

Visualización de Datos

Conceptos e introducción

Visualización de datos:

- 1. ¿Qué es y para qué sirve?**
- 2. Historia**
- 3. Teoría de la visualización de datos**
- 4. Herramientas**
- 5. Productos generados con visualizaciones de datos**
- 6. Creación de visualizaciones**
 - 1. El proceso**
 - 2. El diseño**
 - 3. La publicación**
- 7. Errores de la visualización de datos.**

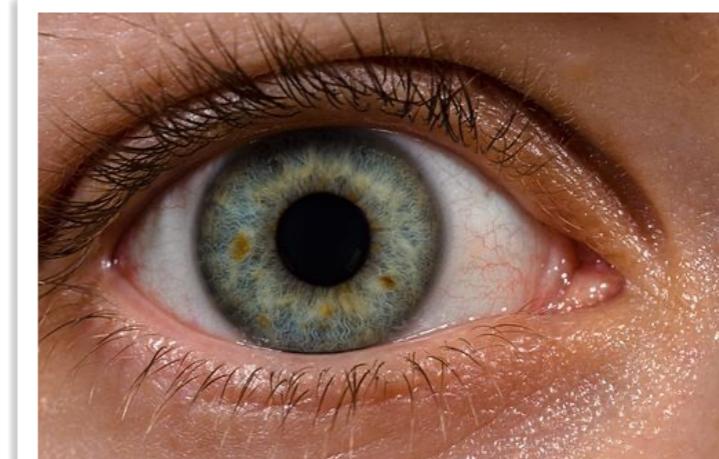
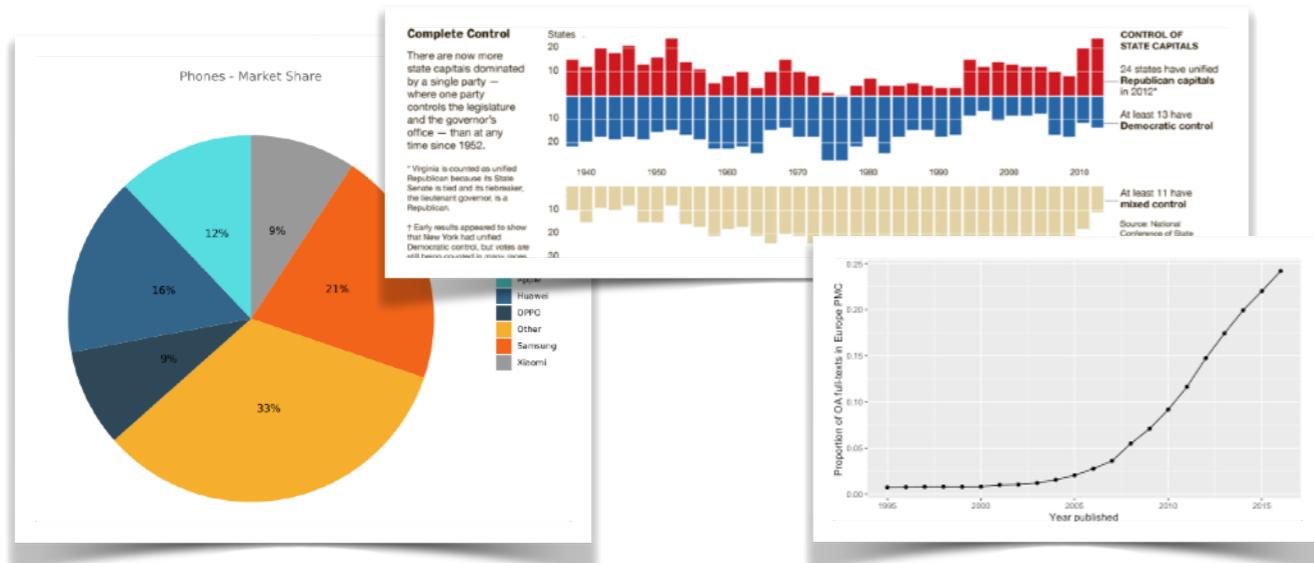
¿Qué es la visualización de datos?

1. Acción y efecto de visualizar datos. (Segasi).
2. Representación gráfica de la información y de los datos (Tableau).
3. Hacer visibles y comprensibles al ser humano aspectos y fenómenos de la realidad que no son accesibles al ojo humano, cuyo conocimiento se encuentra dentro de una base de datos.

¿Por qué necesitamos visualización de datos?

Naturaleza

1. Porque tenemos **límites cognitivos** que nos dificultan procesar, interpretar o recordar la información.
2. Somos **seres visuales**; es mejor entender los datos de manera gráfica, en vez de tenerlos en tablas.



¿Por qué necesitamos visualización de datos?

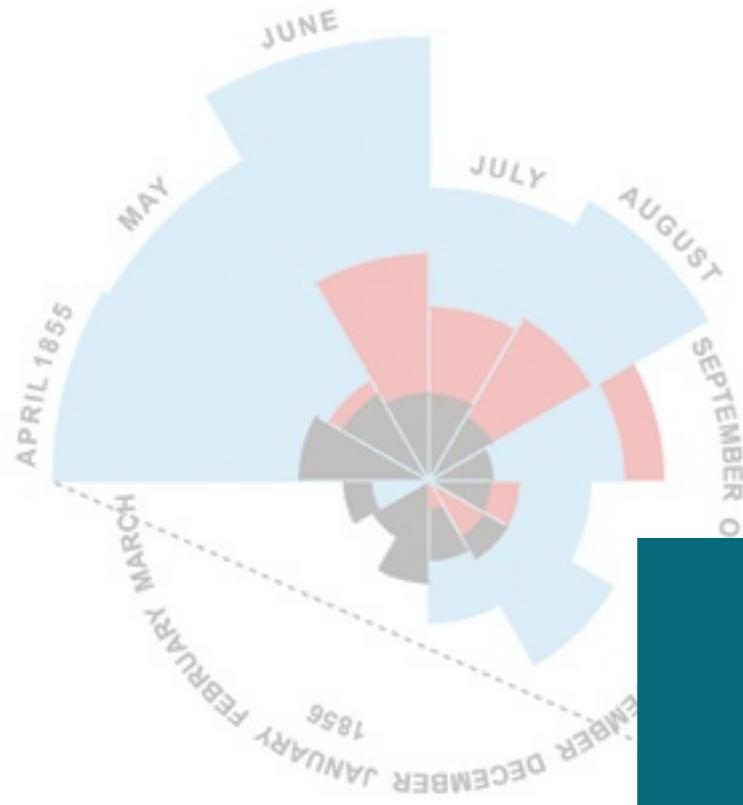
Contexto actual

Porque vivimos en la **era de la información**, y requerimos métodos para analizar grandes cantidades de datos.

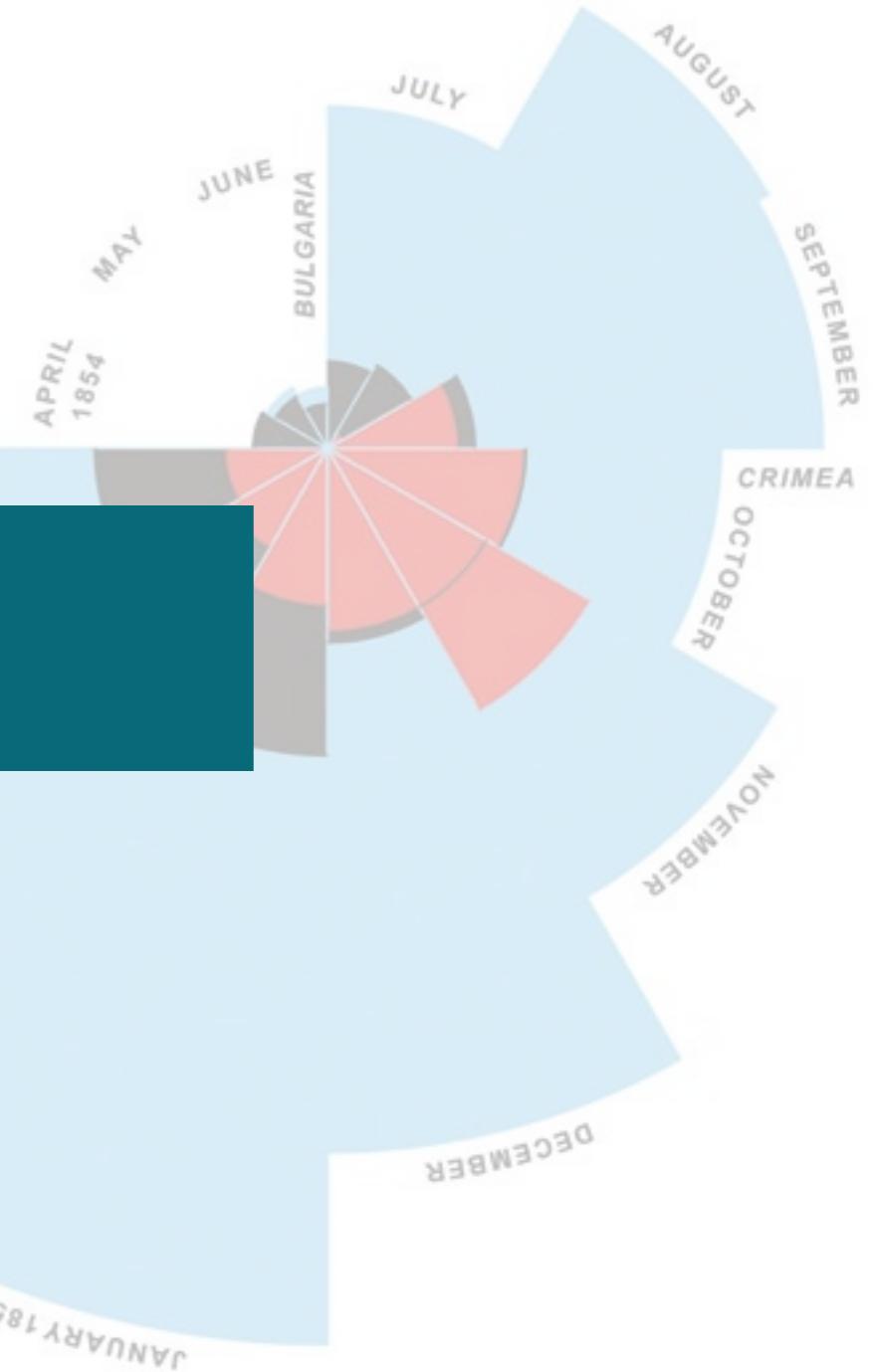


DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.



