



Visualización de Información y Analítica Visual

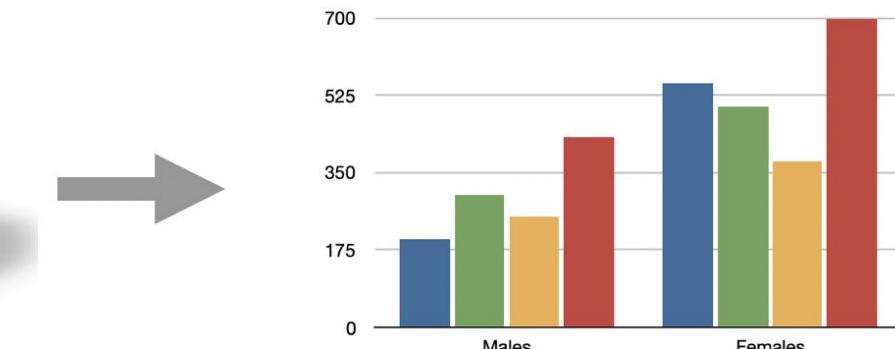
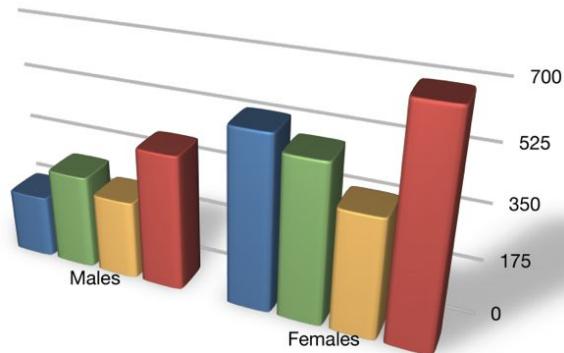
— Hernán Valdivieso López (hfvaldivieso@uc.cl) —

Resumen clase 4

Principios de diseño

Tasa de tinta de datos (*Data ink ratio*)

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



■ 0-\$24,999 ■ \$25,000+ ■ 0-\$24,999 ■ \$25,000+

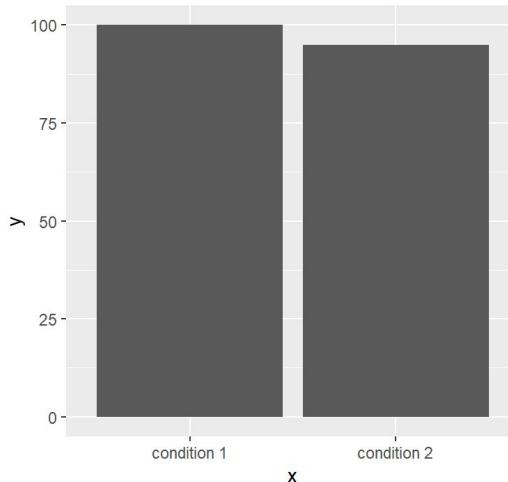
■ 0-\$24,999 ■ \$25,000+ ■ 0-\$24,999 ■ \$25,000+

Principios de diseño

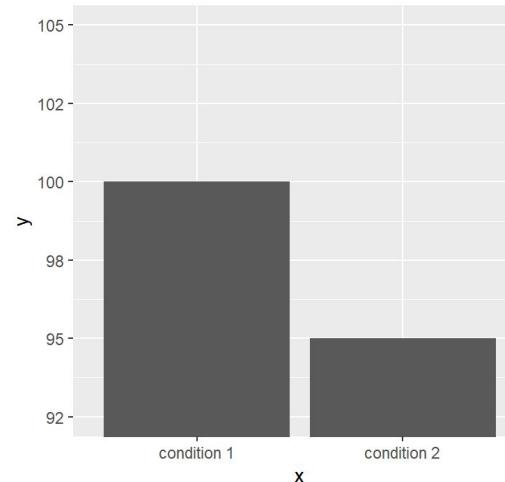
Factor de la mentira (*lie factor*)

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

A lie factor = 1



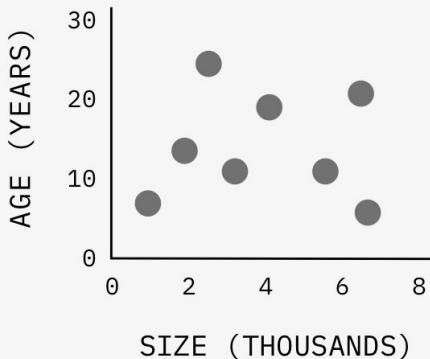
B lie factor ~ 2.8



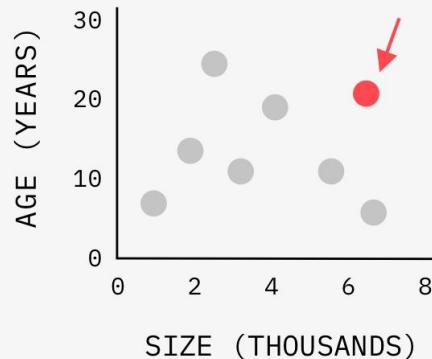
Principios de diseño

Overview first, details on demand

Example: companies by age and size



1. OVERVIEW FIRST



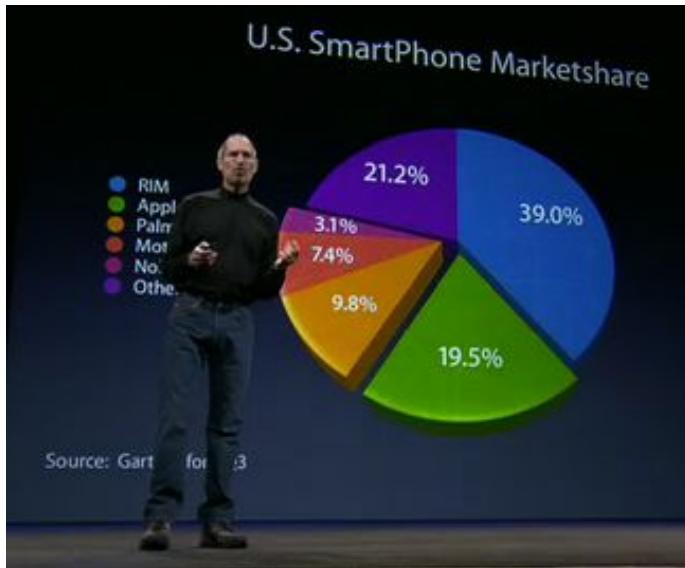
2. ZOOM AND FILTER



3. DETAILS ON DEMAND

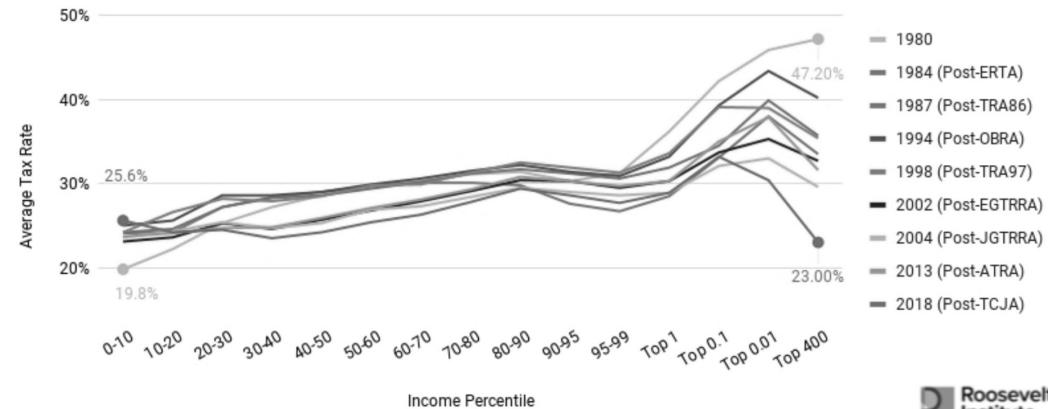
Principios de diseño

Y muchos más....



Average Tax Rates by Income Group after Tax Legislation between 1980 and 2018

While tax rates at the bottom have stayed relatively stable, or slightly increased, since 1980, tax rates for the wealthy have severely declined.



Principios de diseño ...

... relacionados con el uso correcto de canales

- Factor de la mentira (*Lie factor*)
- Ejes engañosos
- No al 3D y 2D injustificado
- Lograrlo en blanco y negro (*Get it right in black and white*)

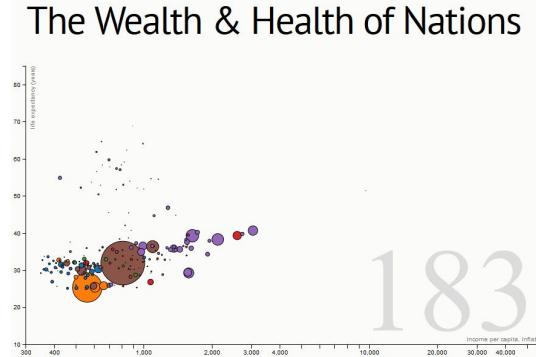
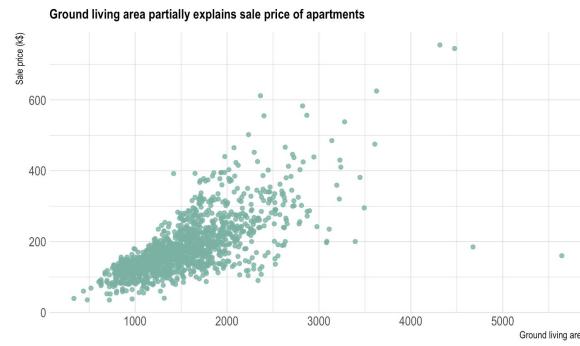
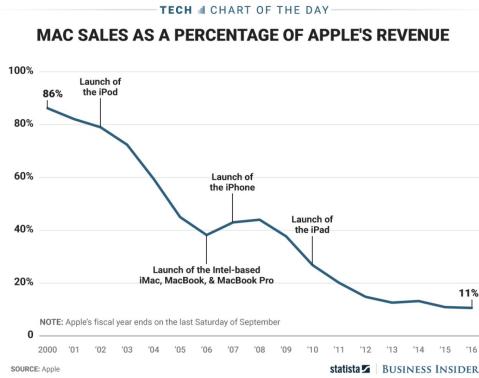
... relacionados con HCI

- *Overview first, details on demand*
- Los ojos le ganan a la memoria (*Eyes beat memory*)
- Tiene que ser receptivo (*Responsive is required*)

... relacionados con el diseño de la visualización

- Tasa de tinta de datos (*Data ink ratio*)
- Consistencia interna y externa
- Autocontención

Gráficos en Matplotlib



Líneas

Dispersión

Burbuja

Gráficos en Matplotlib

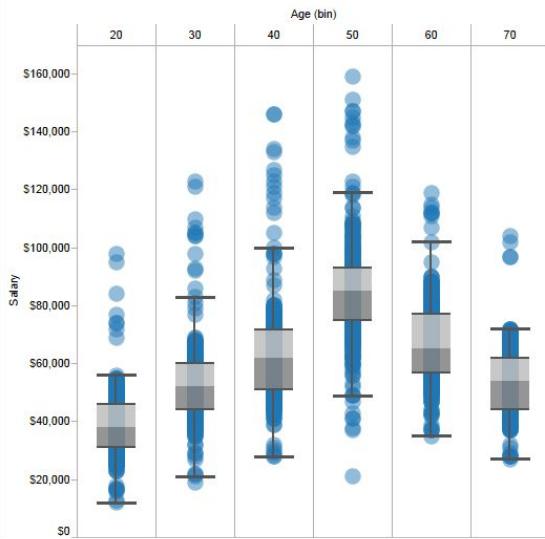


Gráfico de caja

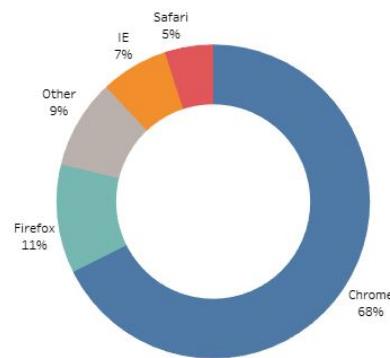
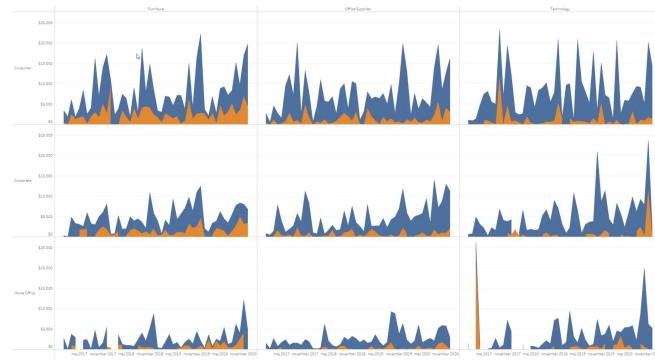


Gráfico de torta/donut



Small Multiples

Clase 5: Teoría del color y visualizaciones en otros dominios (espaciales y texto)

Contenidos

- Teoría del color.
- Visualización de datos espaciales (geográficos y campos espaciales).
 - Proyección cartográfica.
 - Visualizaciones y casos aplicados.
- Visualización de texto.
 - Procesamientos y casos aplicados de visualización de texto.
- Parte práctica: resolver taller no evaluado

Teoría del color

Teoría del color

- ¿Qué es el color?
- Modelos de color
- *Colormaps*
- Formas de escoger colores
- Temas de interés sobre el color

Teoría del color

- **¿Qué es el color?**

- Modelos de color
- *Colormaps*
- Formas de escoger colores
- Temas de interés sobre el color

¿Qué es el color?

🤔 ¿Cuáles son los colores primarios?

🤔 ¿Rojo, azul y amarillo según lo aprendido en el colegio?

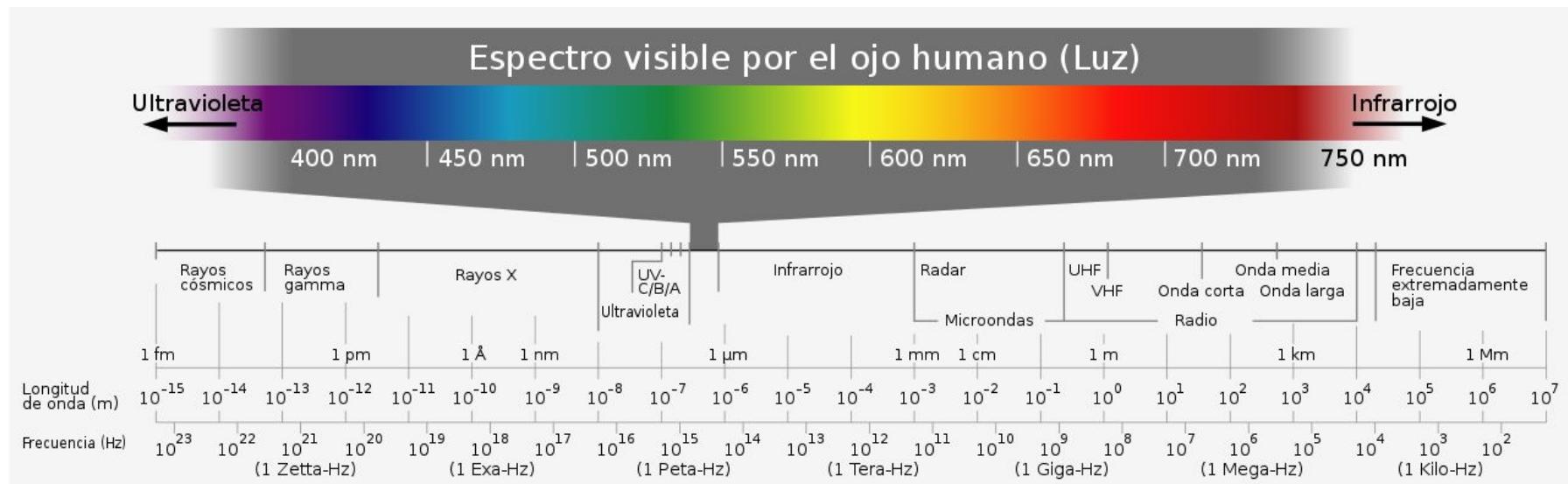
🤔 ¿Rojo, verde y azul según como se define un píxel de una pantalla (RGB)?

