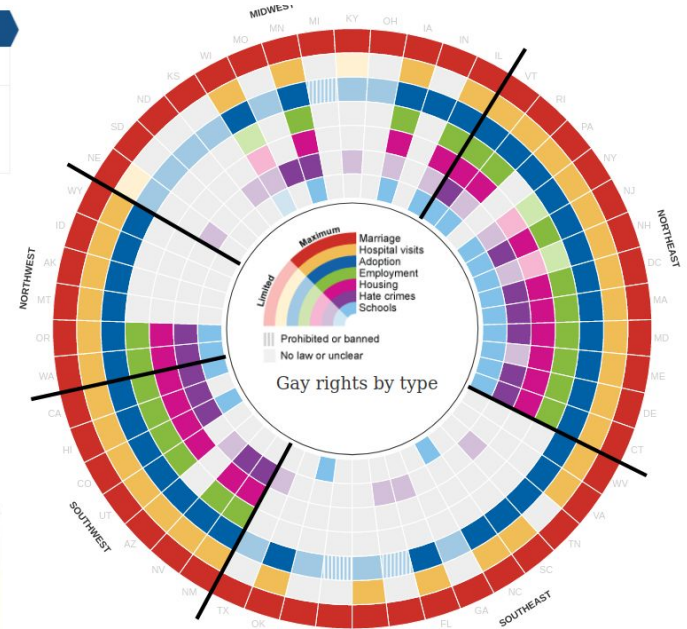
Codificando lo faltante

Representar datos faltantes

- Deja el área vacía
- Usa un símbolo: S/D, ?
- Usa un color o patrón
- Usa anotación o tooltip para dar más información

Espacio negativo es el área vacía alrededor y entre los elementos de un diseño.

- enfatizar elementos clave
- mejorar legibilidad y equilibrio visual
- composición armoniosa
- resaltar faltantes



Tasa de mentiras

$$\text{tasa mentiras} = \frac{\text{tamaño del efecto mostrado en gráfica}}{\text{tamaño del efecto mostrado en datos}}$$

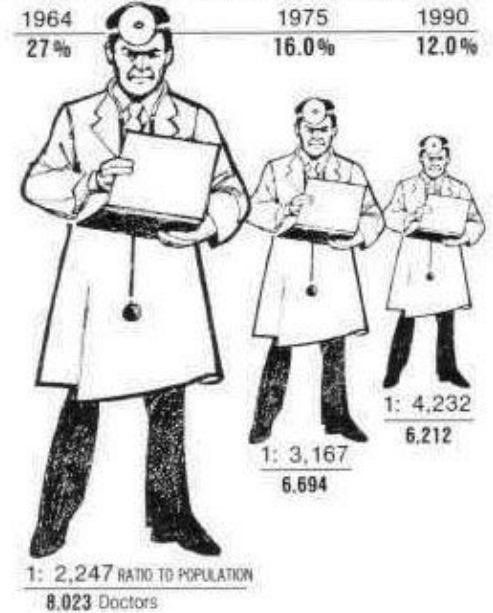
$$\text{tamaño efecto} = \frac{|\text{último valor} - \text{primer valor}|}{\text{primer valor}}$$

¡Considera los valores faltantes en el tamaño del efecto!

THE SHRINKING FAMILY DOCTOR In California

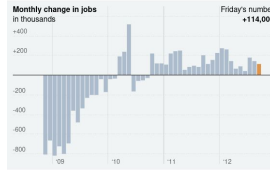
Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27%	16.0%	12.0%

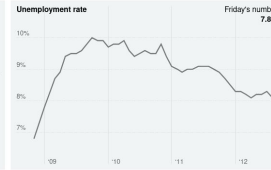


Diferentes perspectivas de una misma gráfica

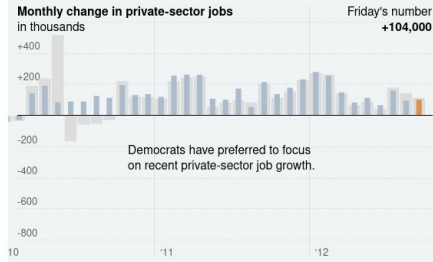
The U.S. economy added 114,000 jobs in September.



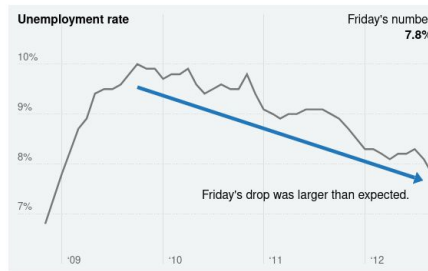
The unemployment rate fell to 7.8%.



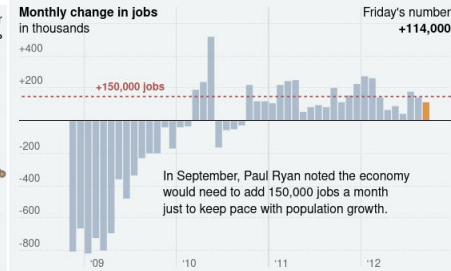
There have been 31 consecutive months of job growth.



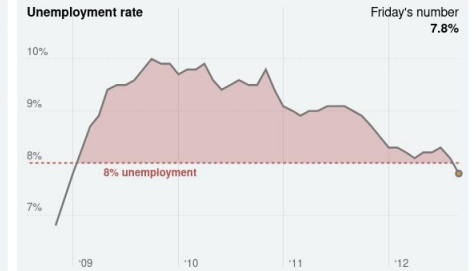
The rate has fallen more than 2 points since its recent peak.