1.
APRIL 1854 TO MARCH 1855.



Historia

The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic Diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes

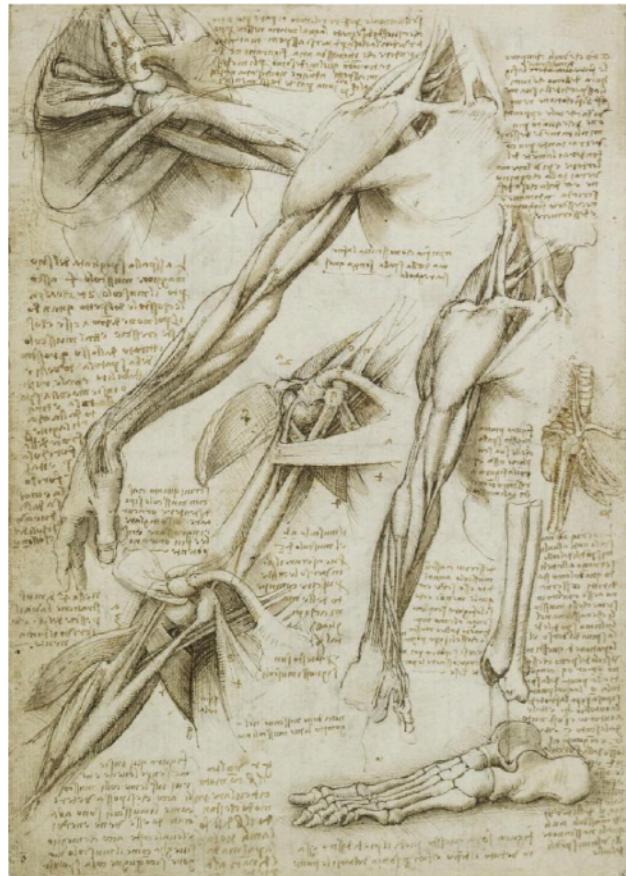
The black line across the red triangle in Nov' 1854 marks the boundary of the deaths from all other causes during the month

In October 1854, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black

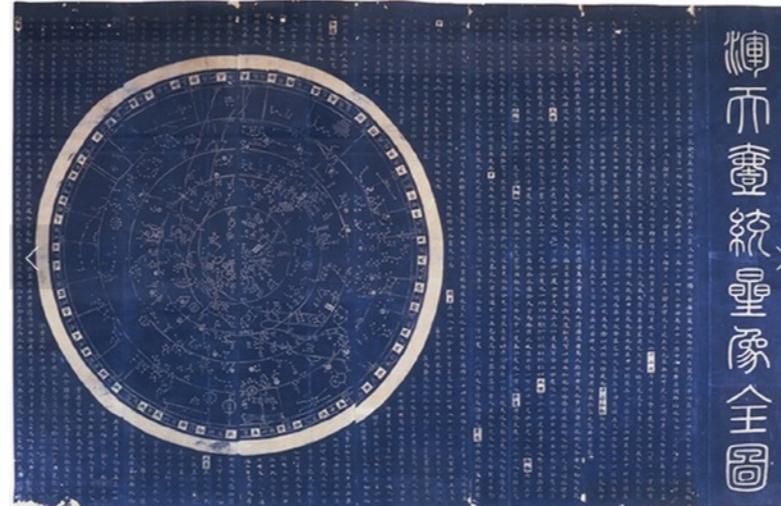
The entire areas may be compared by following the blue, the red & the black lines enclosing them. ©hugh-small.co.uk

Historia

Las visualizaciones surgieron para **registrar datos**.



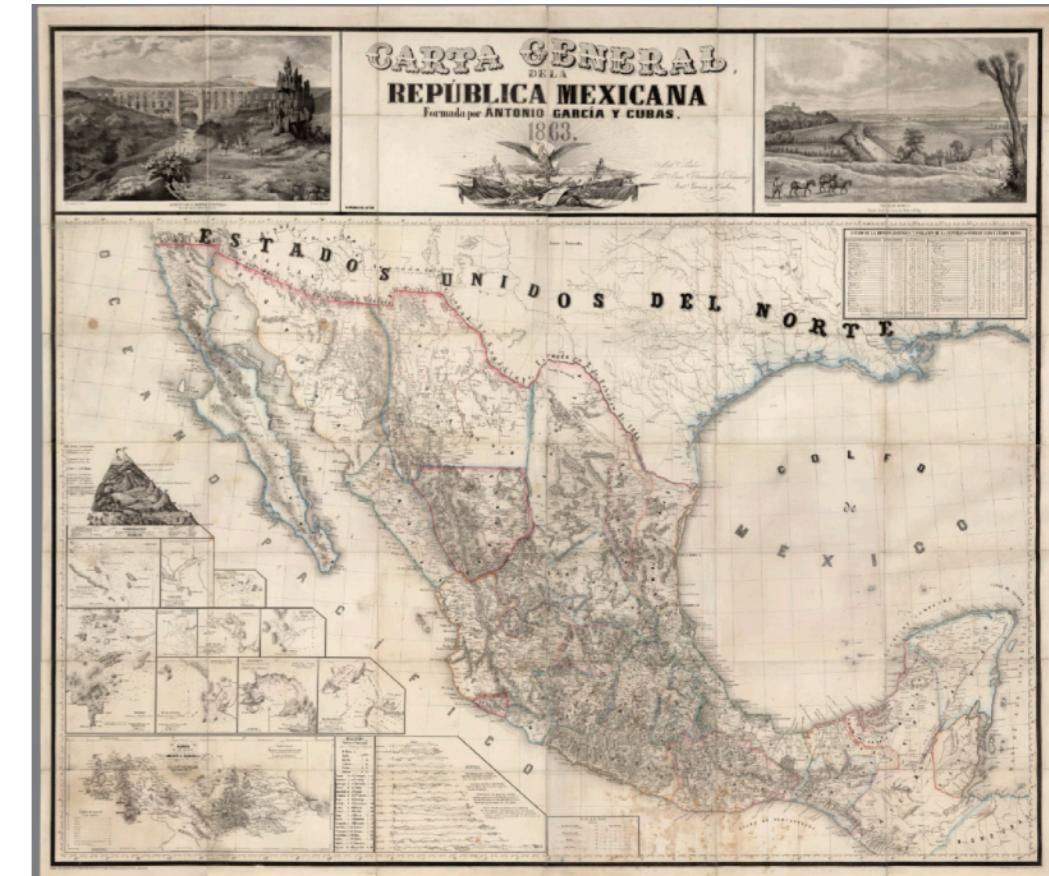
Da Vinci, 1510-1511



Carta estelar china



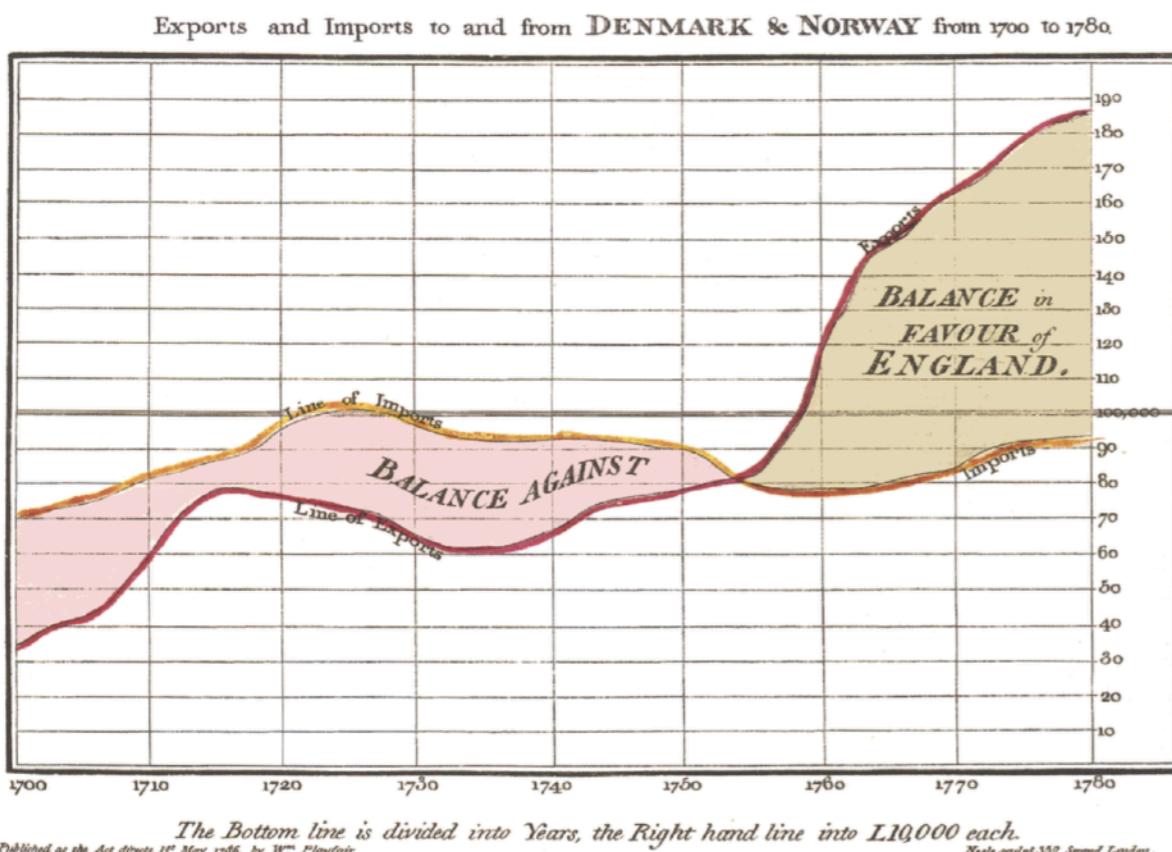
John Ogilby, el camino de Londres a Aberystwith



Carta general de la república mexicana

Historia

Las visualizaciones surgieron para **analizar datos**.