🤔 ¿Cian, magenta y amarillo según los colores que usa la impresora?

Todo es correcto 😊. La idea de colores primarios es **arbitraria** y depende de lo que estemos hablando o mejor dicho, del modelo de colores que estemos usando.

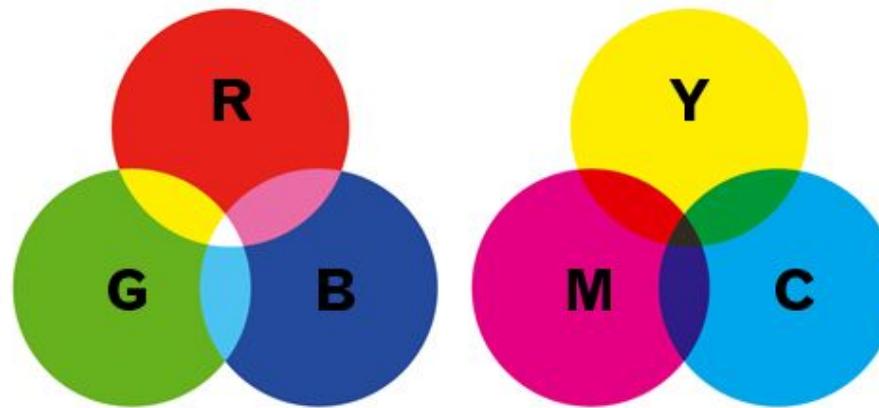
¿Qué es el color? - Desde la ciencia

Luz: radiación electromagnética que recibimos en nuestros ojos y procesamos e interpretamos. Según la radiación que recibamos, será el color que interpretemos.



¿Qué es el color?

- Los colores se experimentan como combinaciones de frecuencias de luz que entran a nuestros ojos.
- **Sistema aditivos:** emisión de luces que suman colores.
- **Sistema sustractivos:** mezcla de pigmentos que restan colores.



Colores aditivos

Colores sustractivos

Teoría del color

- ¿Qué es el color?
- **Modelos de color**
- *Colormaps*
- Formas de escoger colores
- Temas de interés sobre el color

Modelos de color

Abstracciones que buscan representar los colores, muchas veces de forma matemática.

- sRGB(255, 130, 46) o #ff822e
- HSL(24.2, 100%, 59%)
- CMYK(0, 0.489, 0.82, 0)
- CIELAB(67.67, 42.36, 63.4)

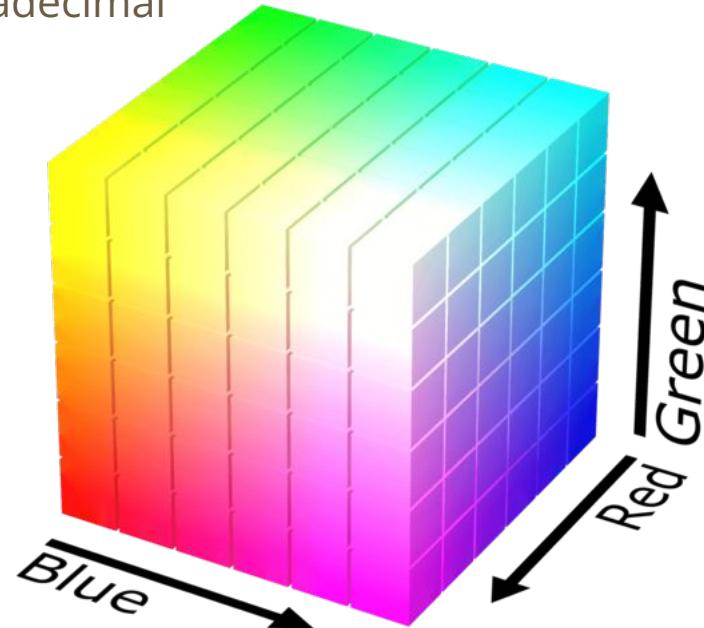


Modelos de color - RGB

Se basa en indicar cuánto rojo (R), verde (G) y azul (B) hay que combinar para generar un color. Se basa en la teoría de los colores aditivos.

Su notación es un conjunto de 3 números o en hexadecimal

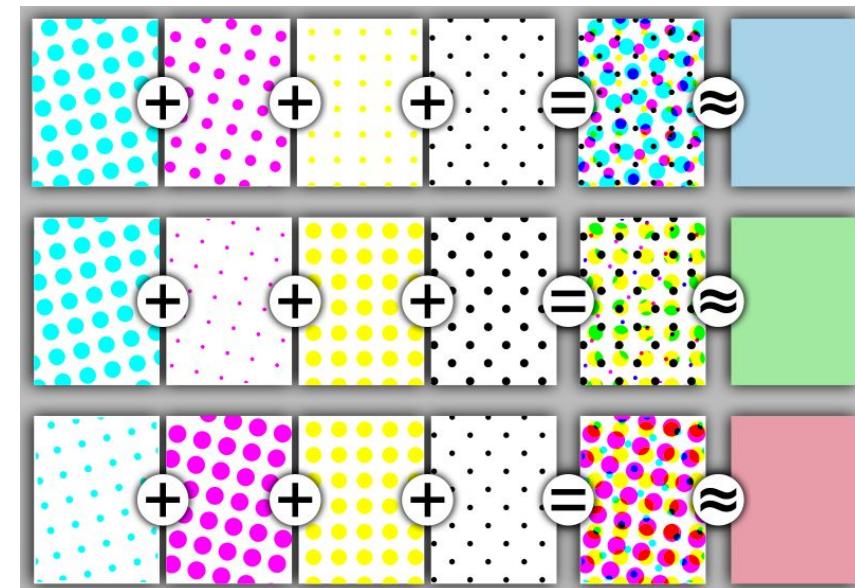
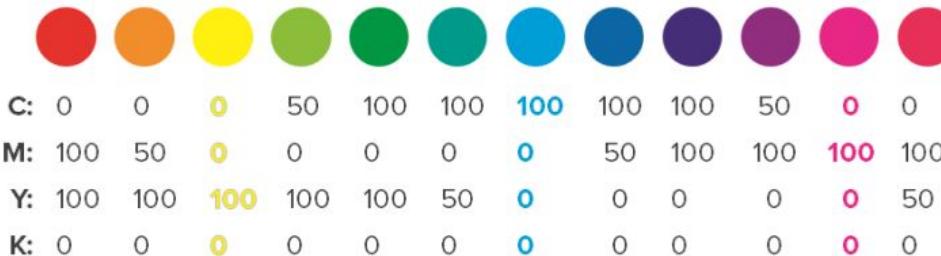
#FF0000	RGB (255, 0, 0)
#FF7F00	RGB (255, 127, 0)
#FFFF00	RGB (255, 255, 0)
#FFFF00	RGB (0, 255, 0)
#0000FF	RGB (0, 0, 255)
#4B0082	RGB (75, 0, 130)
#8F00FF	RGB (143, 0, 255)



Modelos de color - CMYK

Se basa en indicar cuánto cyan (C), magenta (M), amarillo (Y) y negro (K) hay que combinar para generar un color. Se basa en la teoría de los colores sustractivos.

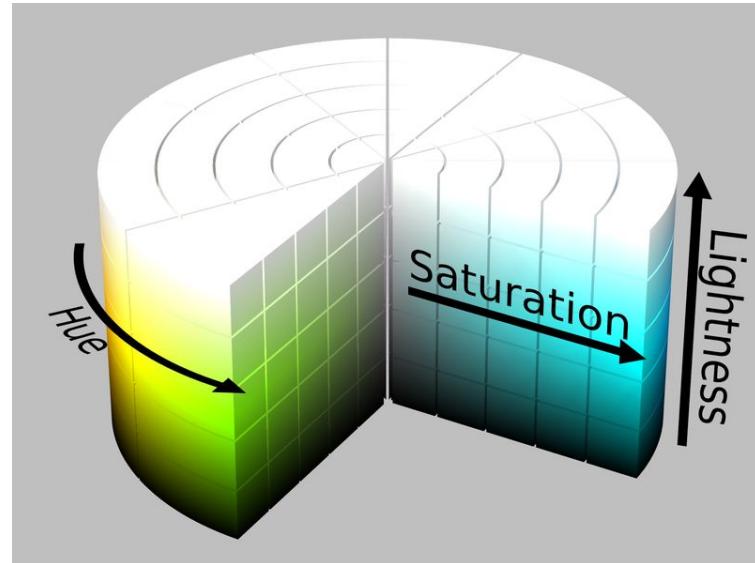
Su notación es un conjunto de 4 números.



Modelos de color - Basado en propiedades (HSL)

Es más intuitivo para nuestra percepción y es usado fuertemente por artistas y diseñadores.

- **Matiz (*hue*)** captura la tonalidad del color. Normalmente conocemos como colores puros, dejando de lado la mezcla del blanco y del negro.
- **Saturación (*saturation*)** especifica la intensidad del color. Qué tan “vivo” o “puro” está.
- **Luminosidad (*lightness*)** especifica la cantidad de luz que recibe el color. Análogo a una linterna.



Teoría del color

- ¿Qué es el color?
- Modelos de color
- ***Colormaps***
 - Formas de escoger colores
 - Temas de interés sobre el color

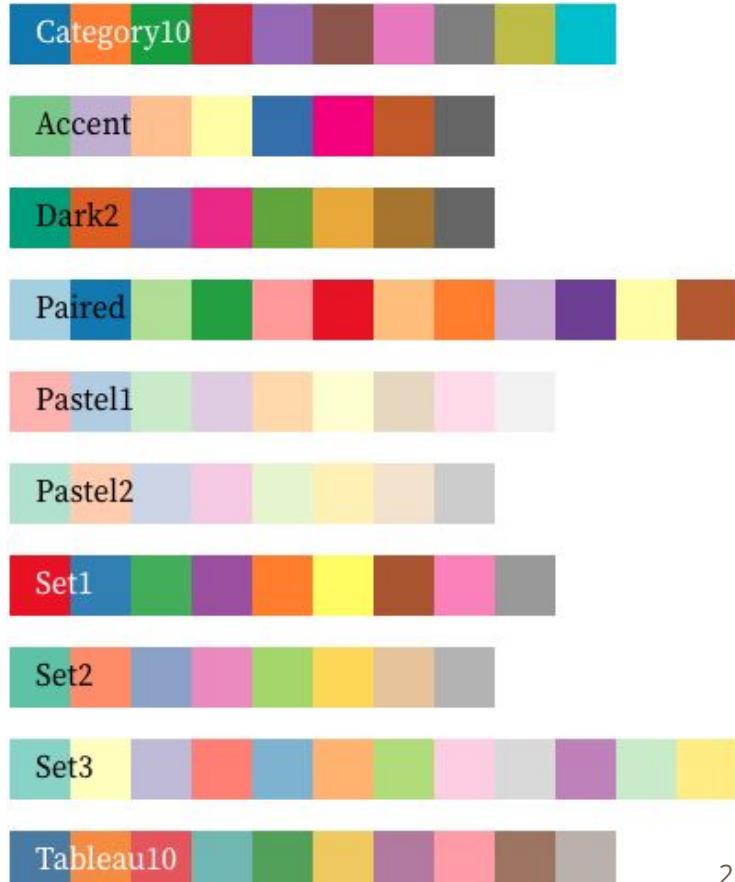
Colormaps

Asociación entre datos y colores. Los principales *colormaps* se categorizan según la secuencia de colores

1. Categóricos
2. Ordenados
 - a. Secuenciales
 - b. Divergentes
 - c. Cílicos

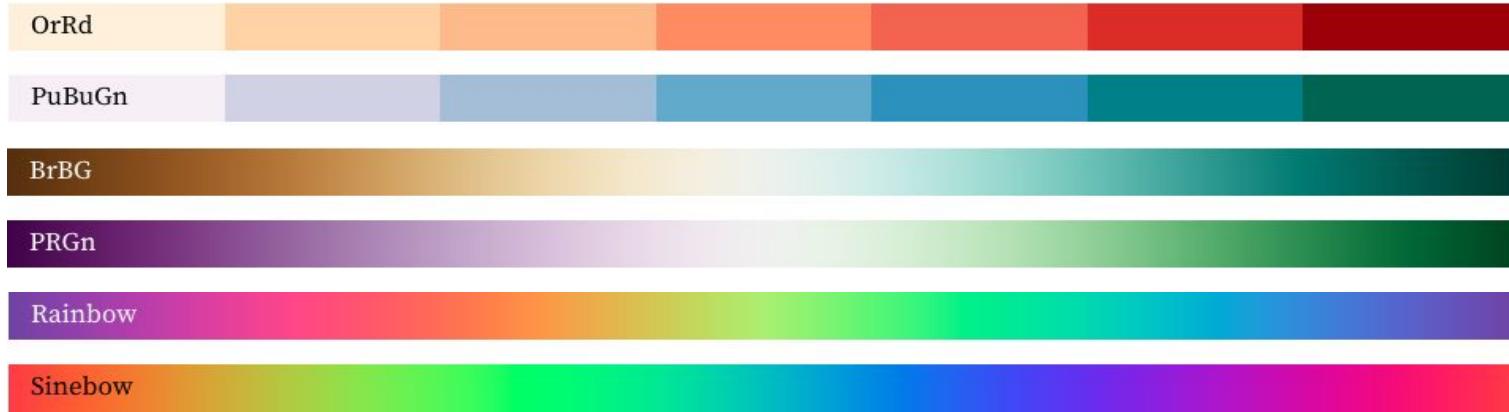
Colormaps - Categórico

- No hay un orden en la secuencia de colores.
- Usado para codificar categorías.
- Limitado a un rango de 6 a 12 colores.
- Se recomienda partir con pocos colores e ir agregando de a poco.



Colormaps - Ordenado

- Se aprecia un orden en la secuencia de colores.
- Se pueden dividir en secuenciales, divergentes y cílicos.



Colormaps - Ordenado secuencial

- La secuencia de colores va en una sola dirección.
- Hay 2 tipos de *colormaps* ordenados secuenciales. Los que ocupan 1 solo matiz en toda la secuencia y aquellos que ocupan más de uno.



Colormaps - Ordenado divergente

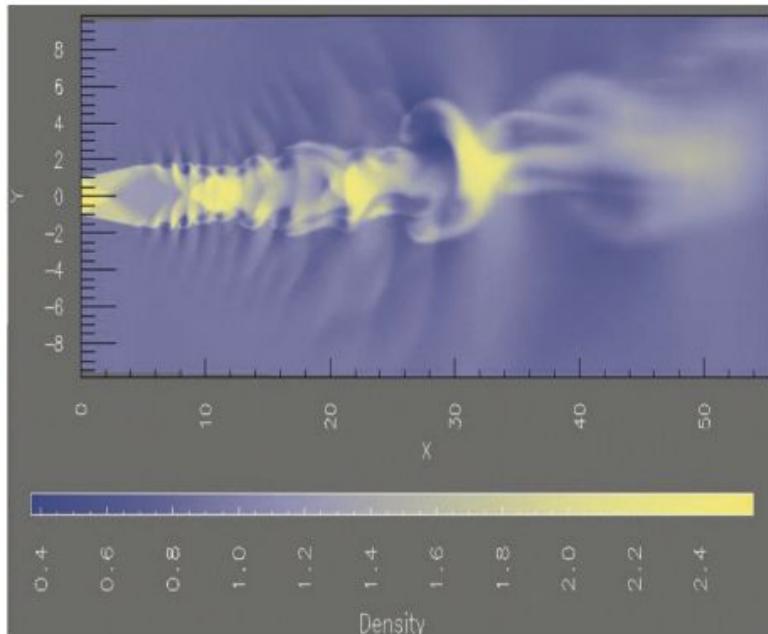
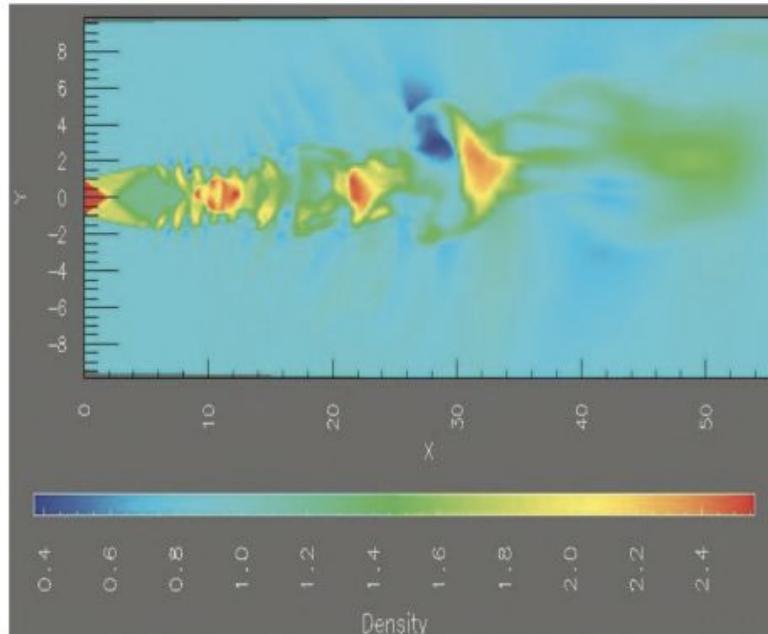
- Se determina un punto intermedio y la secuencia de colores avanza en ambas direcciones.
- En términos de color: tiene dos matices en puntos finales opuestos, y un **color neutral como punto medio**. Puede ser blanco, gris, o uno de alta luminancia como el amarillo.