Job growth has been weak.



The rate was above 8 percent for 43 months.



<https://archive.nytimes.com/www.nytimes.com/interactive/2012/10/05/business/economy/one-report-diverging-perspectives.html>

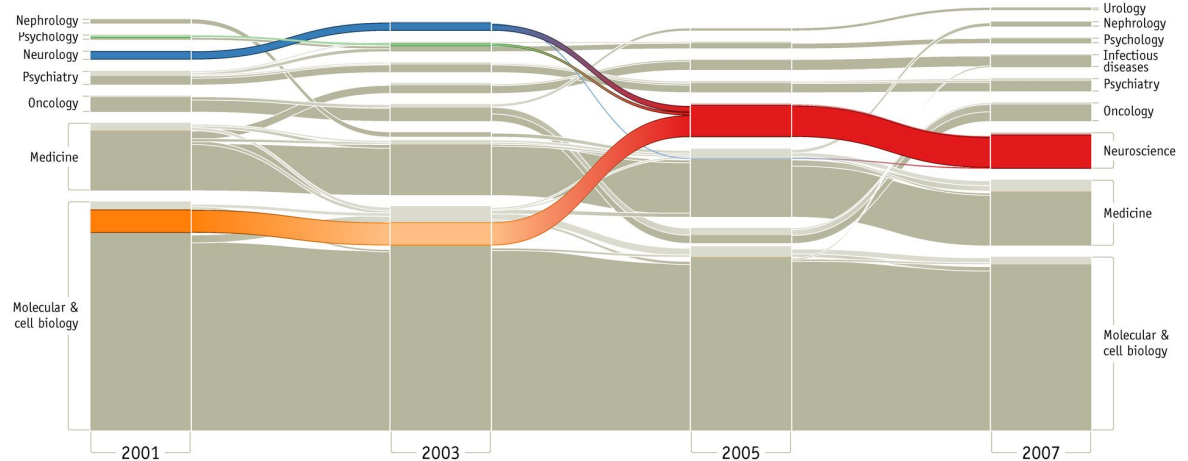


Ejemplos de gráficas



Otros tipos de gráficas

- Tablas
- Mapas
- Diagramas de flujo
- Redes
- Aluviales
- [Bullet Graph](#)
- [Sparklines](#)
- [Cycle Plots](#)
- [Connected Scatter Plots](#)
- [Violin Plots](#)



Tour por el zoológico de las gráficas

<https://queue.acm.org/detail.cfm?id=1805128>

https://en.wikipedia.org/wiki/Alluvial_diagram

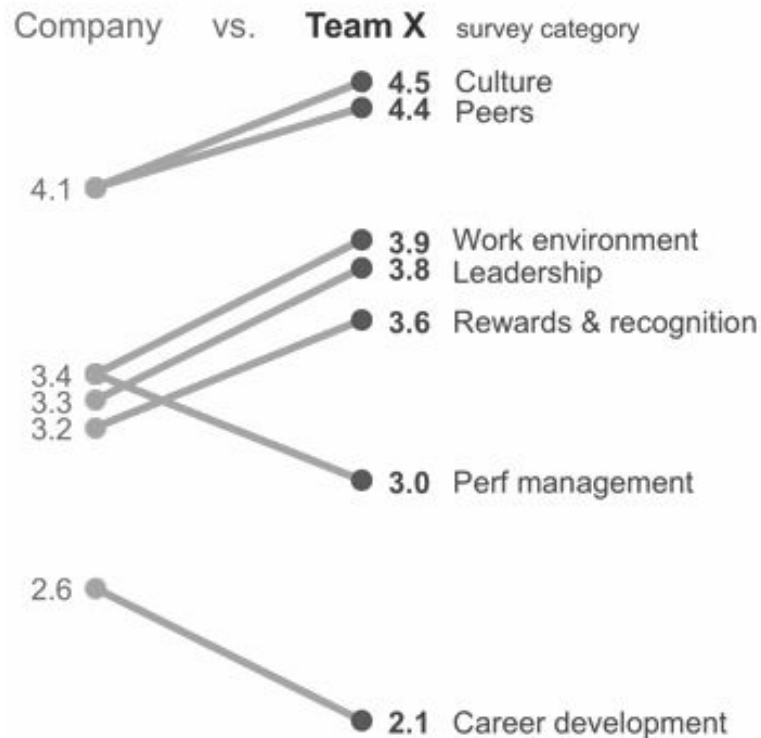
Slope graph (pendientes)

Comparar cambios en valores a lo largo del tiempo o entre dos grupos diferentes.

- ✓ Marcar jerarquías
- ✓ Ilustrar cambios en orden
- ✗ Valores o comparaciones exactos
- ✗ Difícil poner ejes numéricos
- Usar colores para separar grupos
- Usar etiquetas

Employee Feedback

Mean survey score (1=Strongly Disagree, 3=Neutral, 5=Strongly Agree)



Herramientas de visualización

Python

- [Matplotlib](#)
- [Seaborn](#)
- [Plotly](#)
- [Bokeh](#)
- [Altair](#)

Plataformas

- [RawGraphs](#)
- [Tableau](#)
- [PowerBI](#)