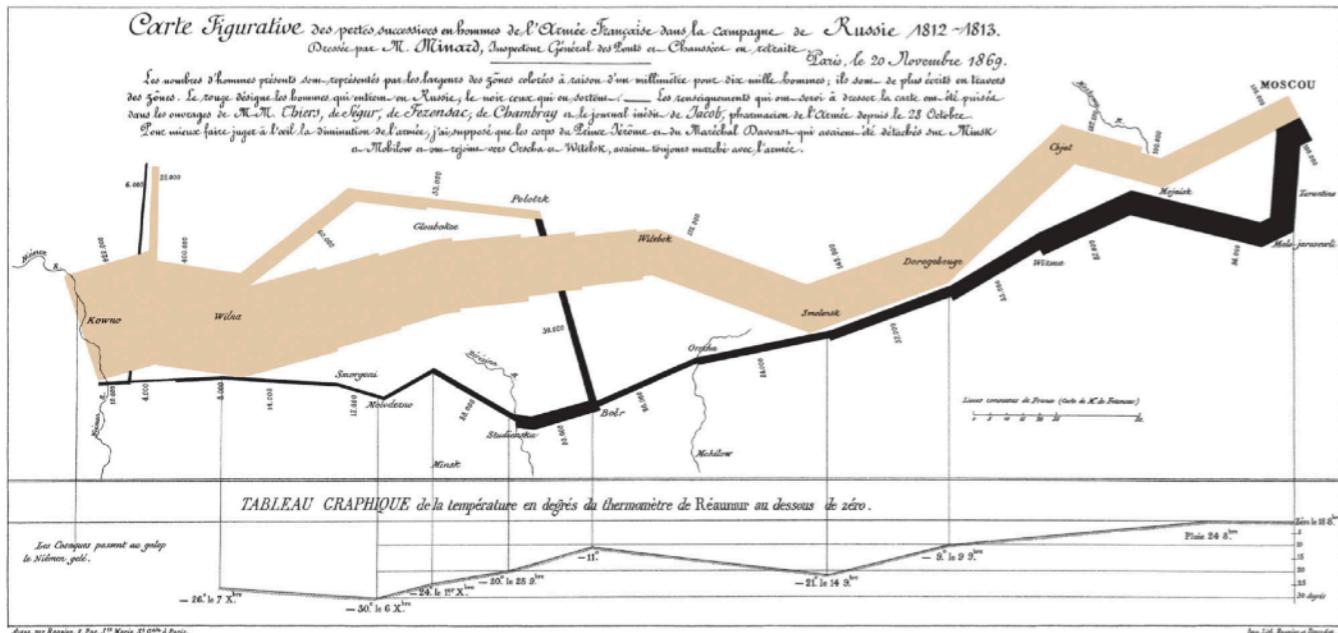


William PlayFair (1786)



John Snow (1854)

Las visualizaciones surgieron para **comunicar**.

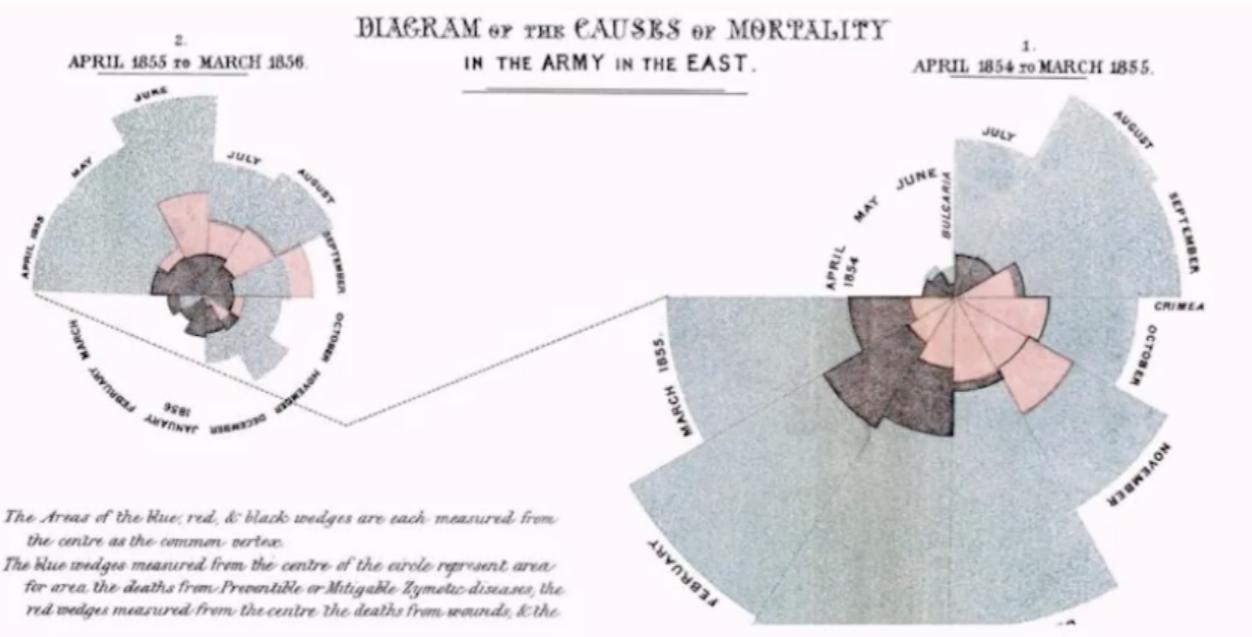


Charles Joseph Minard
(1869)



Railways of London
Harry Beck, 1933

Historia



Caso de Florence Nightingale

→ Position

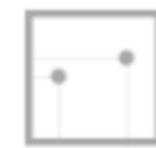
→ Horizontal



→ Vertical



→ Both



→ Color



→ Shape



Teoría

→ Tilt



→ Size

→ Length



→ Area



→ Volume



Figure 5.3. Visual channels control the appearance of marks.

Tipos de datos a representar

Según Segasi, los tipos de datos a representar en visualizaciones son los siguientes:

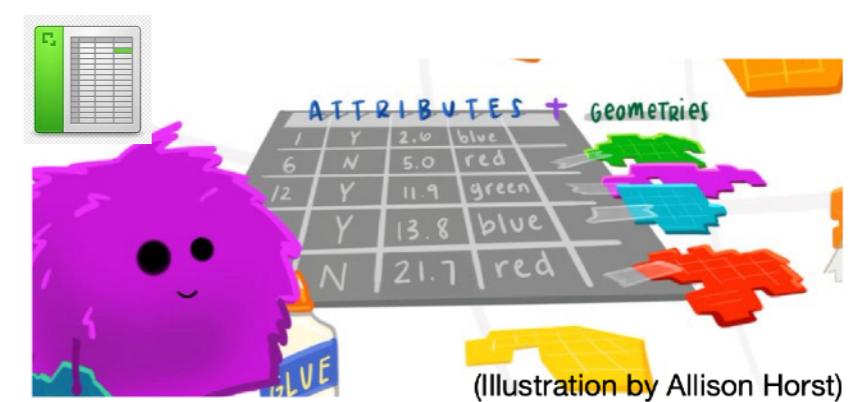
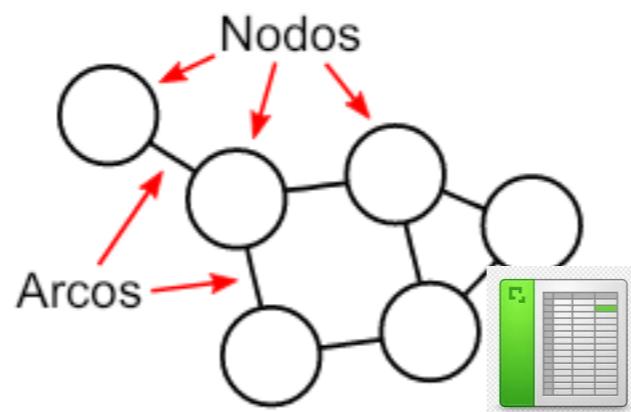
1. **Unidades** (entidades individuales discretas).
2. **Variables** (propiedades medibles u observables).
3. **Enlaces** (relación entre unidades/nodos).
4. **Posición** (ubicación en 2D o 3D).

Bases de datos

Base de datos: Conjunto de información que es objeto de análisis.

Estas pueden ser:

- **Tablas** (unidades @ renglones y variables @ columnas)
- **Redes y árboles** (Unidades o nodos, variables y enlaces).
- **Geográficas** (Unidades, variables, posición y geometrías).



Unidad de observación

Es la unidad descrita por los datos que uno analiza. Es el **individuo o sujeto** sobre el cual habla cada renglón de la base de datos.

Por ejemplo:

- Un municipio en un año determinado
- Un estado en un año determinado
- Personas parte de un experimento
- Individuo representativo dentro de una encuesta

Tipos de variables



Elementos de la visualización

Los elementos de la visualización de datos se pueden clasificar en dos: **marcas** y **canales**.

Las **marcas** son los **elementos gráficos básicos** o primitivos en una imagen.

Los **canales** visuales son los elementos que nos permiten **controlar la apariencia** de las marcas.

Marcas

Las **marcas** son los **elementos gráficos básicos** o primitivos en una imagen. Se clasifican de acuerdo con:

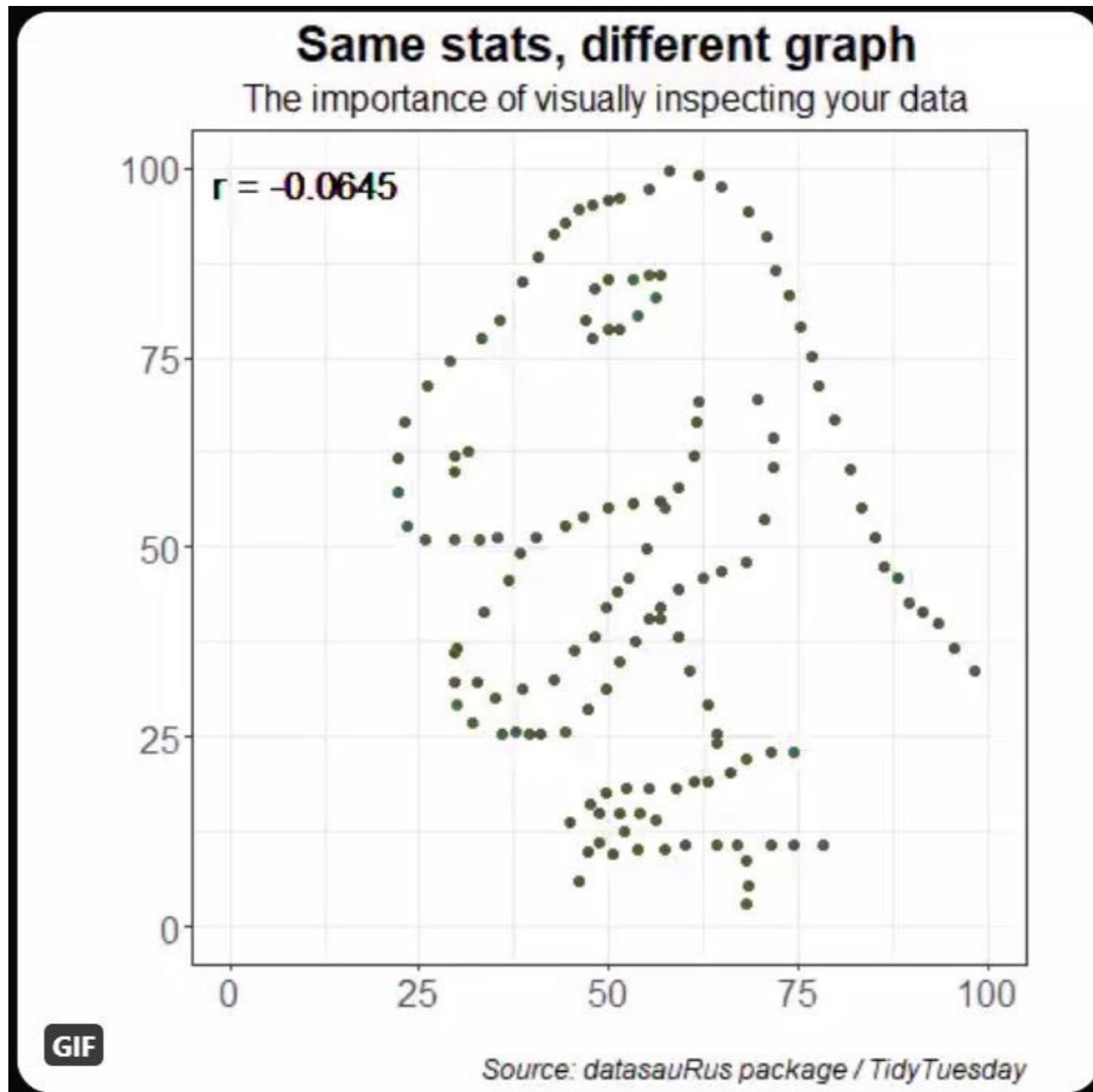
- 1) El **tipo de dato** que representan
- 2) El **número de dimensiones** que requieren

Marcas para **unidades** y **nodos**:



En ggplot2() en R, a las marcas se les conoce como geometrías.

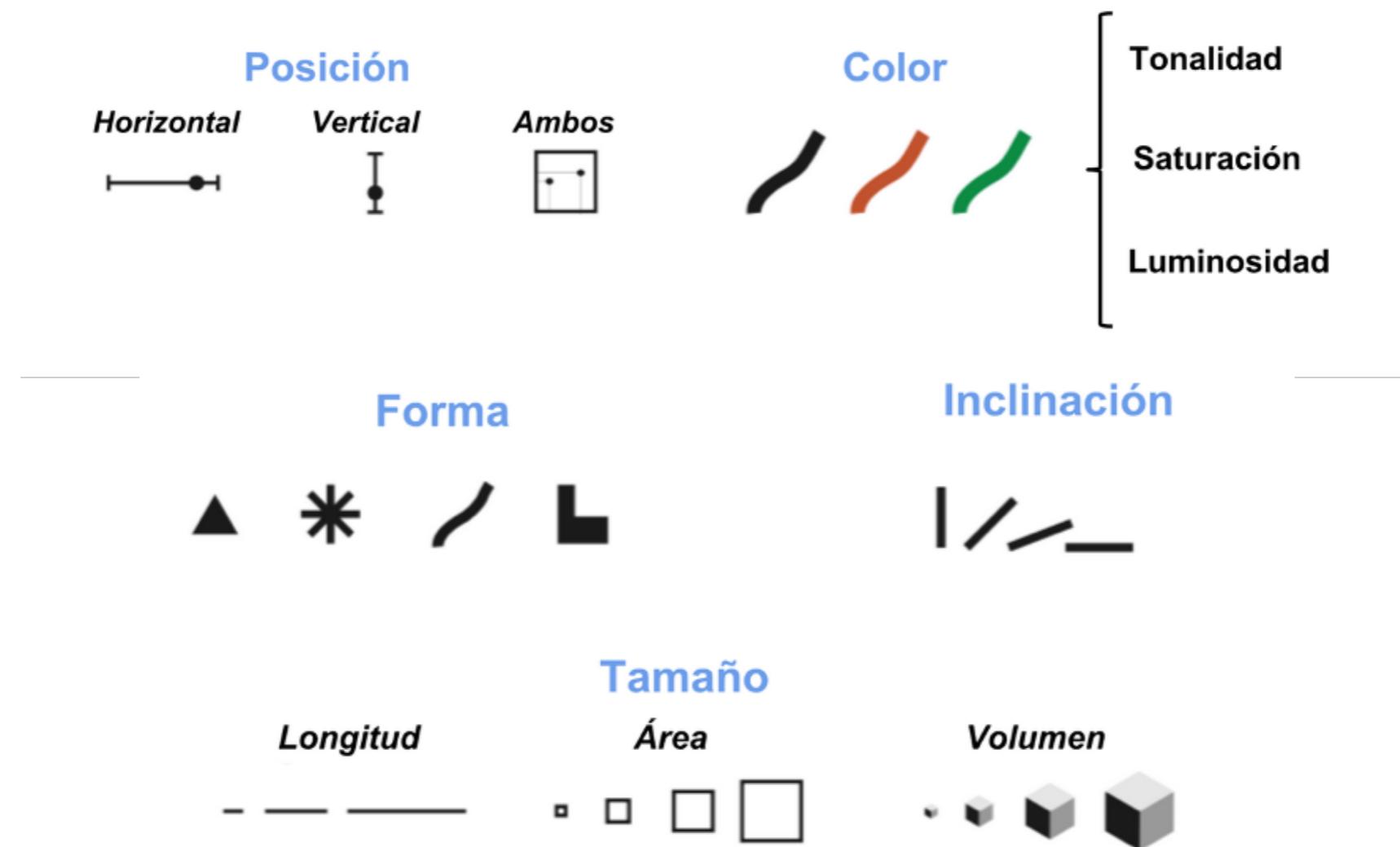
Marcas



En `ggplot2()` en R, a las marcas se les conoce como geometrías.

Canales

Los **canales** visuales son los elementos que nos permiten **controlar la apariencia** de las marcas.



En `ggplot2()` en R, a las marcas se les conoce como elementos estéticos o *aesthetics*.



Publicación de las visualizaciones

Las visualizaciones tienen como fin presentarse en estos medios:

1. En una **infografía**
2. En un **artículo periodístico**
3. Como gráfica **sola** (por ejemplo, en RRSS).
4. En **artículos científicos**
5. En **dashboards**
6. En una **página web**
7. En **video**

Uso de visualización estática o interactiva

Gráfica estática



Gráfica interactiva



- Las gráficas estáticas son útiles cuando no tienes un medio web para presentarlas o cuando las vas a imprimir en papel.

- Las gráficas interactivas se recomiendan cuando se van a presentar en un **medio web** y se quiere mostrar más información que la que se podría mostrar en un medio impreso.

Infografía

Imagen explicativa que combina texto, ilustración y diseño, cuyo propósito es sintetizar información de cierta complejidad e importancia, de una manera directa y rápida.



Artículo periodístico

En el caso de visualizaciones que se van a mostrar en prensa, se requiere que sigan una **identidad gráfica** y que tengan un buen **diseño gráfico** en general.

¿QUÉ SE DIJO EN LA FINAL?

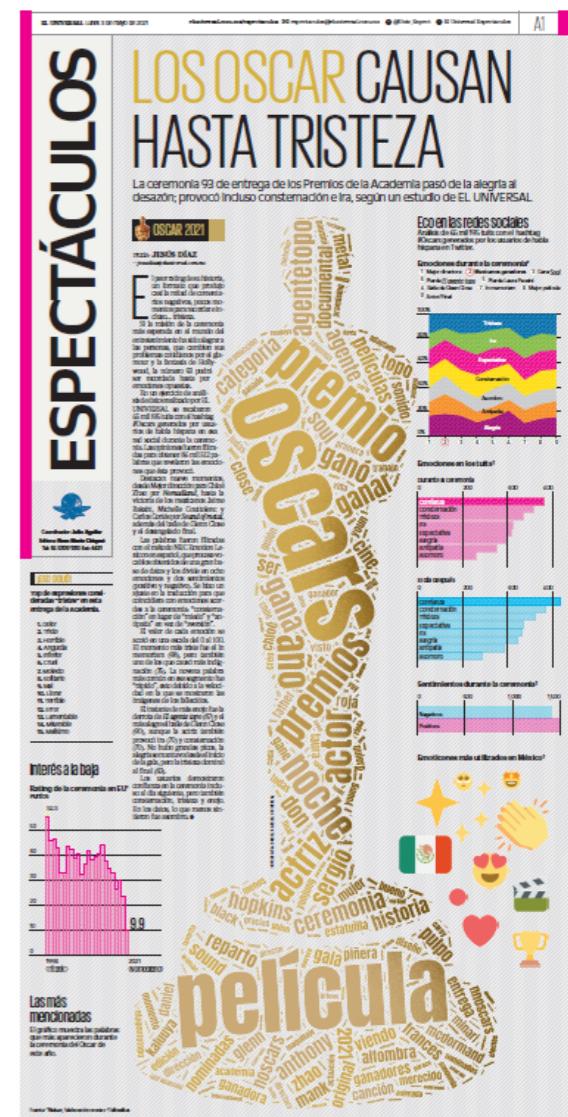
En 23 años, Cruz Azul perdió en seis finales, por lo que los aficionados mostraron el temor de volviera a suceder una derrota. Estas son algunas palabras que se dieron a la par de la más pronunciada: "Campeón".

El 30 de mayo de 2021 será un día imborrable en la memoria de los fans del Cruz Azul. Más de quiso dejar claro su sentir en las redes sociales; los nervios, la emoción, el coraje y hasta el buen humor fueron sentimientos reflejados durante las casi dos horas que duró el encuentro... durante las cuales se publicaron 29,953 tuits con la palabra "Cruz Azul".