Colormaps - Ordenado

🤔 ¿Cuándo usar estos colormaps?

Depende 100% de la tarea visual y lo que deseamos enfatizar.



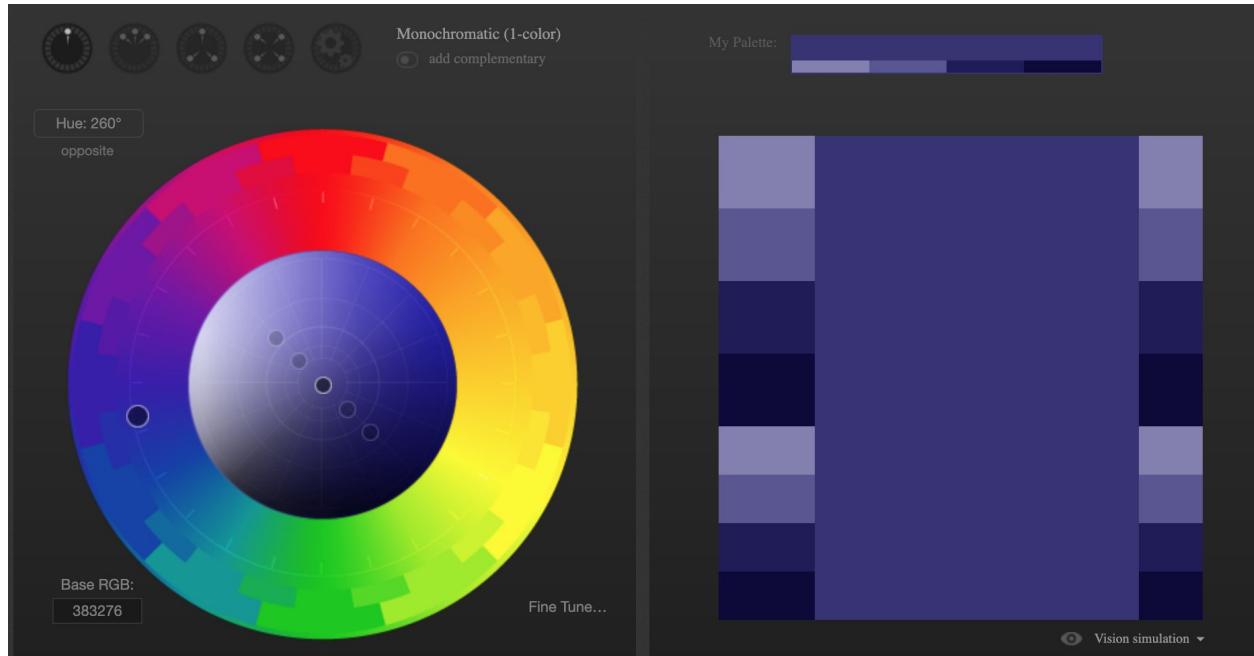
Teoría del color

- ¿Qué es el color?
- Modelos de color
- *Colormaps*
- ***Formas de escoger colores***
- Temas de interés sobre el color

Formas de escoger colores

Apoyándonos del principio de expresividad

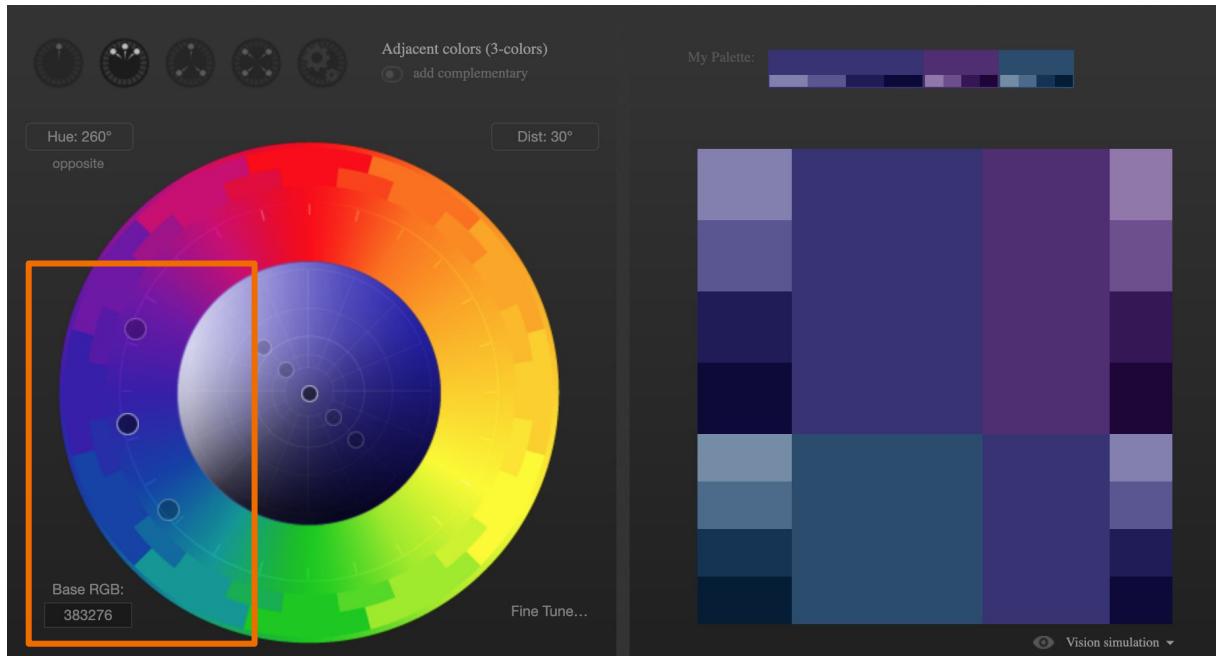
- Si los datos son ordenados. Es buena idea recurrir a un **colormap monocromático**



Formas de escoger colores

Apoyándonos del principio de expresividad

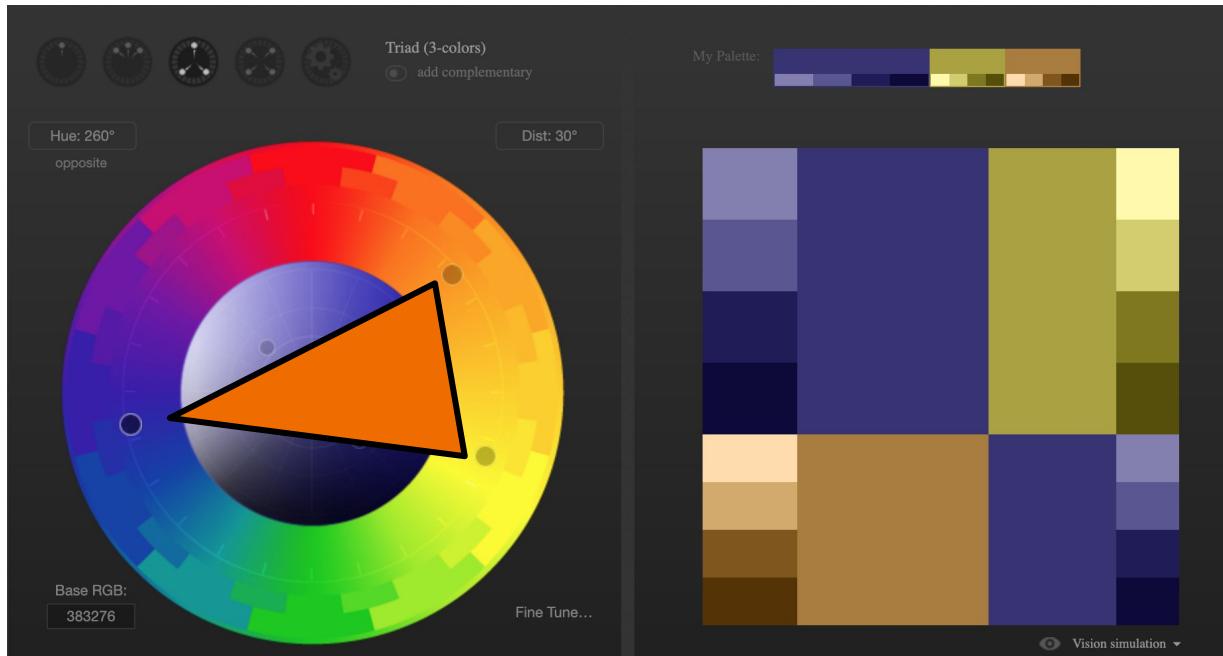
- Si los datos son categóricos. Se puede recurrir a un **colormap adyacente (o análogo)**. Son colores diferentes pero cercanos al color de inicio.



Formas de escoger colores

Apoyándonos del principio de expresividad

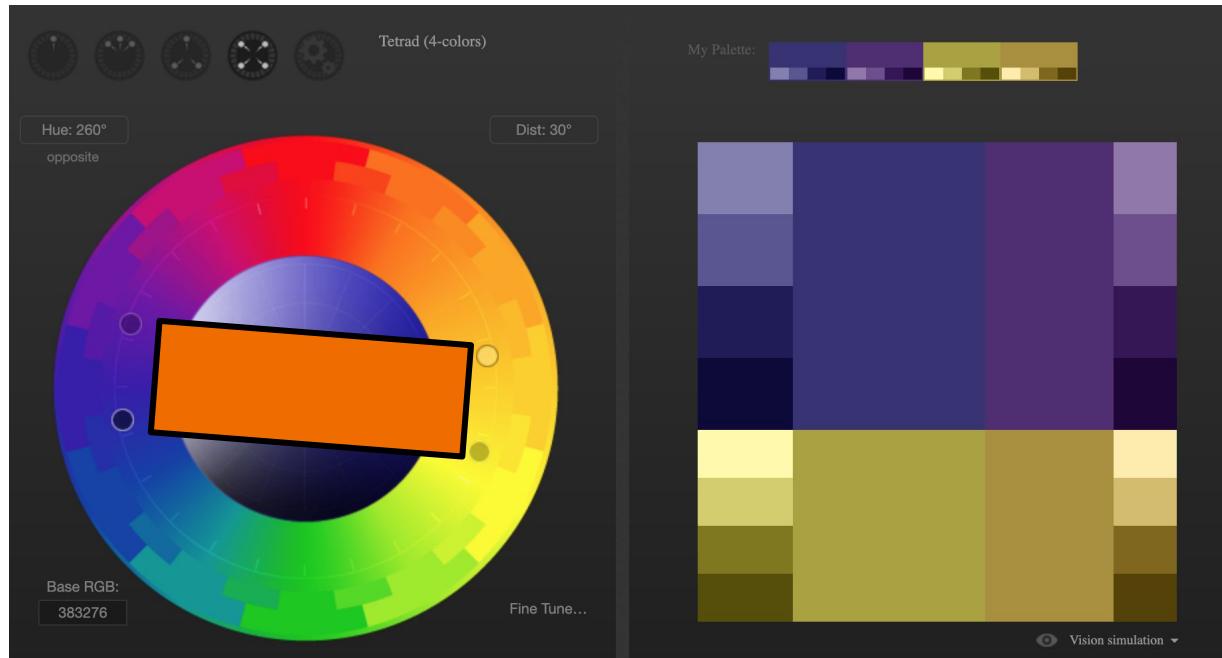
- Si los datos son categóricos. Se puede recurrir a un **colormap de triada**. Se forma un triángulo entre los 3 colores.



Formas de escoger colores

Apoyándonos del principio de expresividad

- Si los datos son categóricos. Se puede recurrir a un **colormap de tetrádicos**. Se forma un cuadrado entre 4 colores debido a que son 2 pares de colores complementarios.



Teoría del color

- ¿Qué es el color?
- Modelos de color
- *Colormaps*
- Formas de escoger colores
- **Temas de interés sobre el color**

Psicología del color

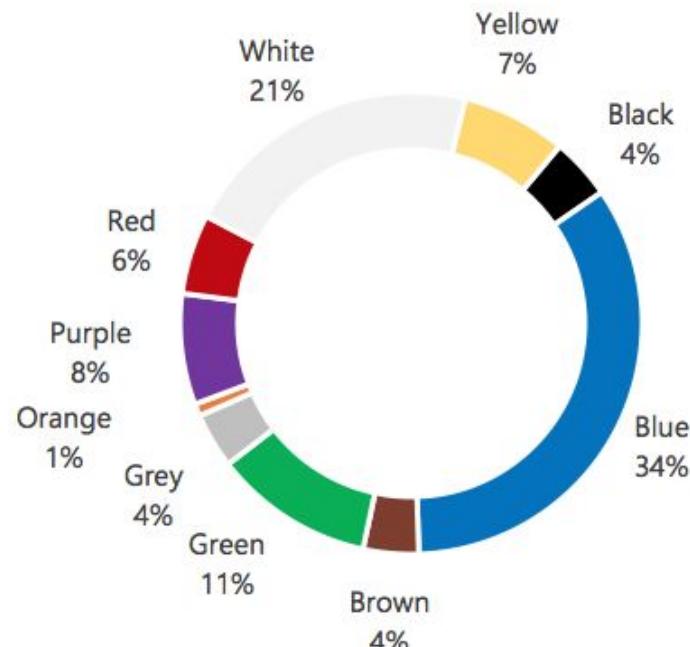
Psicológicamente asociamos ciertos colores a algunos sentimientos o conceptos



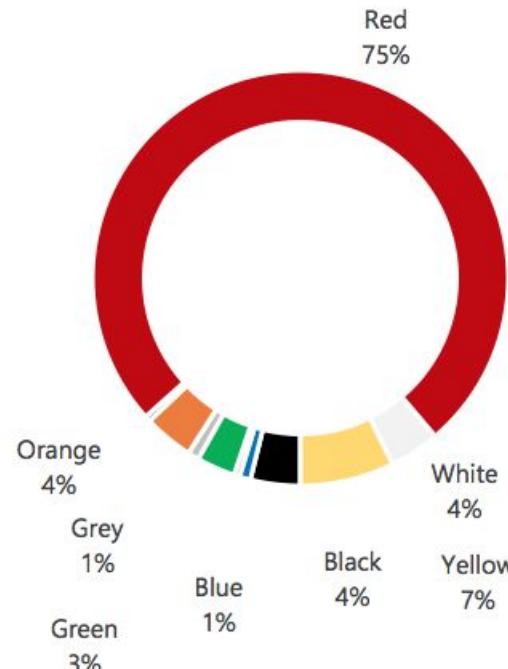
Psicología del color

Psicológicamente asociamos ciertos colores a algunos sentimientos o conceptos

Trust

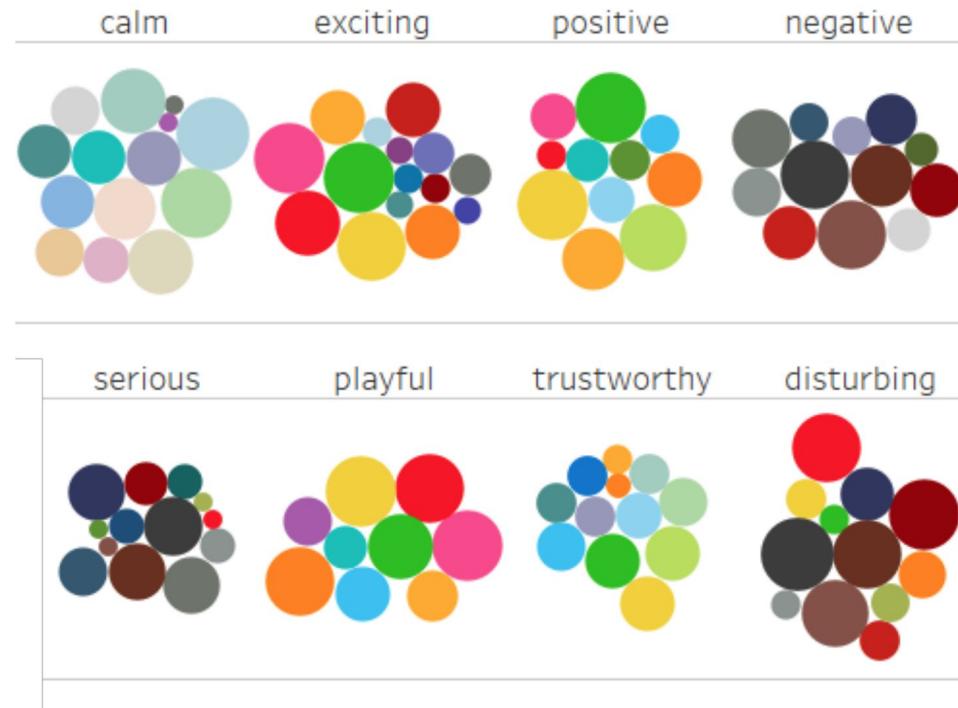


Speed



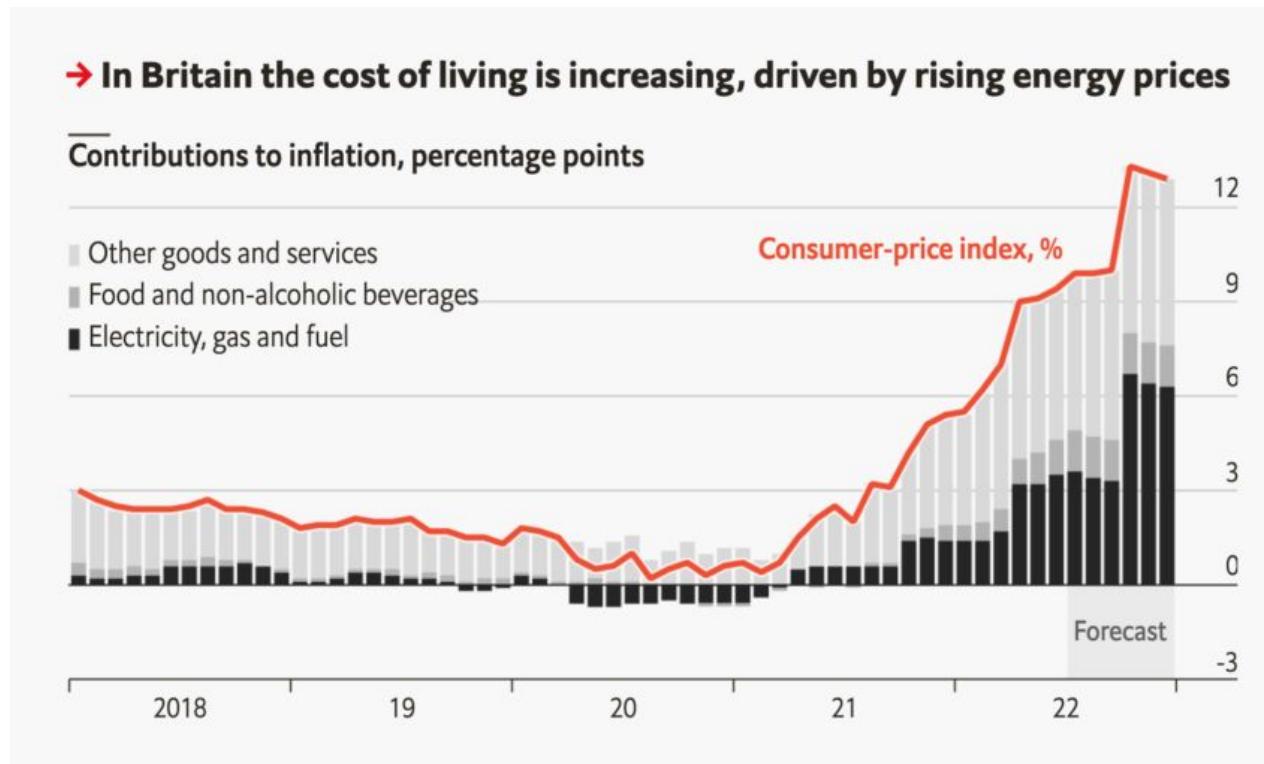
Psicología del color

Psicológicamente asociamos ciertos colores a algunos sentimientos o conceptos



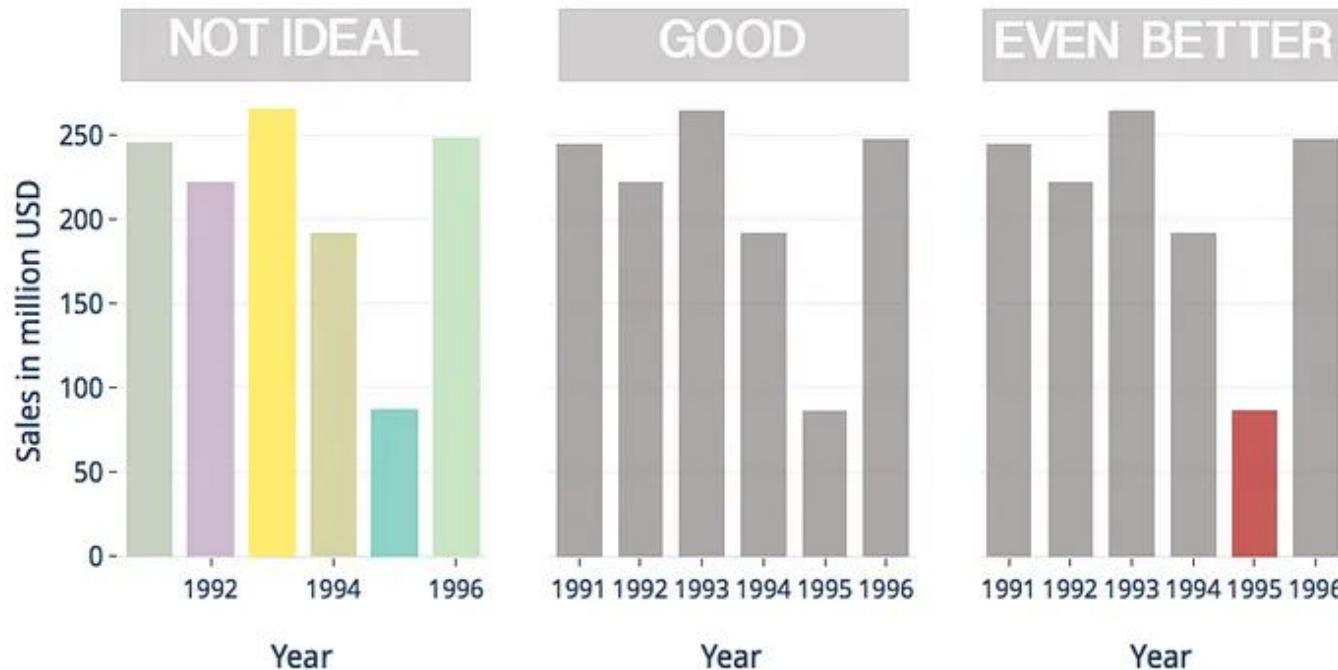
Psicología del color

Hay colores, como el gris, que pasan más desapercibidos.



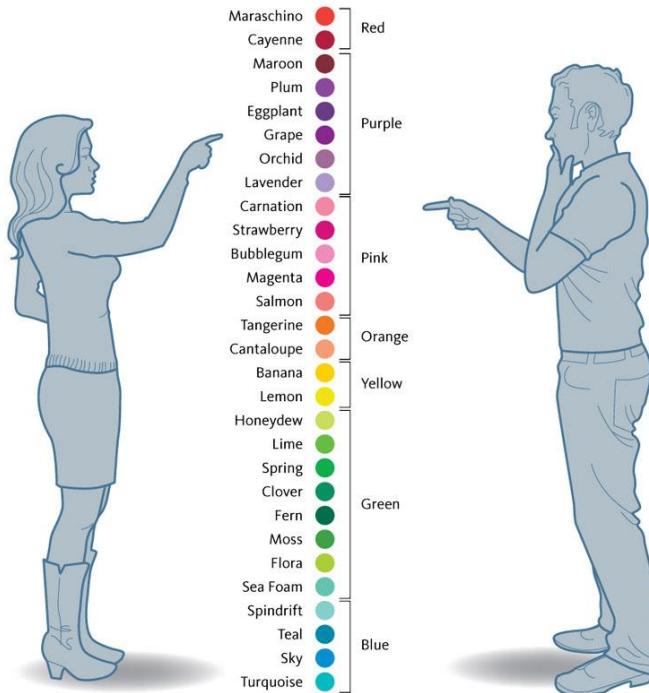
Psicología del color

Hay colores, como el gris, que pasan más desapercibidos.



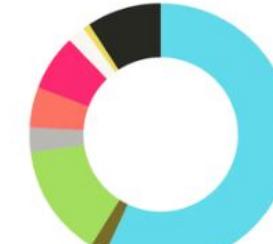
Percepción del color entre hombres y mujeres

Estudios muestran que las mujeres y los hombres tienen diferencias de percepción en los colores, tanto física como psicológicamente.



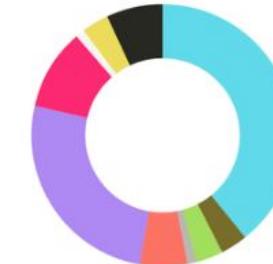
COLORES FAVORITOS ENTRE HOMBRES Y MUJERES

HOMBRES



- Azul 57%
- Rojo 7%
- Marrón 2%
- Verde 14%
- Gris 3%
- Naranja 5%

MUJERES



- Azul 35%
- Morado 23%
- Rojo 9%
- Marrón 3%
- Verde 14%
- Gris 1%
- Amarillo 3%
- Naranja 5%

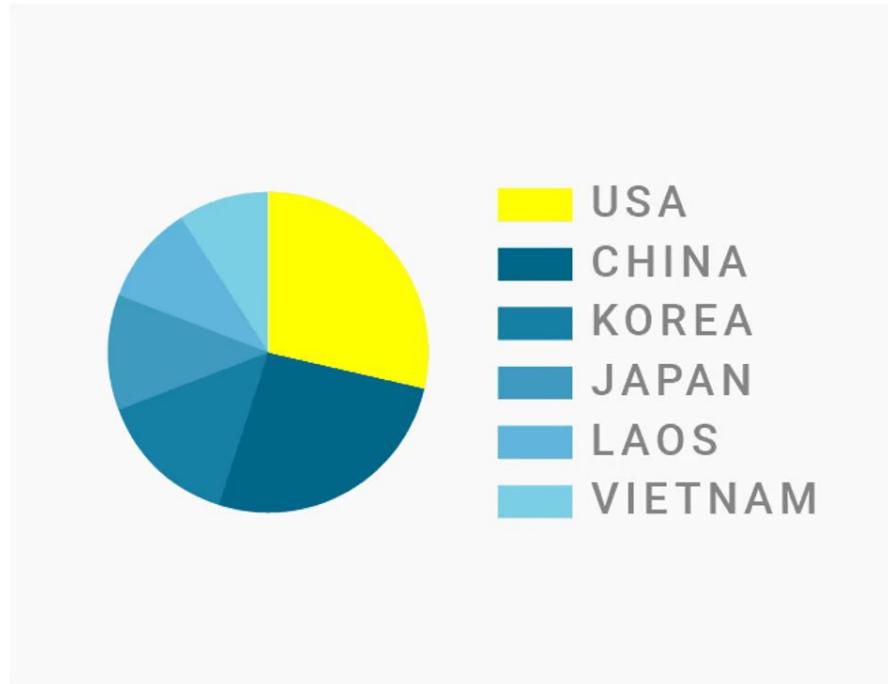
Daltonismo

- Alteración de origen genético que afecta a la capacidad de distinguir los colores.
- Recomendación para hacer visualizaciones más accesibles:
 - Evaluar si es estrictamente necesario usar color.
 - En caso de hacerlo, diseñar no solo con matiz, si no con saturación y luminosidad.
 - Simular efectos de daltonismo sobre elecciones y reconocer efectividad.

Fatiga ocular y colores saturados

Cuando recurrimos a colores muy saturados o muy brillantes, aceleramos el proceso de **fatiga ocular**.

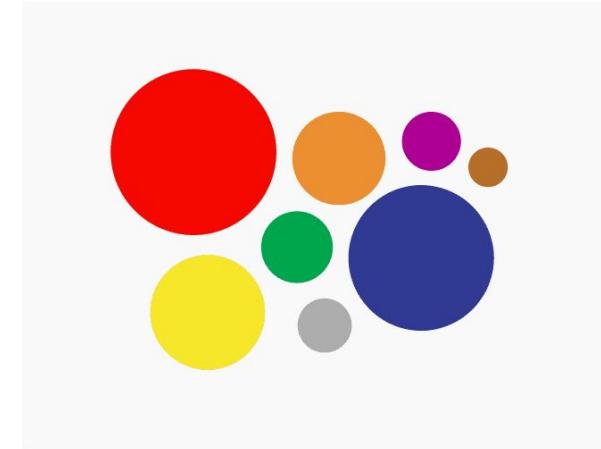
Esto produce que el usuario canse su vista y ya no esté totalmente concentrado en ver la pantalla/visualización



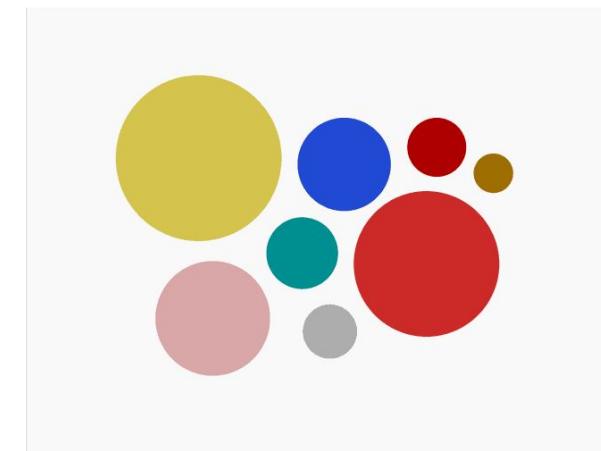
Fatiga ocular y colores saturados

Para evitar esta situación, se aconseja:

- Utilizar **colores de tonos pastel** en lugar de colores vivos como el rojo, fucsia, amarillo y naranja en una misma gráfica.
- También hay que tener en cuenta el **contexto del uso de color**. Si es para destacar un valor, tonos saturados será mejor.