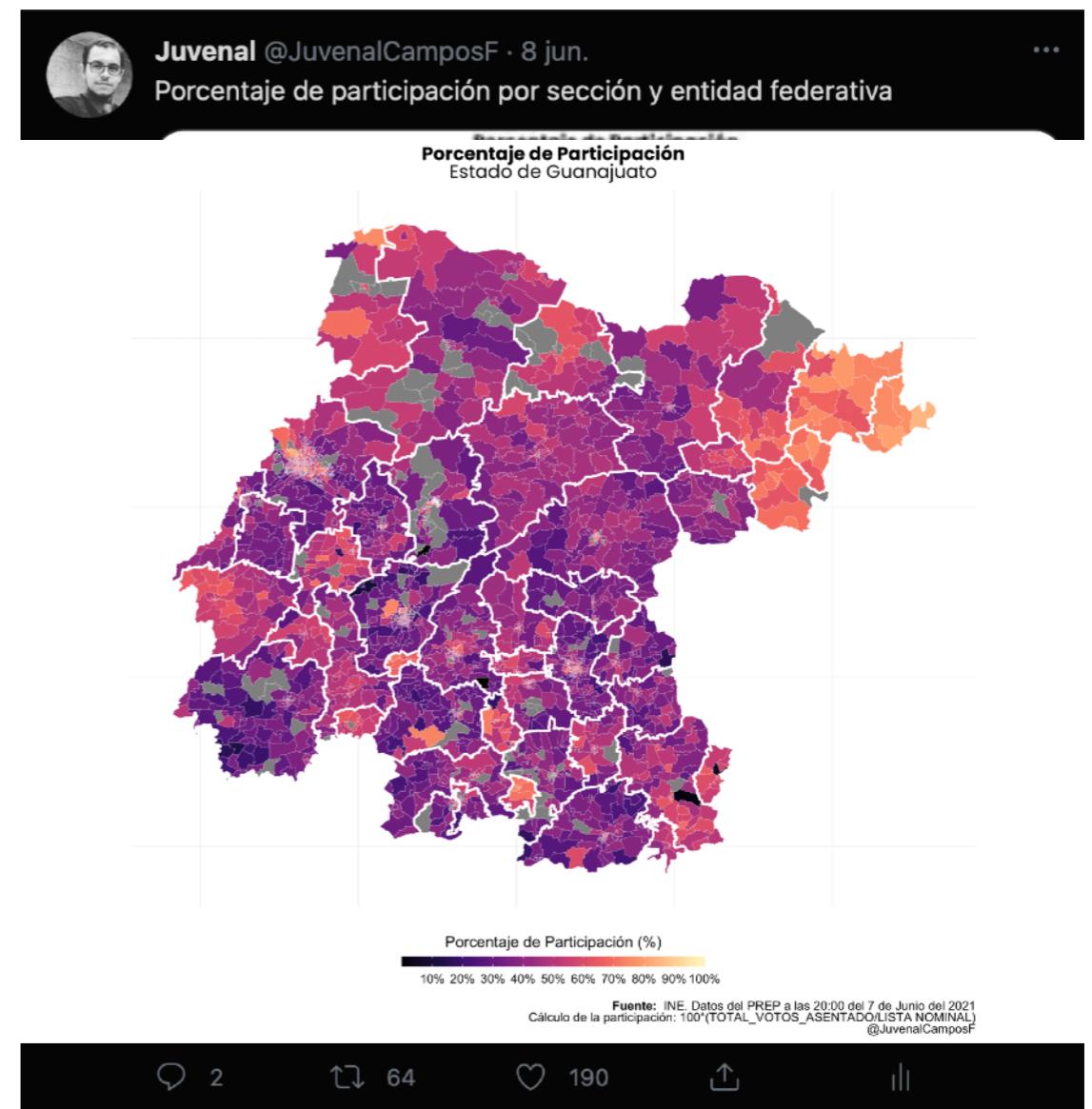
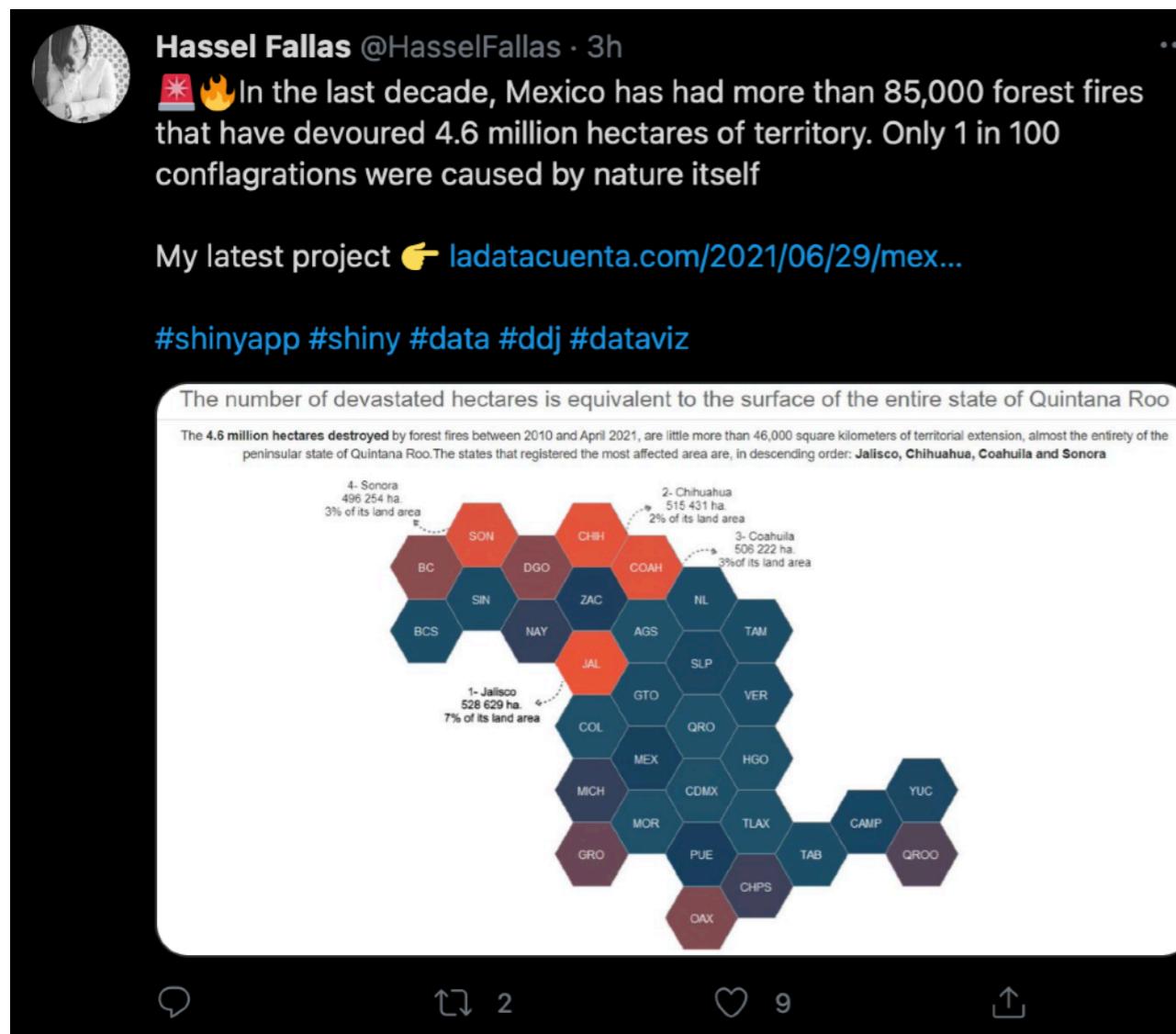
Quitando el nombre del equipo, la palabra más mencionada de la noche fué "campeón" (2,468 veces), seguida de "gana" (2,101) y, por supuesto, "go!" (1,839). De hecho, éste se gritó fuerte: "go!" (dicho 1,839 veces), "goof!" (20), "goool!" (43), e incluso "goooooool!", así con seis ceros, dicho más de 40 veces; sin olvidar el "gozalo" (220) y sus variantes.

Cruz de las 200 palabras más comunes.
Faltaron en 83% de los Díaz.