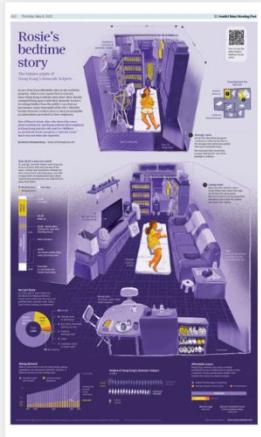


NOT IDEAL



BETTER

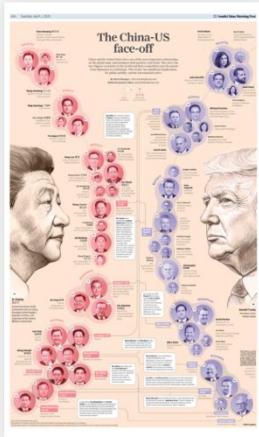
Fatiga ocular y colores saturados



May 8, 2025

Rosie's bedtime story

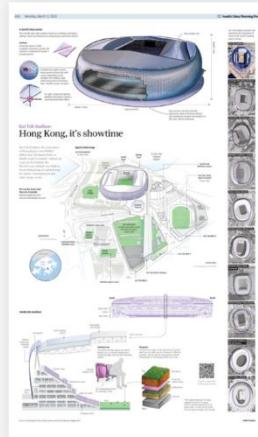
Davies Christian Surya



April 2, 2025

The China-US face-off

Rocio Marquez, Jessica Folker



March 3, 2025

Kai Tak Stadium: Hong Kong, it's showtime

Lau Ka-kuen, Marcelo Duhalde



January 27, 2025

Get ready for the Year of the Snake

Victor Sanjinez



January 20, 2025

Trump's immediate agenda

Victor Sanjinez, Andrew London

Rainbow palette

Tambien en el área de color hay decisiones de diseño que no están totalente claras

- [Rainbow Colormaps Are Not All Bad](#)
- [Rainbow Color Map Critiques: An Overview and Annotated Bibliography](#)
- [Rainbow color map distorts and misleads research in hydrology – guidance for better visualizations and science communication](#)

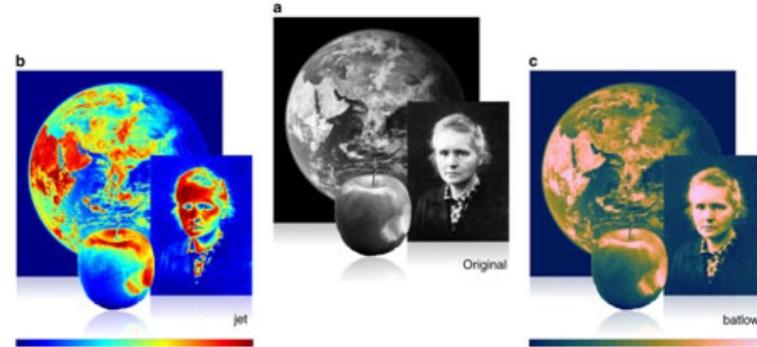
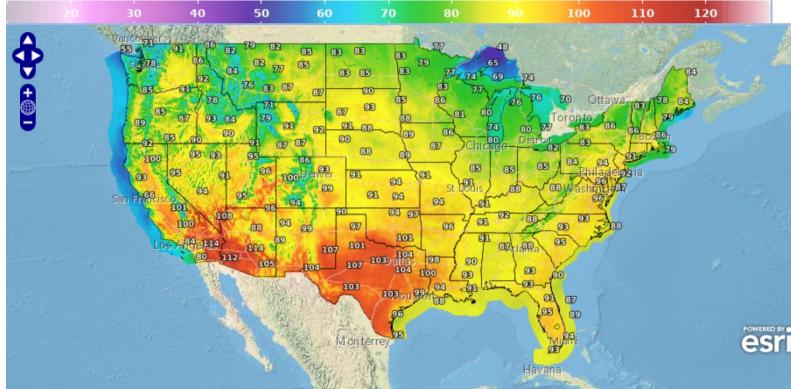


FIGURE 4. Three images demonstrating the effect of different colormaps from Crameri et al.⁷

Links de interés

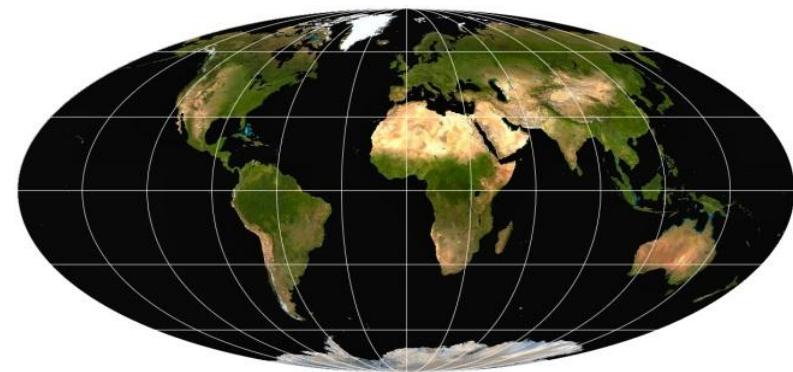
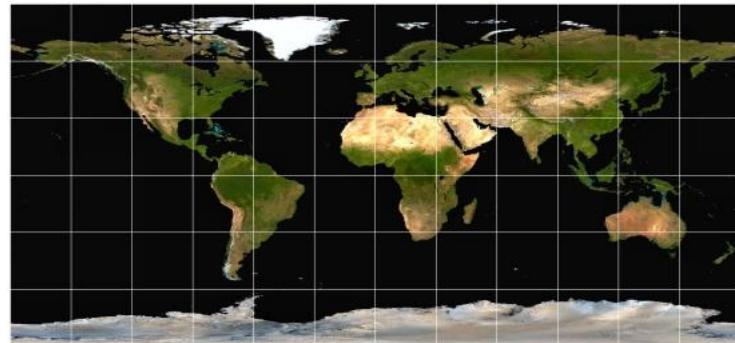
- [Colorizer.org](#)
- [Colors - Generador de paletas de colores](#)
- [Extensión de chrome: Colorblindly](#)
- [ColorBrewer](#)
- [Herramienta para verificar paleta de colores antes de usar](#)
- [Paletton](#)
- [Colorfle \(juego\)](#)

Visualización de datos espaciales

Proyección Cartográfica

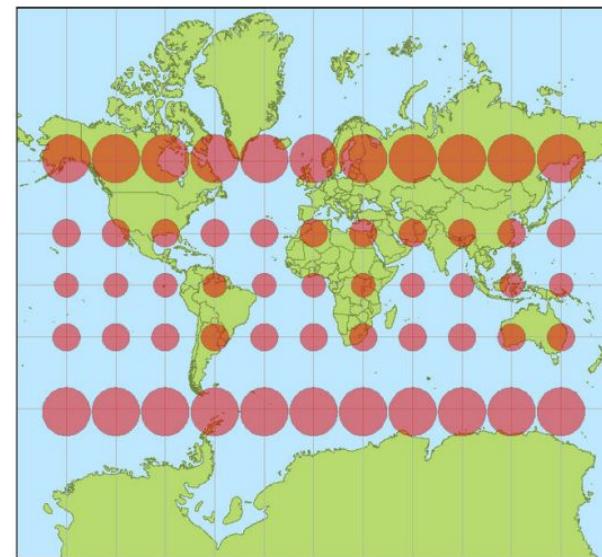
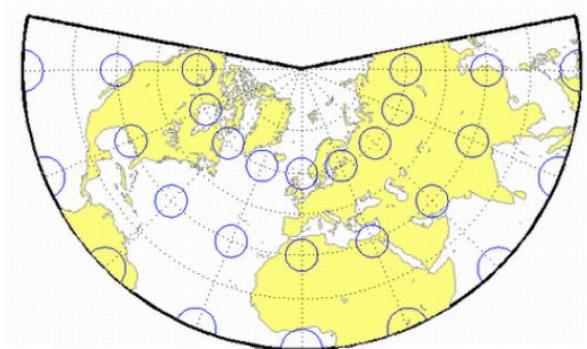
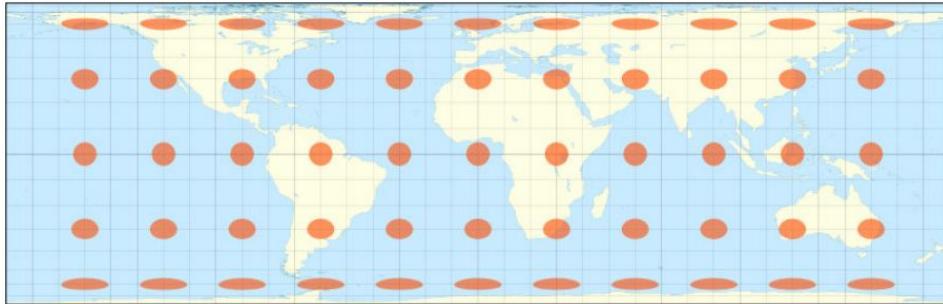
Proyección cartográfica

- Mapeo de posiciones en el globo (esfera) a posiciones en la pantalla (superficie plana).
- Software gratuito **G Projector** de la NASA. Explora una gran variedad de estas proyecciones.



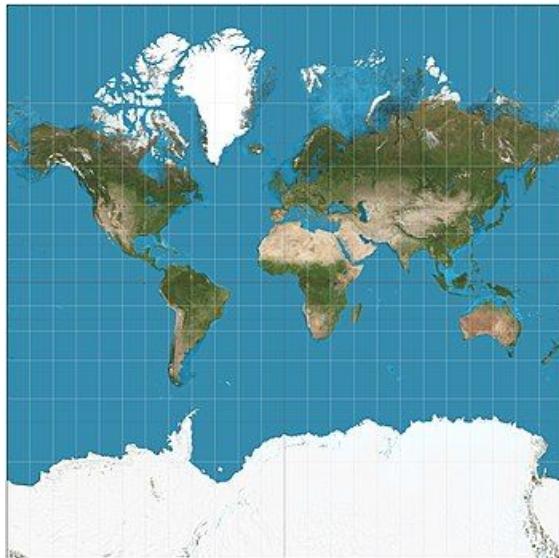
¿Por qué ver proyecciones?

Pasar de 3D a 2D no se logra del todo. **Se debe ceder** en algún punto.



Proyección de Mercator (1569) - Conforme

Diseñada para facilitar el **trazado de rutas** en la superficie terrestre dado que mantiene ángulos (deformación conforme).

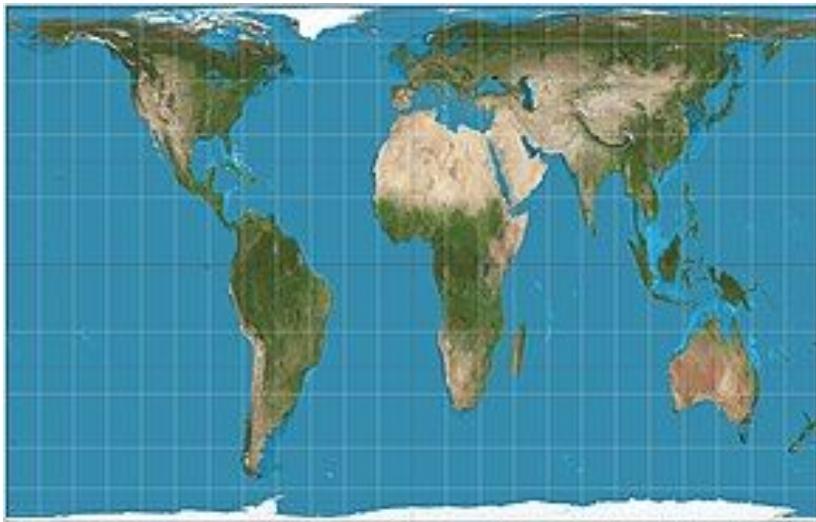


Las superficies se **deforman** según se aproximan a **latitudes polares**.

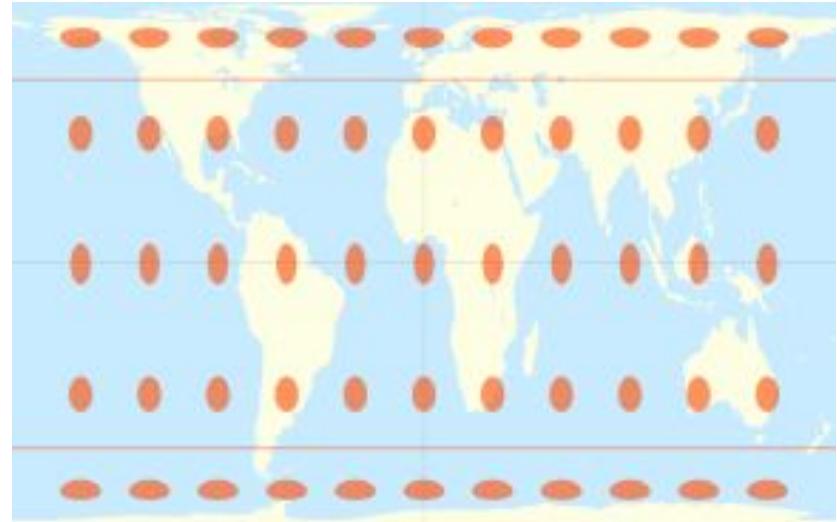


Proyección de Peters (1855) - Equivalente

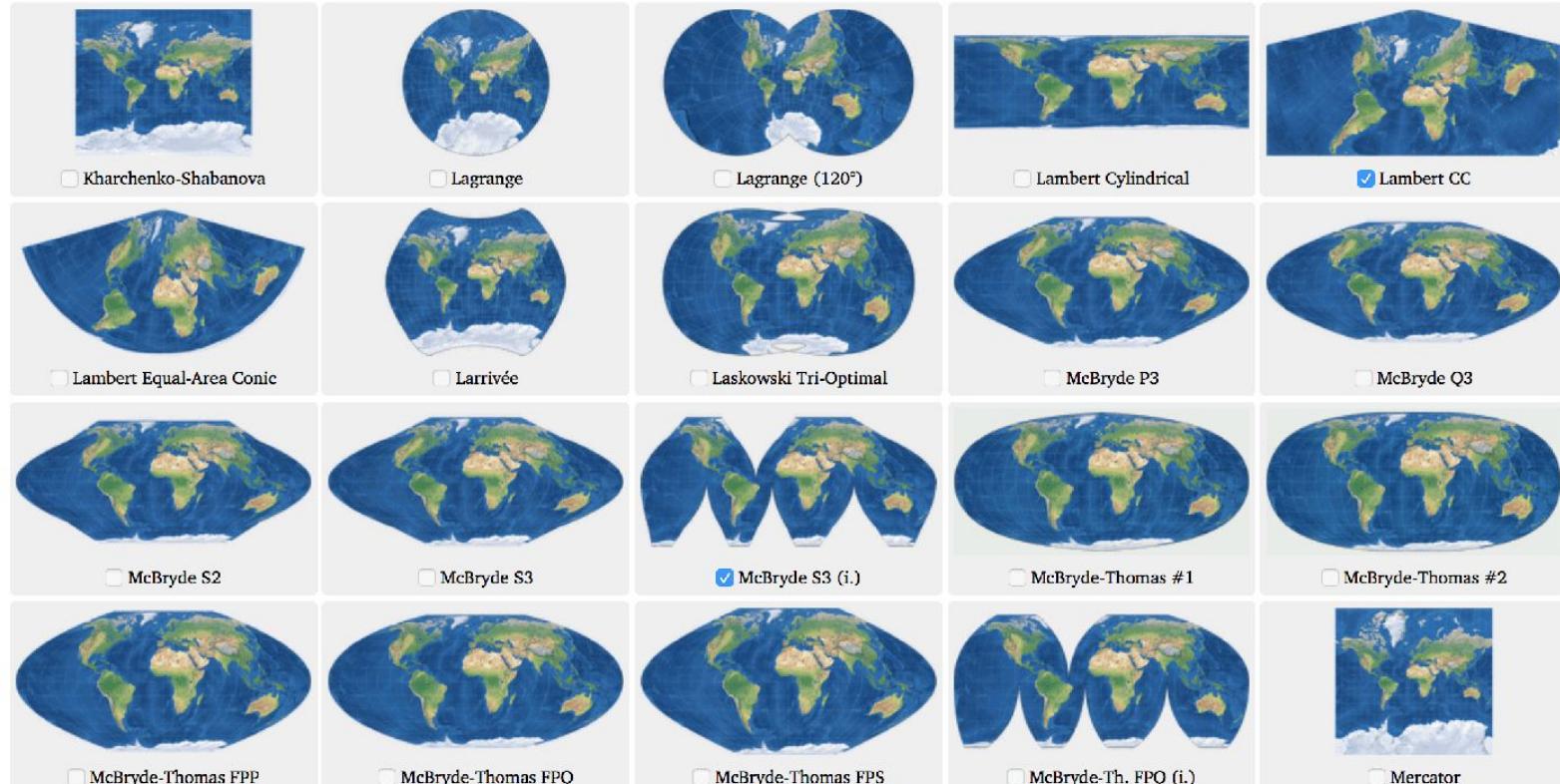
Al contrario de la proyección de Mercator, esta proyección se asegura de **mantener las áreas** (deformación equivalente).



A costa de mantener la proporción de áreas, se **distorsiona la forma** real de los países.



Y hay muuuchas más



[Lista de proyecciones](#)

Links de interés

- [The True Size Of ...](#)
- <https://engaging-data.com/country-centered-map-projections/?c=152>
- [Google Map ocupa una variante de Mercator \(Web Mercator\)](#)

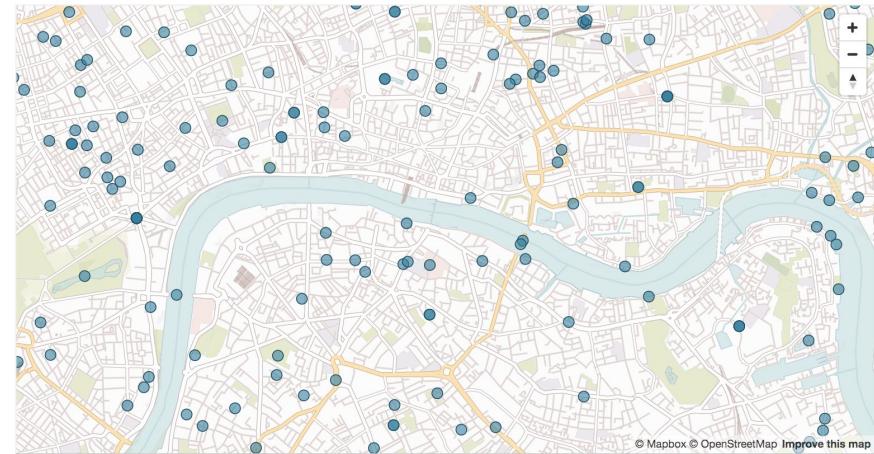
Visualización de datos espaciales

Visualizaciones y casos
aplicados

Visualización de datos geoespaciales

Mapa de puntos

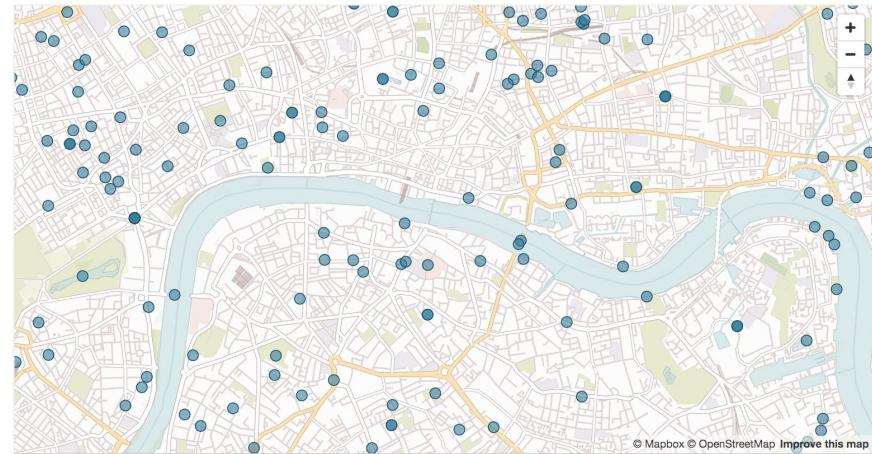
- Visualizan fenómenos puntuales colocando un símbolo o pixel donde se produce el fenómeno.
- Se usan círculos, barras, cuadrados, etc. El valor está codificado por tamaño o color.



Visualización de datos geoespaciales

Mapa de puntos

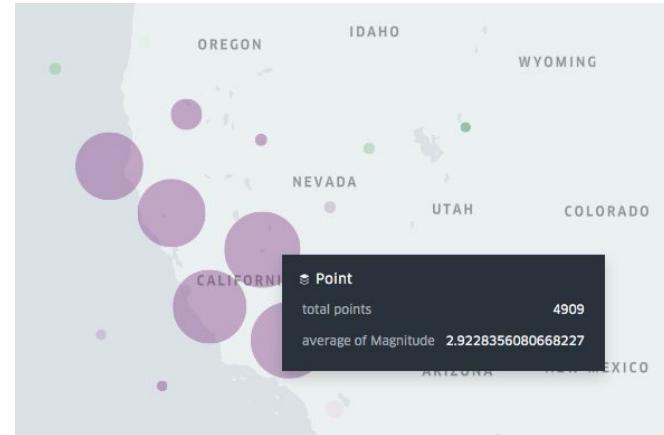
- **Ventaja:** gran facilidad de comprensión e ilustran la variación de una cierta densidad espacial.
- **Desventaja:** la superposición de puntos dificulta comparar zonas con alta densidad de eventos.



Visualización de datos geoespaciales

Cluster map

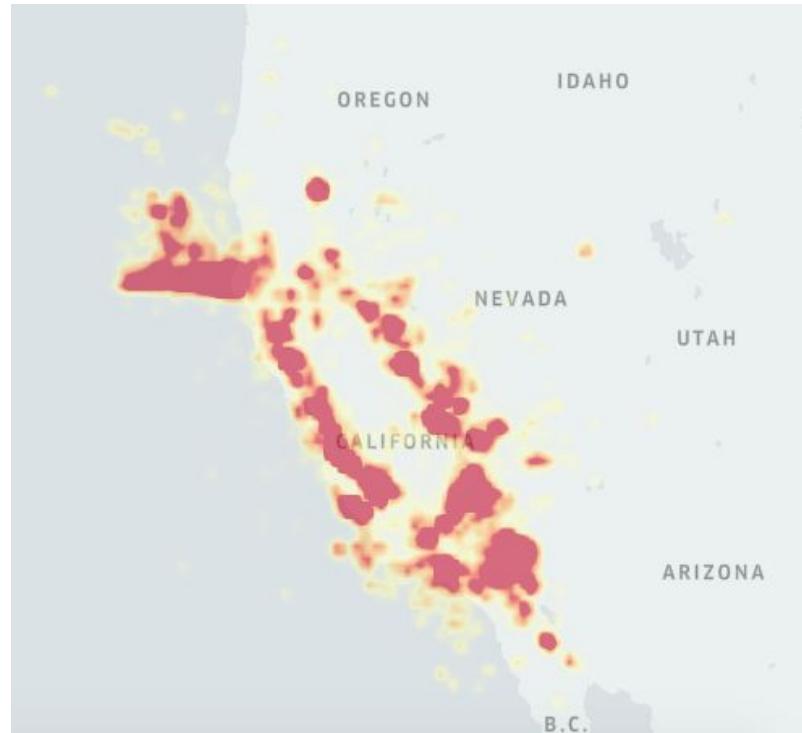
- Agrupa todos los puntos que estén a cierta distancia entre ellos. El tamaño del círculo puede ser en función de la cantidad de puntos o de algún otro atributo
- **Ventaja:** Permite identificar zonas de altas concentraciones donde los puntos se podrían superponer.
- **Desventaja:** Requiere realizar zoom si deseamos ver alguna ubicación puntual de los datos.



Visualización de datos geoespaciales

Heatmaps

- Superpone un mapa de calor sobre el mapa geográfico. El color puede representar la densidad de puntos en dicha zona o algún otro atributo.
- **Ventaja:** Permite identificar zonas de altas concentraciones de puntos de un atributo. El uso de color destaca más que el uso de área.
- **Desventaja:** Comparar la información entre 2 zonas se vuelve más difícil.



Visualización de datos geoespaciales

Choropleth map

- Muestra áreas geográficas divididas o regiones coloreadas en relación con una variable numérica. Se requiere disponer de las divisiones.
- **Ventaja:** Permite estudiar cómo evoluciona una variable a lo largo de un territorio.
- **Desventaja:** Las regiones con tamaños más grandes tienden a tener un mayor peso en la interpretación del mapa, que incluye un sesgo.

What Is the Healthy Life Expectancy After Retirement?

Average number of years of healthy life after retirement in European countries in 2022



EU-27 countries + Switzerland and Norway. Data based on average effective age of exit from the labor market and disability-free life expectancy at age 65 for both sexes.

Sources: Eurostat, OECD, Statista calculations

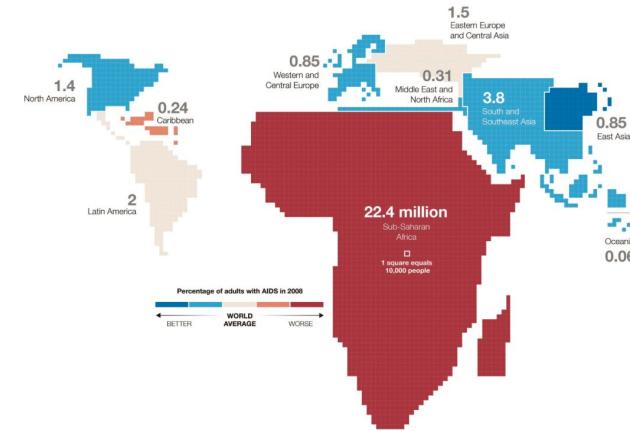


statista

Visualización de datos geoespaciales

Cartograma

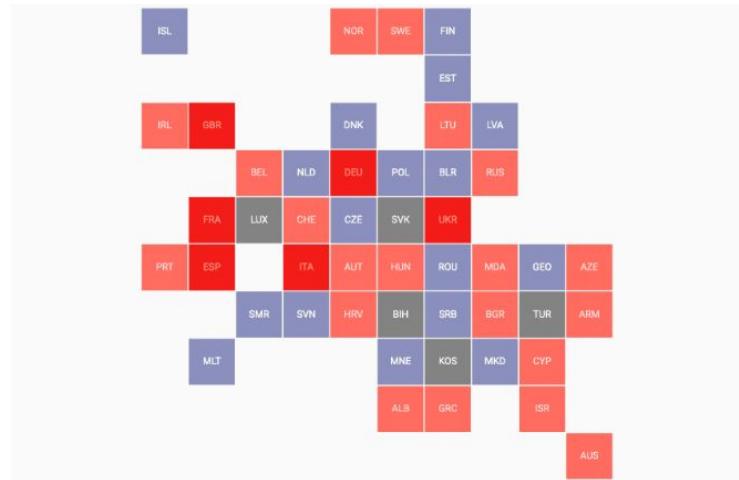
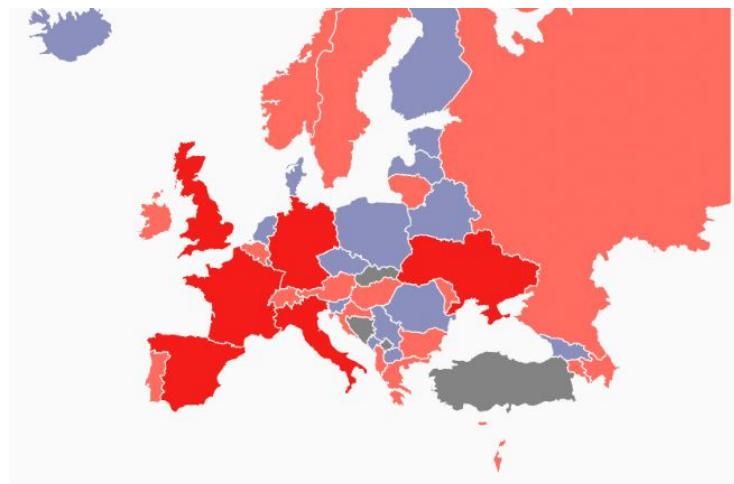
- Tipo específico de transformación de mapa, donde las regiones se redimensionan de acuerdo con una variable.
- Las regiones se colorean de acuerdo con otro atributo numérico o categórico.
- **Ventaja:** Evitan el problema de los mapas coropléticos a través de la distorsión no controlada.
- **Desventaja:** Distorsiona los límites reales y, por lo tanto, hace que el mapa sea más difícil de identificar



Visualización de datos geoespaciales

GridMap

- Se asigna una celda de igual tamaño para cada región.
 - Ya no hay distorsionar en el mapa como lo haría un cartograma o áreas de diferentes tamaños como el mapa de coropletas.



*gif

COUNTRIES WITH THE MOST PAID VACATION DAYS

Resume.io reviewed the annual statutory paid leave and paid public holidays laws in 197 countries, referring to government websites, the OECD, the International Labour Organization, etc. They then ranked the countries based on the combined number of statutory paid leave days and paid public holidays per year.

Paid leave refers to the statutory minimum annual time off bookable by the individual. Vacation days refers to the combined total of paid leave days and paid public holidays.

Top 10 Countries by Total Paid Vacation

Iran	53
Yemen	45
Bahrain	44
Bhutan	44
Madagascar	43
Niger	43
Togo	43
Azerbaijan	42
Kuwait	42
Mali	42



DATA SOURCE: [Resume.io](#)

Note:

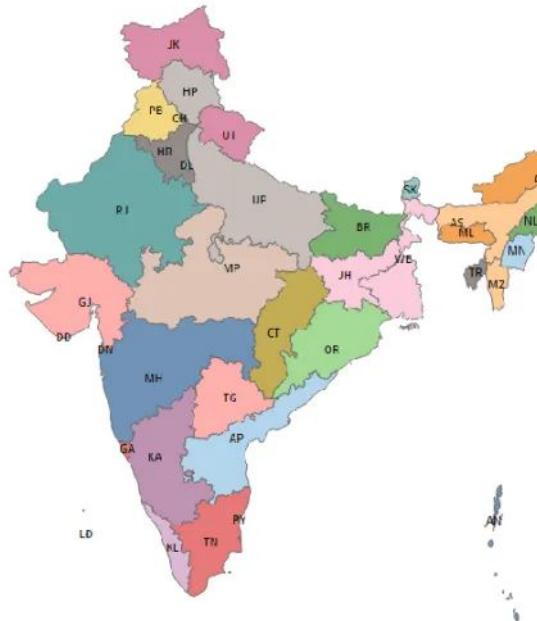
The tile/grid map mostly manages to keep nice recognisable countries shapes. But it's neither a true tile map nor a true geographical filled map.

DESIGNED BY: AAKARSH R | [@aakarsh_r6](#)

Visualización de datos geoespaciales

Mapa Hexagonal

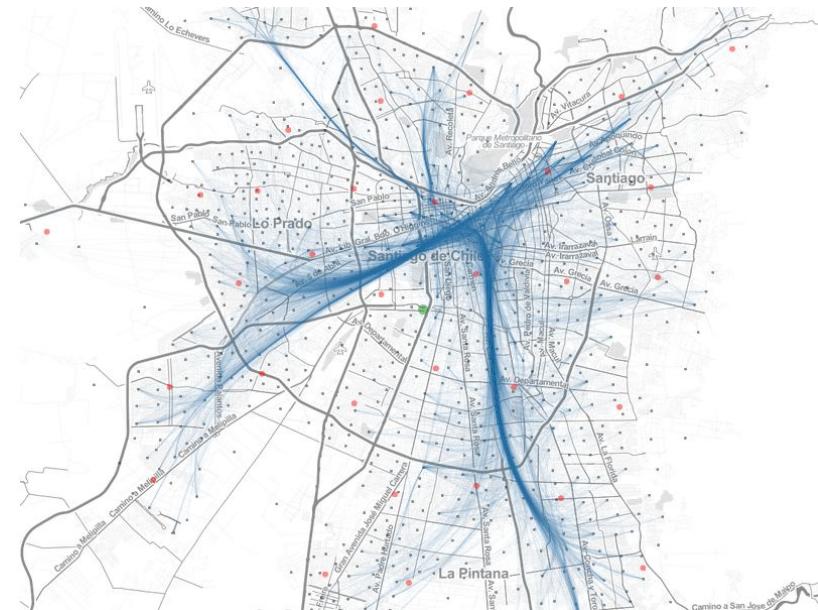
- Igual que un mapa de grilla, pero usando otra forma (hexágonos en este caso).