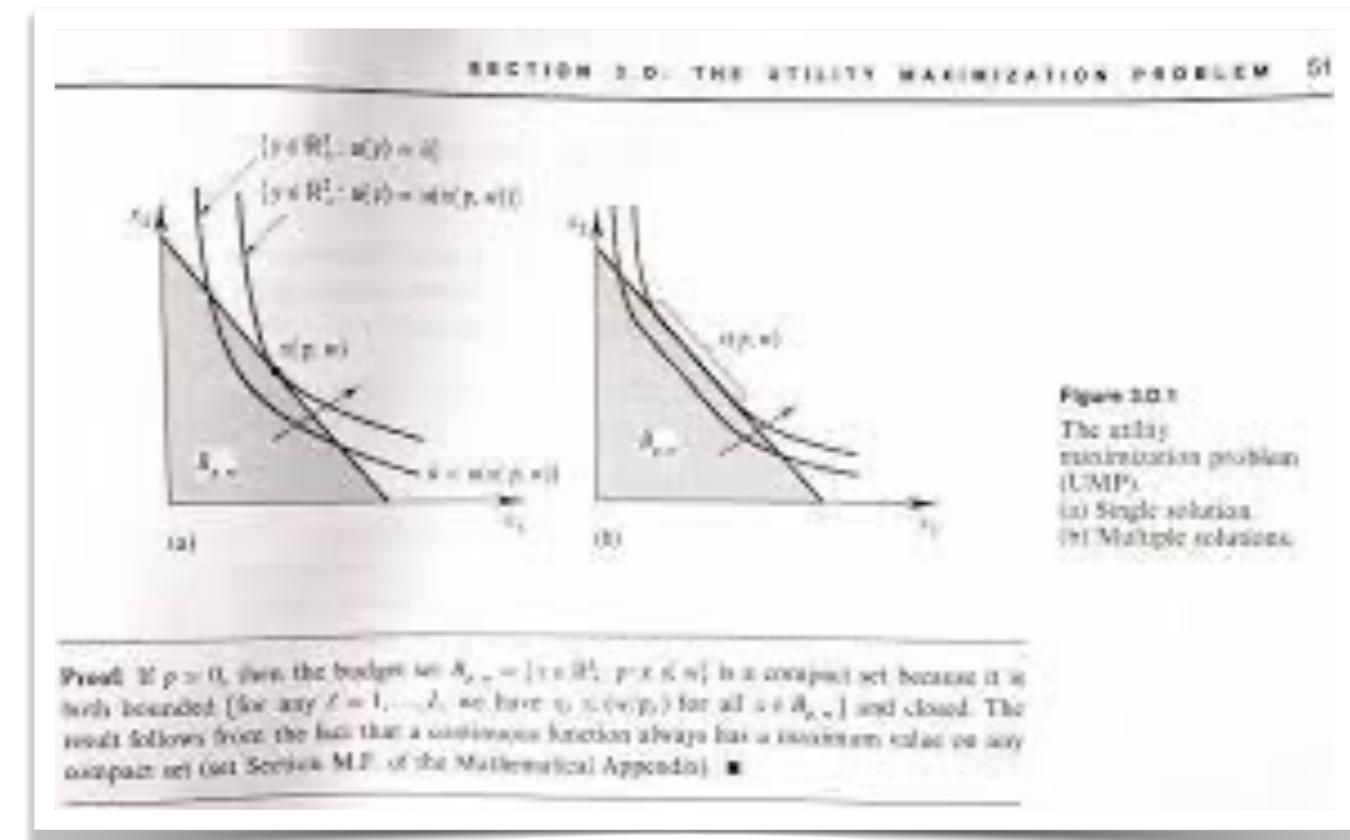
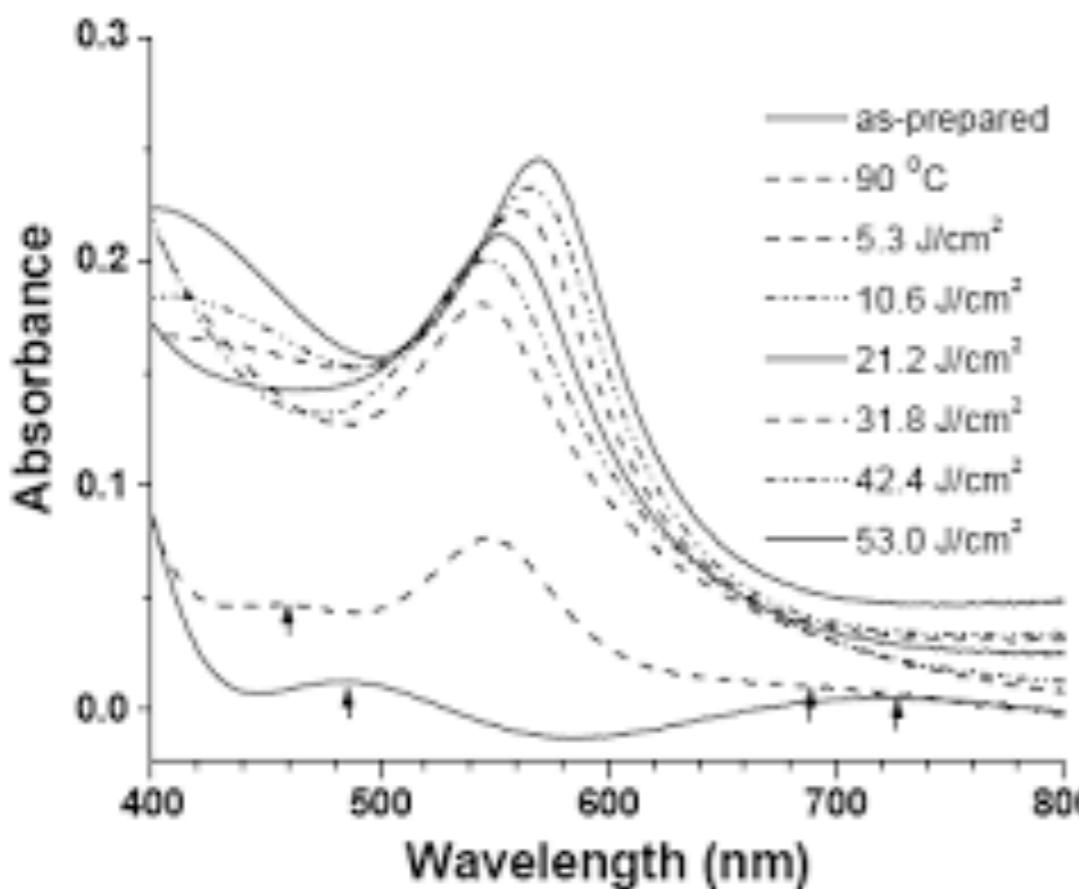
Gráfica sola (@RRSS)

En redes sociales, tenemos generalmente que diseñar **gráficas estáticas, gif o videos** que muestren el dato que queremos indicar.



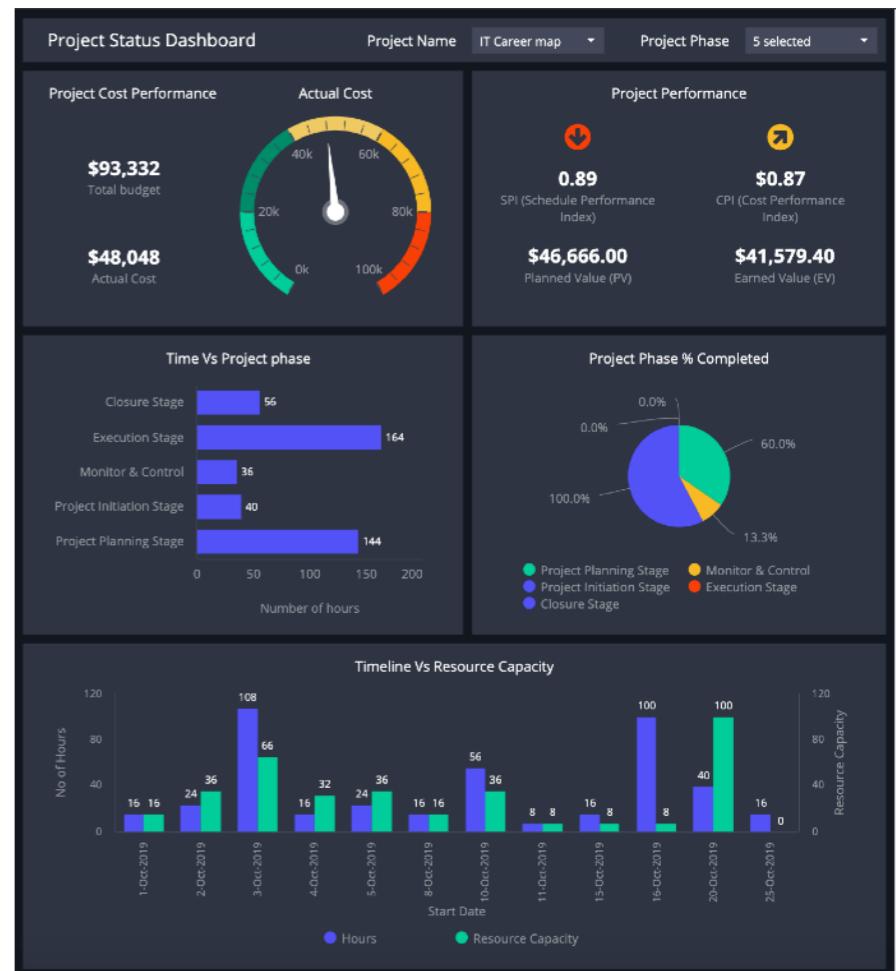
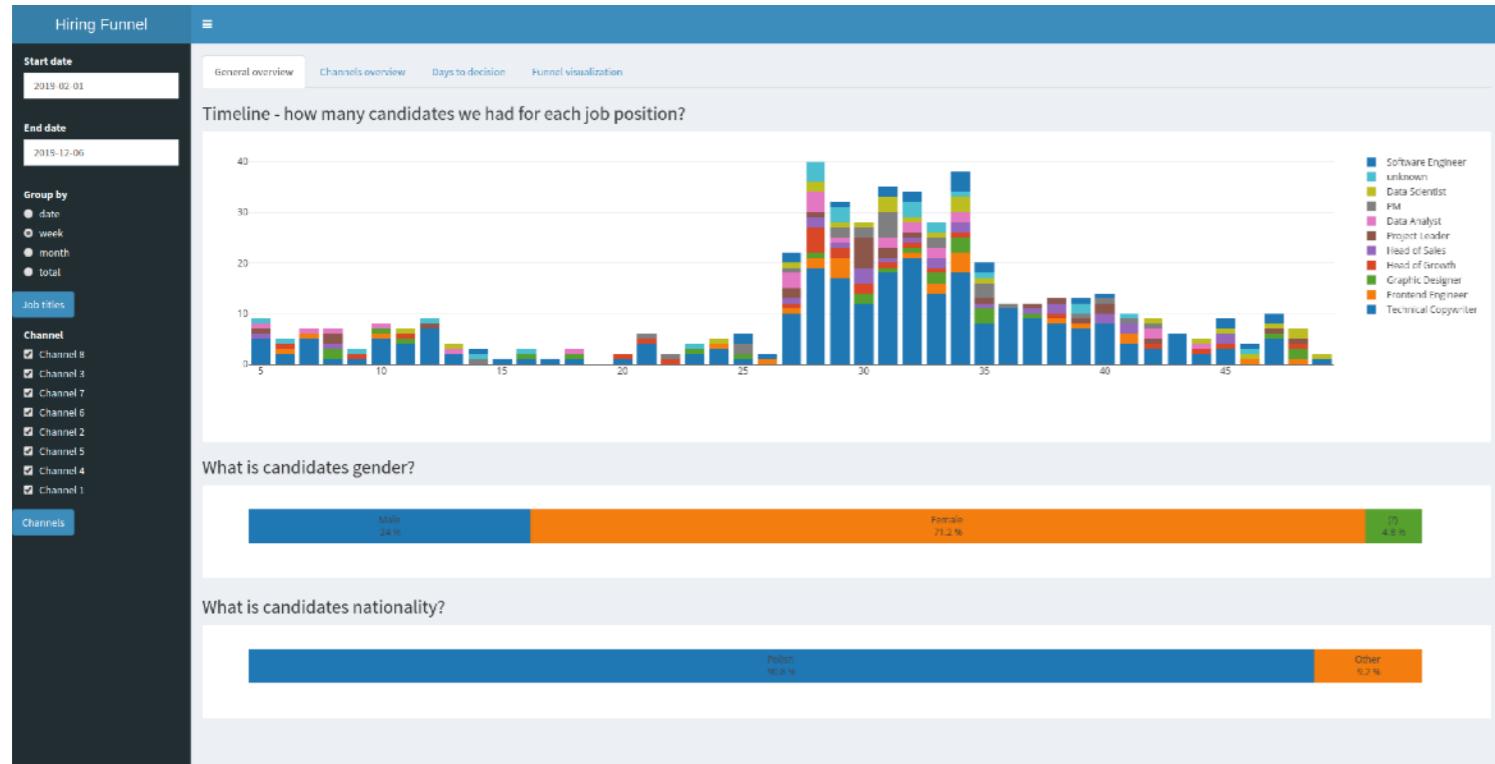
En artículos científicos

En este caso, tenemos restricciones de publicación (**colores, tamaños, uso de formas, formatos de archivos específicos**) que hay que tener en cuenta.



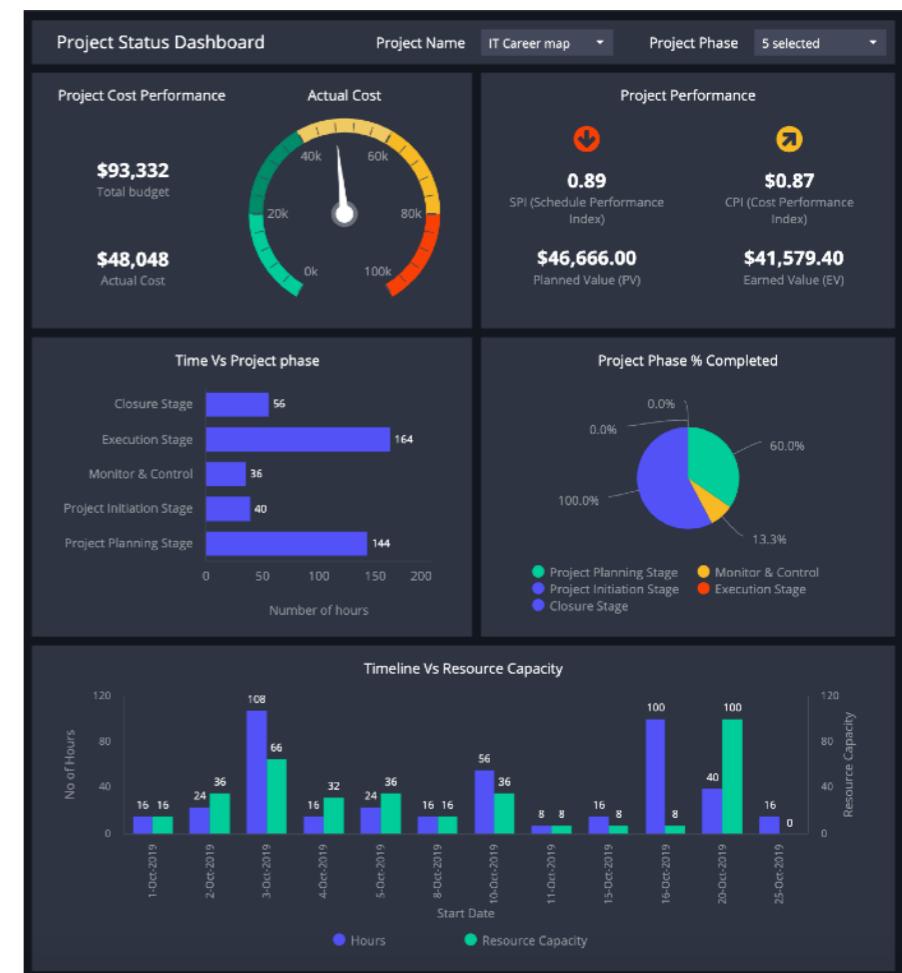
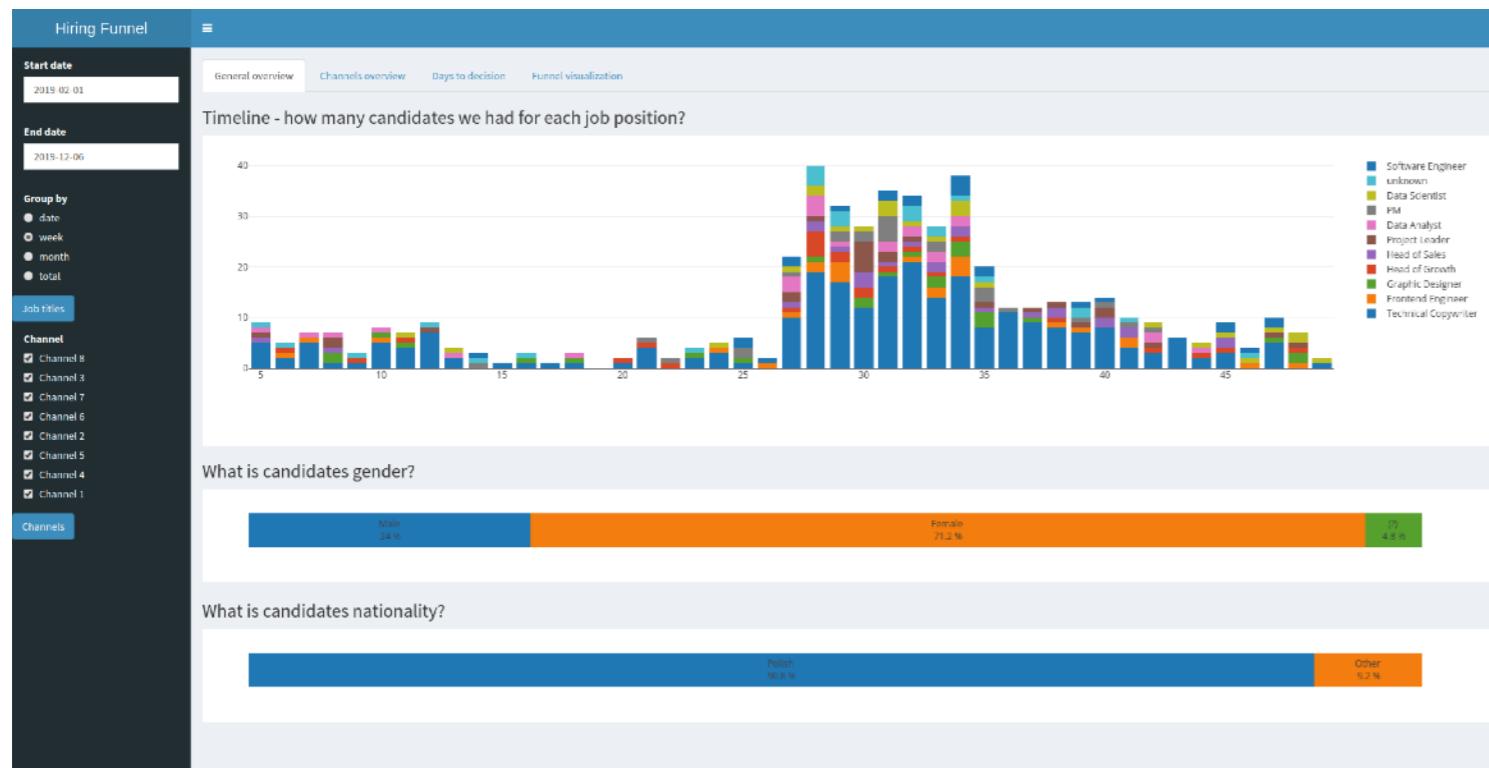
En dashboards

Un dashboard es una **herramienta de gestión** de la información que monitoriza, analiza y muestra de manera visual los **indicadores clave de desempeño (KPI)**, métricas y datos fundamentales para hacer un **seguimiento** del estado de una empresa, un departamento, una campaña o un proceso específico.



En dashboards

El reto de diseñar las visualizaciones en un dashboard es que **se diseñan varias visualizaciones** que deben mostrar información relevante, deben **trabajar en conjunto** y deben **adaptarse a cambios en los datos**.

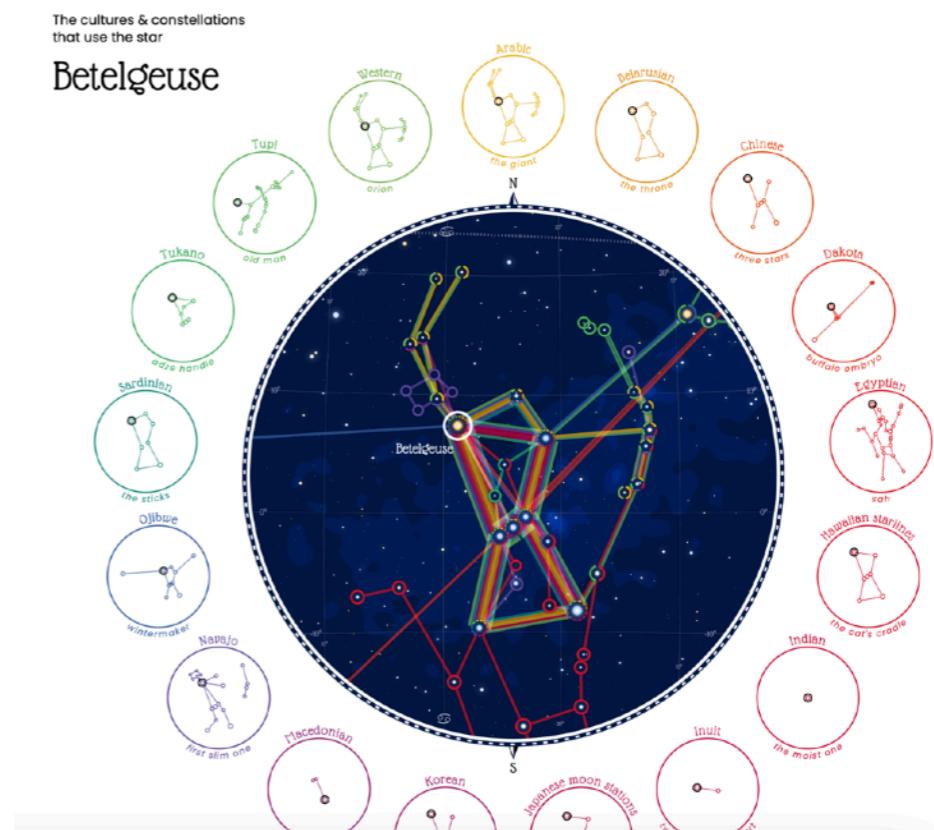


En una página web

Uno de los medios más flexibles para publicar las visualizaciones de datos. Se puede hacer uso de **formatos interactivos y estáticos**, así como de **efectos** y **animaciones**.