Visualización de datos geoespaciales

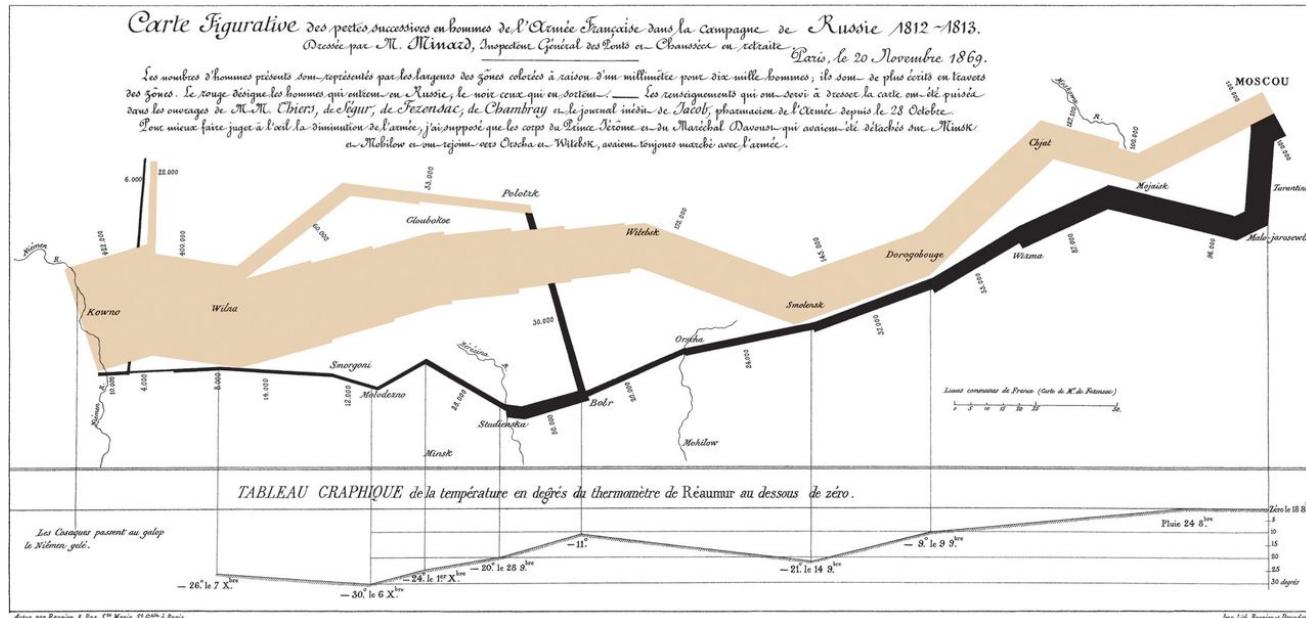
Mapa de flujo

- Muestran movimientos lineales. Se usan líneas en forma de flecha indicando dirección y sentido del flujo.
- El uso de línea permite representar el tipo de movimiento que se da y la cantidad de movimiento que se está dando.
- El ancho de las líneas generalmente codifica la cantidad de movimiento.



Ejemplo aplicado: Invasión de Napoleón

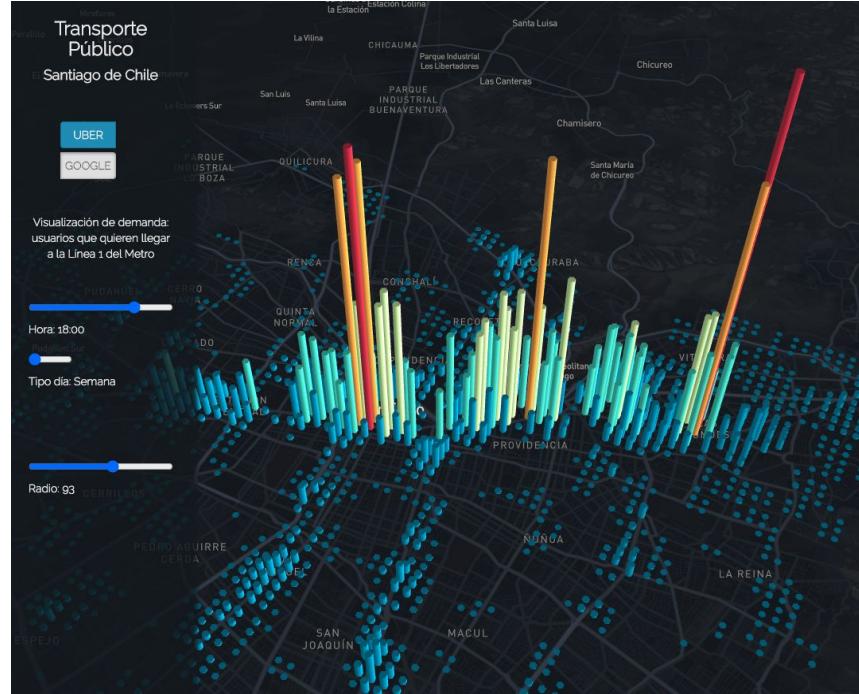
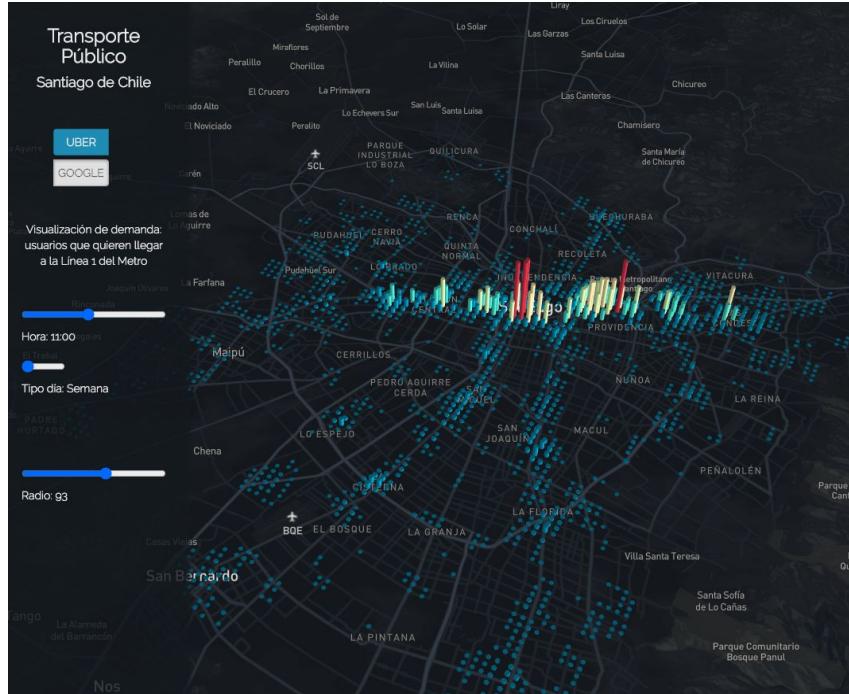
La visualización espacio-temporal más antigua y famosa es el mapa de Minard (1869) que representa la invasión de Napoleón a Rusia.



Se puede ver en el mapa de flujos información de la ida, del regreso, con eventos como batallas y otros hitos de la campaña.

Ejemplo aplicado: Transporte público

Visualización espacio-temporal de uso de estaciones de metro, comparando tráfico 11 am versus 6 pm.



Ejemplo aplicado: *Clustering* de barrios en Chile

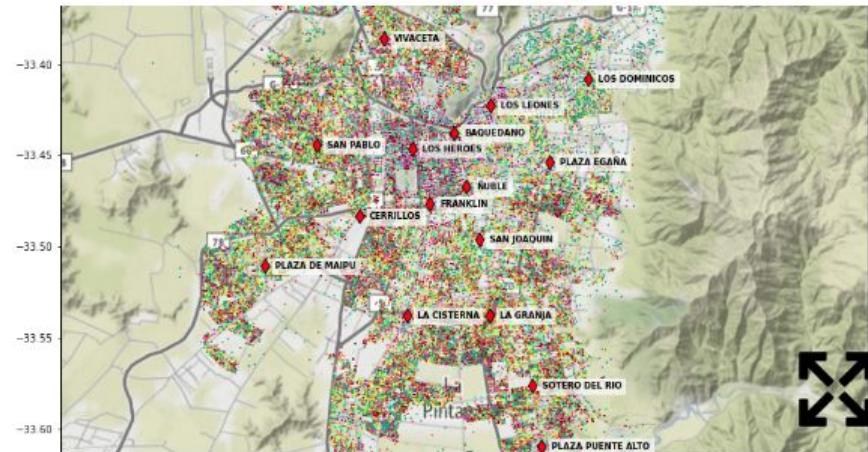
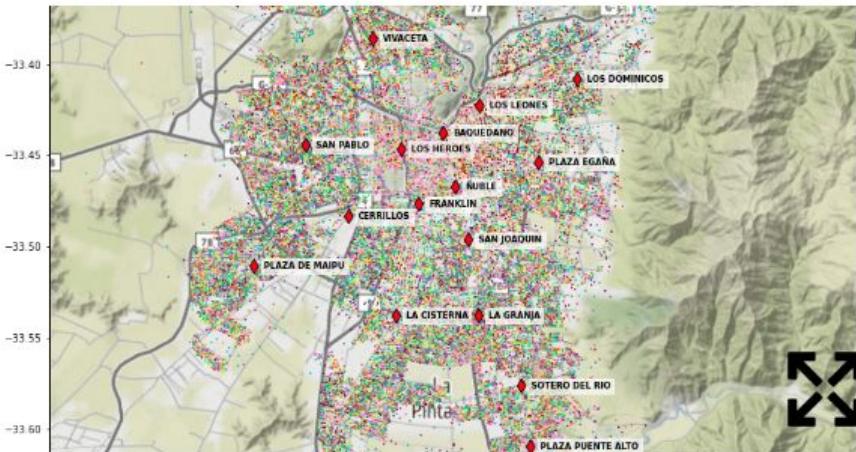
- ¿Qué define a un barrio en Chile?
- ¿Cómo podemos identificarlo automáticamente?
- ¿Qué atributos de las calles son vitales para definir un barrio?
- ¿Hay un método automatizado para identificar barrios de forma más efectiva?

Ejemplo aplicado: *Clustering* de barrios en Chile II

- Pedir a usuarios que comparan métodos de clustering automatizado mediante el uso de visualizaciones.

Vera, Camila & Lucchini, Francesca & Bro, Naim & Mendoza, Marcelo & Löbel, Hans & Gutiérrez, Felipe & Dimter, Jan & Cuchacovic, Gabriel & Reyes, Axel & Valdivieso, Hernán & Alvarado, Nicolás & Toro, Sergio. (2022). [Learning to cluster urban areas: two competitive approaches and an empirical validation](#). EPJ Data Science. 11. 10.1140/epjds/s13688-022-00374-2.

Selecciona la imagen que mejor representa los barrios de del Gran Santiago.



Visualización de texto

**Procesamientos y casos
aplicados de visualización de
texto.**

Visualización de texto

¿Por qué estudiar visualizaciones de texto?

- Porque son datos no estructurados así que no se aplica directamente todo lo que hemos visto.
- Porque es uno de los tipos de datos más comunes.
- Porque visualizar texto nos puede dar muchos “*insights*”.
- Porque es bueno saber qué tipo de “modificaciones” pudo sufrir el texto antes de ser visualizado

Visualización de texto

¿Por qué estudiar visualizaciones de texto?

	Atrib. 1	Atrib. 2	Atrib. 3	Atrib. 4	Atrib. 5	Atrib. 6	Atrib. 7	Atrib. 8
Obj. 1	0	0	1	2	7	0	1	0
Obj. 2	1	1	5	8	0	0	1	0
Obj. 3	1	1	0	0	5	9	3	1
Obj. 4	3	7	3	6	3	8	2	2

I. Loomings

C
all me blessed. Some years ago—never mind how long
precisely—having little or no money in my purse, and
nothing particular to interest me on shore, I thought I
would sail about a little and see the watery part of the world. It is
a very I have of drifting off the sphere and regulating the circulation.
Whenever I find myself growing giddy about the mouth,
whatever it may be that has been on my soul, whenever I
find myself suddenly going before circumstances, and
bringing up the rear of every friend I meet, and especially
whenever my hopes get such an upper hand of me, that it re-
quires a strong moral principle to prevent me from deliberately
stepping into the street, and publicly knocking people down
all over. I account it high time to get to sea as soon as I can.
This is my soliloquy for plot and bell. With a philosophical
flourish I call chancery upon the board. I quickly take to
the water. There was no wind, but the sun had known it,
drew it out in their depths, some clear or other, through
nearly the same feelings towards the ocean with me.

There now is just another city of the Mahabharat, belied bound
by wharves as Indian suds by coast roads—commerce surrounds
it with her net. Right and left, the streets take you watered. In
extreme down-tow in the bounces where that noble mole

is washed by waves, and cooled
by breakers, which a few hours
previous were out of sight of
land. Look at the creek of
water-gamers there.

“Circumambulate the city of
a day,” Sabatini’s afternoon.
Go from Corfu to Messina—Cor-
sica. Stop at the port there,
by Ventimiglia, northward.
What do you see?—Pointed blue
silhouettes all around the
town, stand thousands upon
thousands of ancient men fixed
in ocean invasions. Some leaning
against the spire, others leaning
against the tower, the poor brutes
some looking over the bulk
waists of ships from China;
some high aloft in the rigging,
as if trying to get a still better
survived pose. But these are all
landlubians of weak days past
up in bats and plates—not to
mention masts and benches
climbed to rocks. How then is
that? Are the green fields green?
What do they know?

The look here cause more
crowds, pacing straight for the
water, and seemingly bound
for a dry strand! Nothing
will content them but the en-
tire front of the land, loiter-
ing under the shade of
young water-lilies will not
suffice. No. They must get into
as high the water as they pos-
sibly can without falling in, and
then they stand under a
green canopy. Indeed all
they come from boats and al-
leys, alleys—east, east, south, and
west. Set here they all smile.
Tell me, does the magnetic
value of the needles of the
compasses of these ships at-
tract them so much?

Once more, for you are in
the country, in some high land
of lakes. Take almost any path
you please, and in one it carries you down in a ditch, and
lays you there by a pool in the stream. There is magic in
the earth. Let the most desolate
of spots, and the most desolate
of men—stand in his deep
water reverse—stand that man
on his legs, set his feet a going,
and he will infallibly lead you
to water, if water there be in all
that region. Should you ever be
aboard in the great American:

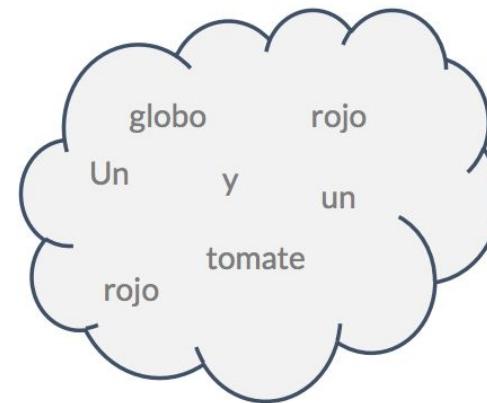
72

Visualización de texto

¿Cómo estructuramos un poco el texto?

Representamos un documento como una **bolsa de palabras**, sin considerar el orden de éstas

Un tomate rojo y un globo rojo



Visualización de texto

¿Cómo estructuramos un poco el texto?

Podemos representar la bolsa de palabras de forma numérica, en una matriz

Documento 1: Un auto rojo

Documento 2: Un tomate rojo y un globo rojo.

Documento 3: Un plátano amarillo y un tomate verde.

	1	2	3	4	5	6	7	8	9
Doc 1	0	1	0	0	1	0	1	0	0
Doc 2	0	0	1	0	2	1	2	0	1
Doc 3	1	0	0	1	0	1	2	1	1

ID	palabra
1	amarillo
2	auto
3	globo
4	plátano
5	rojo
6	tomate
7	un
8	verde
9	y

Visualización de texto

Pre-procesamiento típicos en el texto

- Normalizar (pasar todo a minúsculas, todo mayúsculas).
- Eliminación de *stop-words* (eliminar ciertas palabras).
- Stemming o Lematización (llevar un conjunto de palabras distintas a una igual).
- N-grams (separar la frase en conjuntos de N palabras).
- TF-IDF para no utilizar frecuencias.

Visualización de texto

Stopword

- Consiste en **eliminar del documento palabras muy comunes** que no aportan información.
- Estas palabras ocurren en casi todos los documentos, y puede ser preferible eliminarlas.

Sample text with Stop Words	Without Stop Words
GeeksforGeeks – A Computer Science Portal for Geeks	GeeksforGeeks , Computer Science, Portal ,Geeks
Can listening be exhausting?	Listening, Exhausting
I like reading, so I read	Like, Reading, read

Visualización de texto

Stemming

- Conservamos la raíz (literal) de la palabra
- Cambio, cambios, cambió, cambiando, cambiado → **Cambi**

Lematización

- Hacer una reducción apropiada a una palabra raíz (más inteligente).
- Soy, es, eres, somos, son → **SER**

Visualización de texto

N-gramas

• N = 1 : Esta es una oracion

unigramas

- Esta,
- es,
- una,
- oracion

• N = 2 : Esta es una oracion

bigramas

- Esta es,
- es una,
- una oracion

• N = 3 : Esta es una oracion

trigramas

- Esta es una,
- Es una oracion

Visualización de texto

TF-IDF

- TF : *Term Frequency* : Número de veces que está la palabra
- DF : *Document Frequency* : Número de documentos que tiene la palabra

Calculamos $TF * 1/DF$

Documento 1: Un auto rojo

Documento 2: Un tomate rojo y un globo rojo.

Documento 3: Un plátano amarillo y un tomate verde.

	1	2	3	4	5	6	7	8	9
Doc 1	0	1	0	0	1	0	1	0	0
Doc 2	0	0	1	0	2	1	2	0	1
Doc 3	1	0	0	1	0	1	2	1	1

ID	palabra
1	amarillo
2	auto
3	globo
4	plátano
5	rojo
6	tomate
7	un
8	verde
9	y

Visualización de texto

TF-IDF

- TF : *Term Frequency* : Número de veces que está la palabra
- DF : *Document Frequency* : Número de documentos que tiene la palabra

Calculamos $TF * 1/DF$

Documento 1: Un auto rojo

Documento 2: Un tomate rojo y un globo rojo.

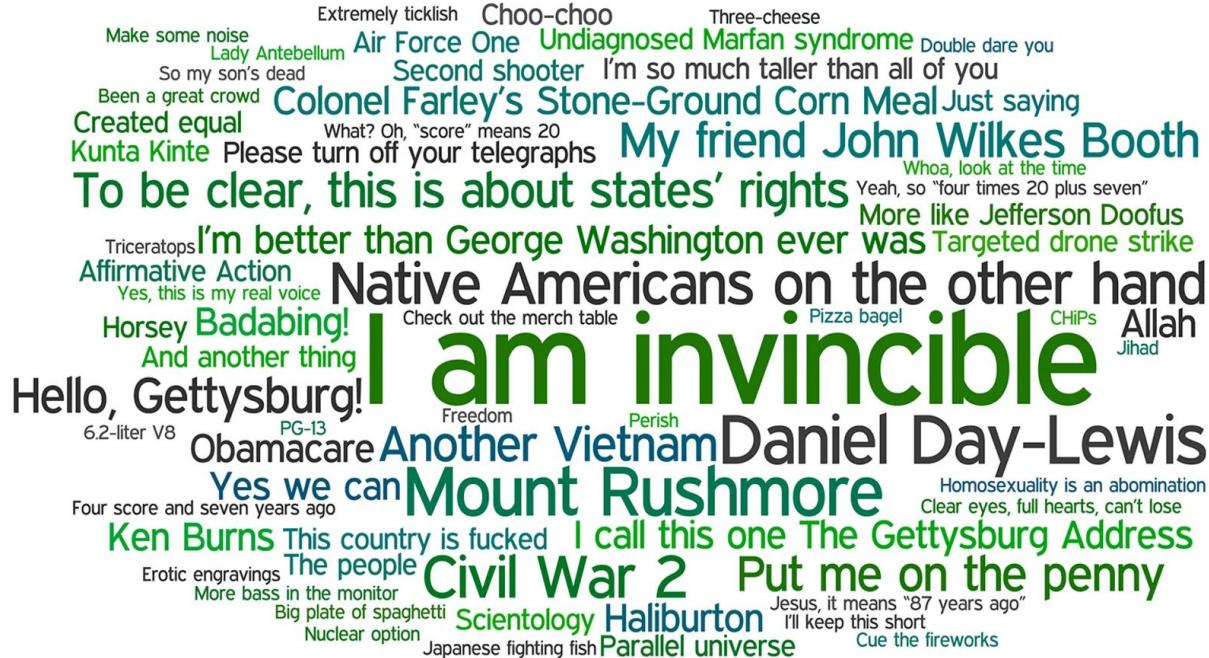
Documento 3: Un plátano amarillo y un tomate verde.

	1	2	3	4	5	6	7	8	9
Doc 1	0	1 (1)	0	0	0.5 (1)	0	0.33 (1)	0	0
Doc 2	0	0	1 (1)	0	1 (2)	0.5 (1)	0.66 (2)	0	0.5 (1)
Doc 3	1 (1)	0	0	1 (1)	0	0.5 (1)	0.66 (2)	1	0.5 (1)

ID	palabra
1	amarillo
2	auto
3	globo
4	plátano
5	rojo
6	tomate
7	un
8	verde
9	y

Visualizaciones: Nubes de palabras

Se pueden ocupar
N-gramas y no sólo 1
palabra.



Visualizaciones: Nubes de palabras

Tambien, se pueden ordenar las palabras y agregar más elementos visuales.

Democratic Verbs

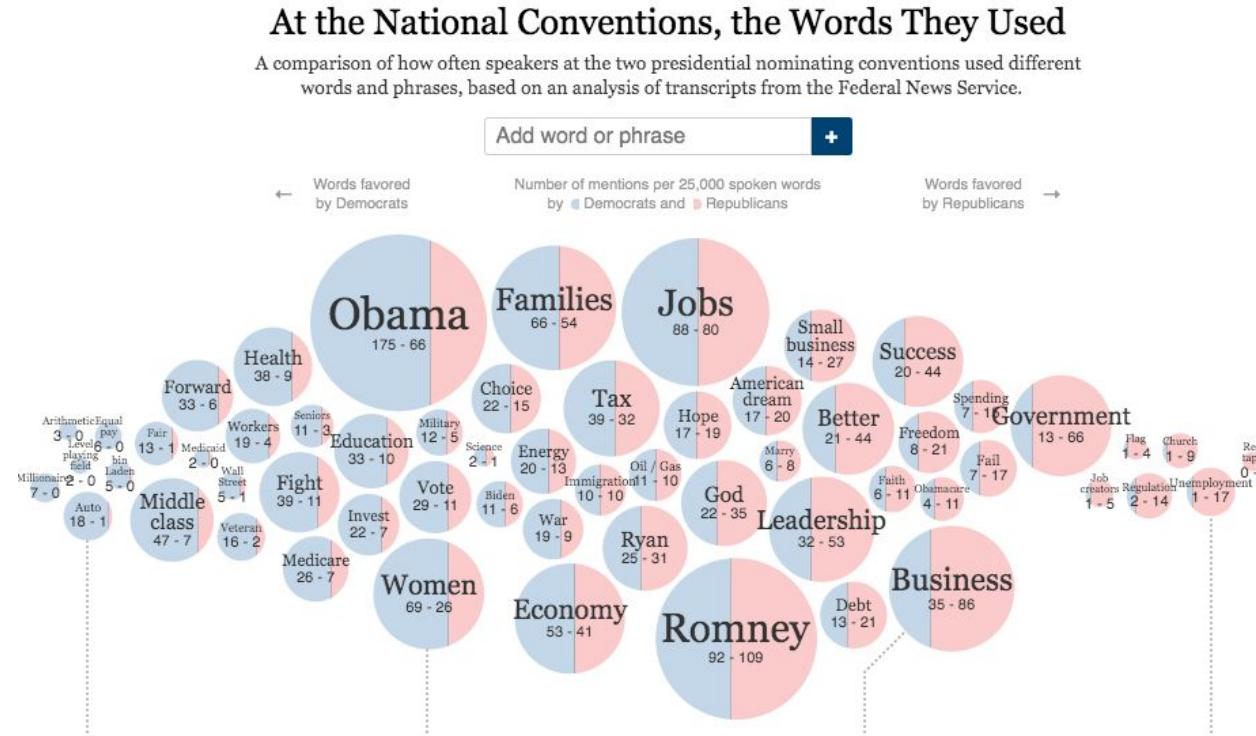
think⁹⁷ going⁷⁸ want⁷⁷
need⁷¹ did⁶⁶ get⁶¹ know⁶⁰
been⁵⁵ say⁵¹ had⁵¹ said⁴⁸ make⁴⁷
let⁴⁴ were⁴² go³⁹ believe³⁸ thank³⁵ take³²
look²⁹ does²⁸ got²⁸ talking²³ am²³ done²² put²² being²⁰
voted²⁰ come²⁰ made¹⁸ saying¹⁷ bring¹⁷ says¹⁵ doing¹⁵ support¹⁴
address¹⁴ move¹⁴ deal¹³ getting¹³ gonna¹³ talk¹³ stand¹² respond¹² coming¹²
trying¹² looking¹¹ having¹¹ hear¹¹ happen¹¹ help¹¹ heard¹¹

Republican Verbs

think¹⁴⁶ want¹⁴²
know¹³⁹ going¹²¹ said¹⁰⁸
need¹⁰⁶ thank⁹⁸ get⁹¹ did⁸⁸
say⁸¹ make⁸⁰ go⁷⁸ been⁷⁸ had⁶⁷ let⁵¹
put⁵⁰ does⁴⁹ take⁴⁸ done⁴⁵ were⁴⁵ tell⁴⁴ like⁴¹
talking³⁸ says³⁷ believe³⁵ got³⁴ give³⁴ look²⁹ come²⁸ see²⁷
respond²⁷ talk²⁶ bring²⁵ made²⁵ saying²⁴ am²³ use²² fight²¹ coming²⁰
went¹⁹ called¹⁸ being¹⁸ quote¹⁸ live¹⁸ agree¹⁷ mean¹⁷ ask¹⁷ doing¹⁷ work¹⁶
having¹⁶

Visualizaciones: *Bubble Cloud*

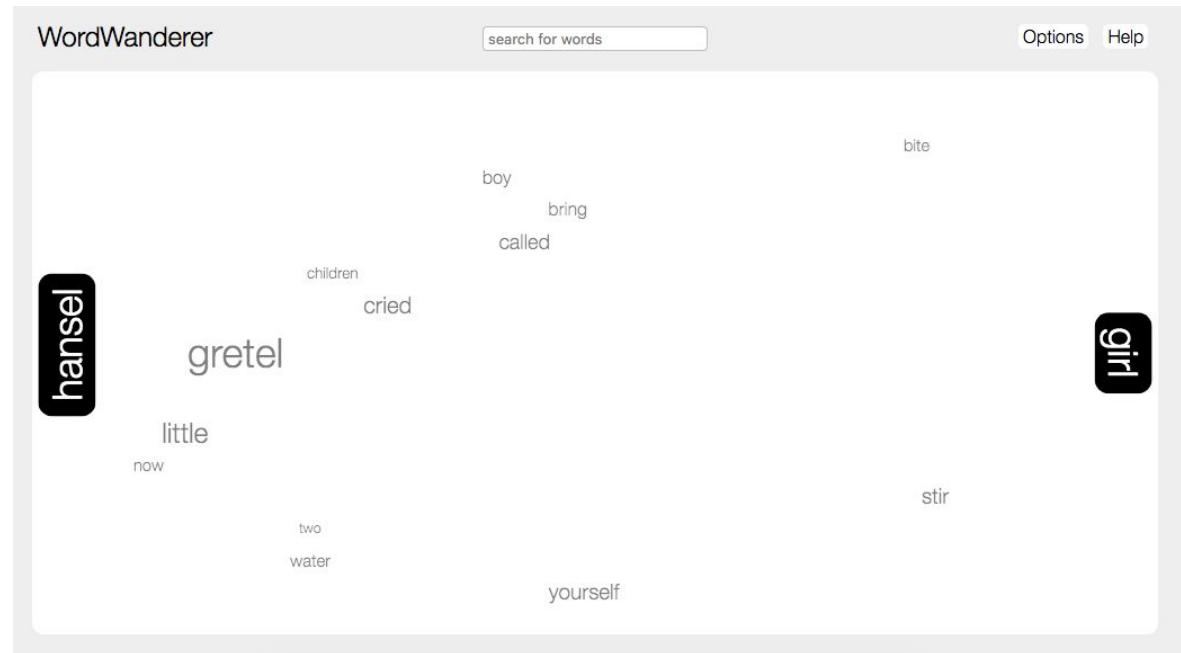
Se pueden mezclar con otros canales para entregar más información.



Fuente: At the National Conventions, the Words They Used - Interactive Feature - NYTimes.com

Visualizaciones: *WordWanderer*

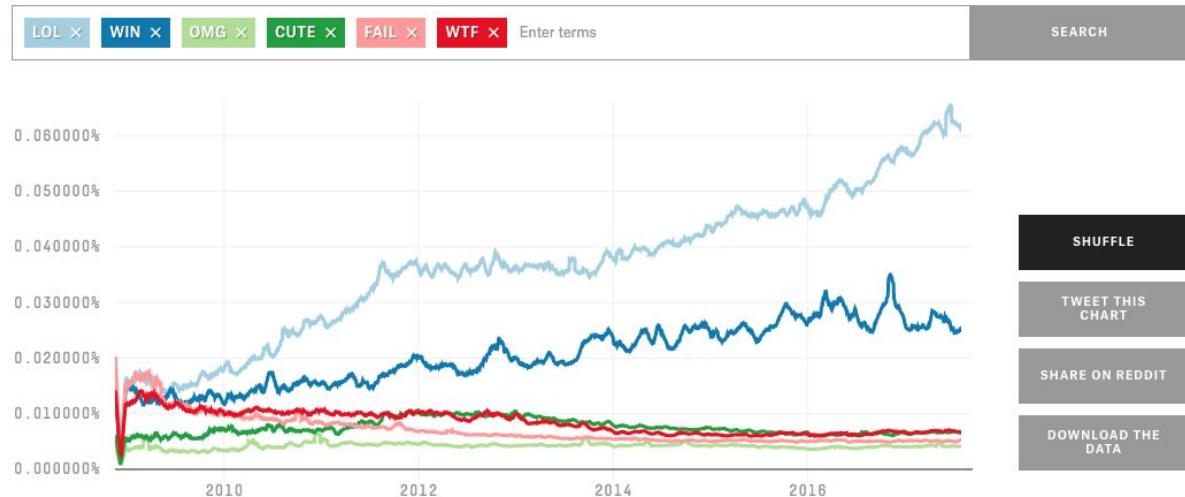
Uso de la posición cómo *encoding* de alguna información



Visualizaciones: *How The Internet* Talks*

Cómo cambia la búsqueda de ciertas palabras en el tiempo.

(datos hasta septiembre del 2017)



Visualizaciones: SHAP - *Text visualization*

Explicar por qué un modelo indica que este comentario es positivo en función de las palabras que contiene.

Azul: palabras negativas

Rosado: palabras positivas

```
[10]: # plot the first sentence's explanation  
shap.plots.text(shap_values[3])
```



This is easily the most underrated film inn the Brooks cannon. Sure, its flawed. It does not give a realistic view of homelessness (unlike, say, how Citizen Kane gave a realistic view of lounge singers, or Titanic gave a realistic view of Italians YOU IDIOTS). Many of the jokes fall flat. But still, this film is very lovable in a way many comedies are not, and to pull that off in a story about some of the most traditionally reviled members of society is truly impressive. Its not The Fisher King, but its not crap, either. My only complaint is that Brooks should have cast someone else in the lead (I love Mel as a Director and Writer, not so much as a lead).

Taller No evaluado

Solución en vivo



Visualización de Información y Analítica Visual

— Hernán Valdivieso López (hfvaldivieso@uc.cl) —