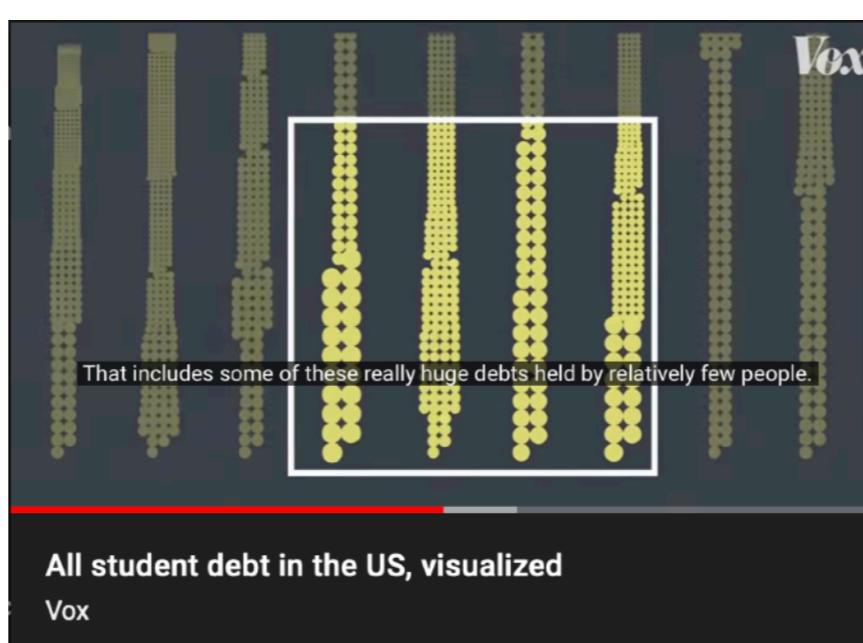
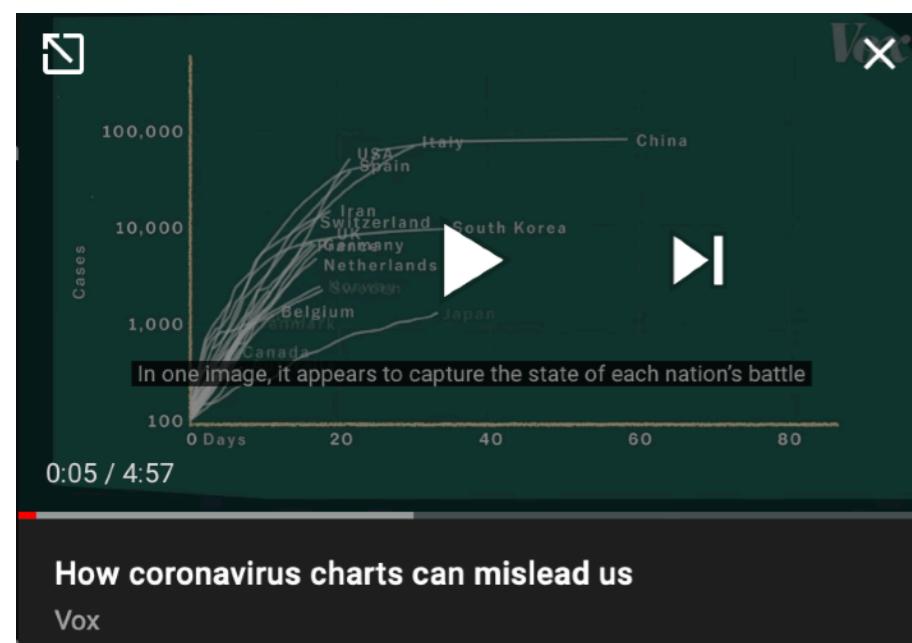
<https://whydocatsanddogs.com>



<https://figuresinthesky.visualcinnamon.com>

Video

En el caso del video o formatos adicionales, **se espera** que las visualizaciones vengan acompañadas de **animaciones, transiciones y mensajes** aclaratorios.



```
funciones.R x styles.css x app.R x server.R x
abstenci Next Prev All Replace All
In selection Match case Whole word Regex Wrap
1 # Librerias ----
2 library(tidyverse)
3 library(leaflet)
4 library(sf)
5 library(shiny)
6 library(ggtext)
7 library(DT)
8 library(shinyCSSloaders)
9
10 # Objetos ----
11
12 # Funciones ----
38:92 # Aplicacion
```

Creación

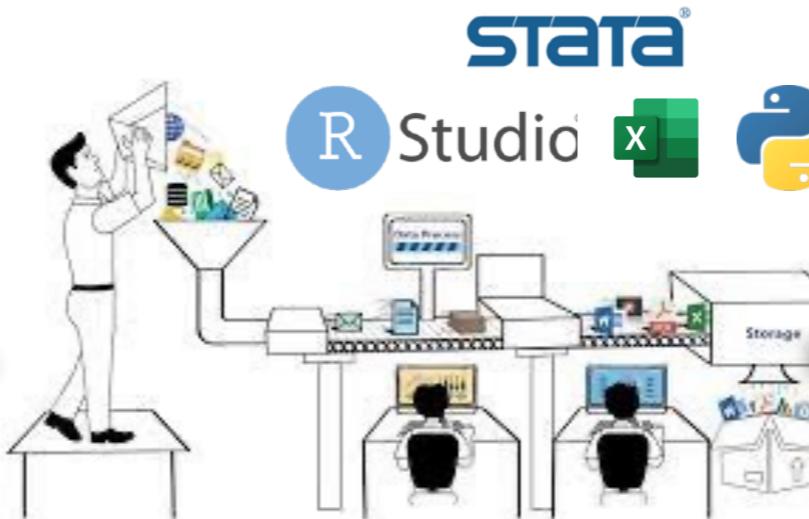
```
Console Terminal x Deploy x Jobs x
~/Desktop/Datos Electorales/07_ShinyApp_Municipal/ ↵
[1] "01001"
> runApp()
Listening on http://127.0.0.1:6852
Joining, by = c("id", "entidad", "municipio", "CVE_INE", "VXM", "JHH", "TOTAL_VOTOS_CALCULADOS", "nom_mun", "nom_ent")
[1] "01001"
> |
```

¿Cómo crear visualizaciones de datos?

Proceso:

1. **Obtención** de los datos crudos.
2. **Procesamiento** de la información.
3. **Diseño y creación** de la visualización.
4. **Post-producción**
5. **Publicación**

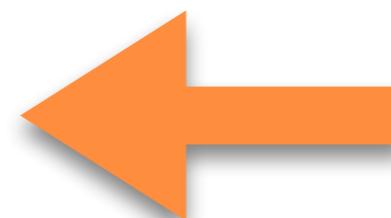
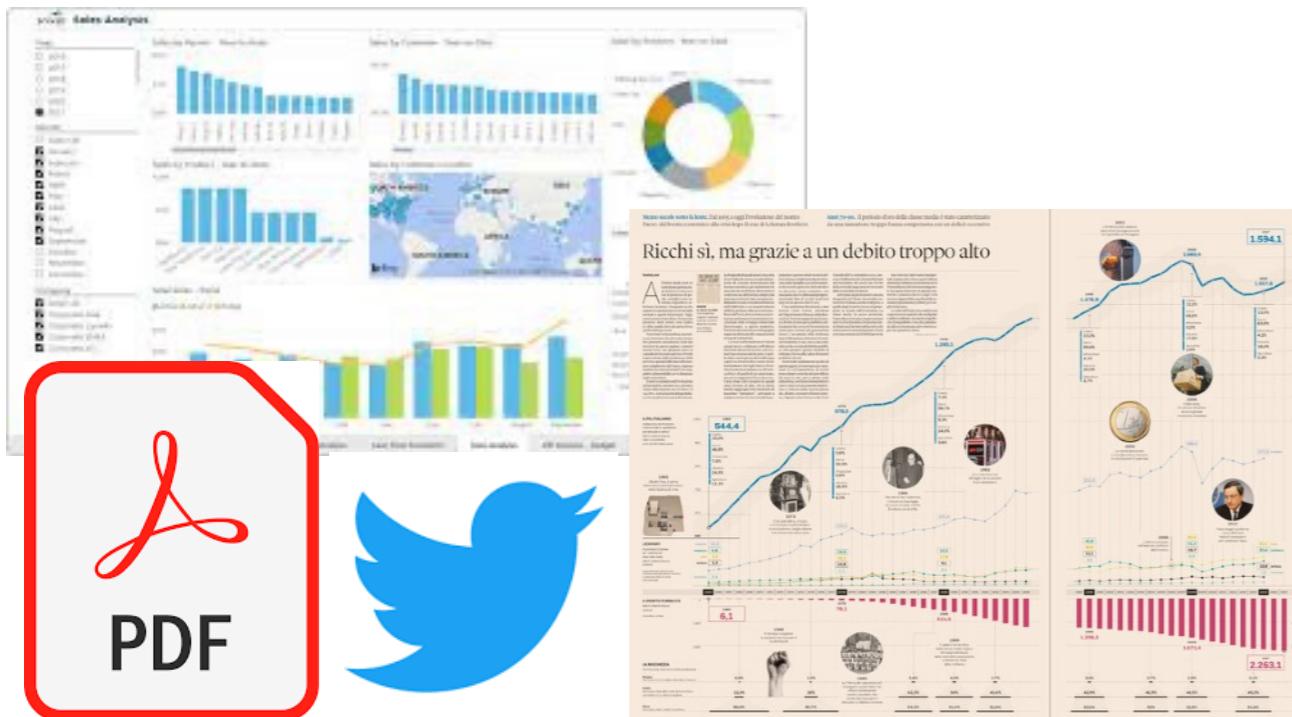
¿Cómo crear visualizaciones de datos?



Procesamiento



Diseño y creación



Publicación

Post-producción

Herramientas de trabajo

Análisis, procesamiento y visualización

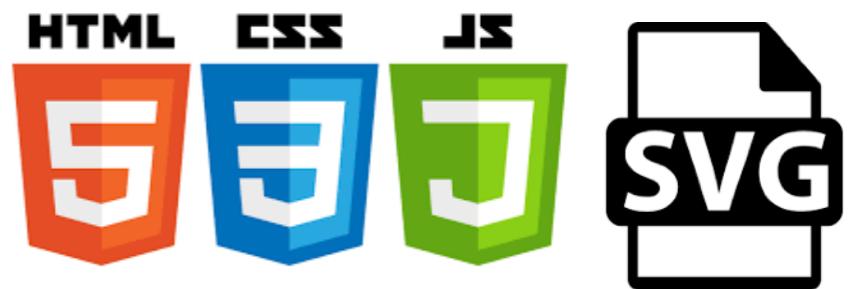


Creación de Dashboards

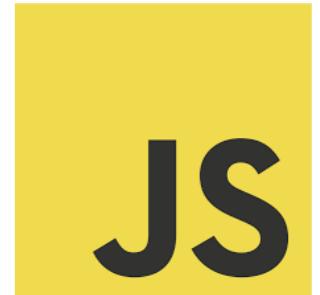


Google Data Studio

Tecnologías web



Librerías de visualización



Herramientas de post-producción y edición



Diseño

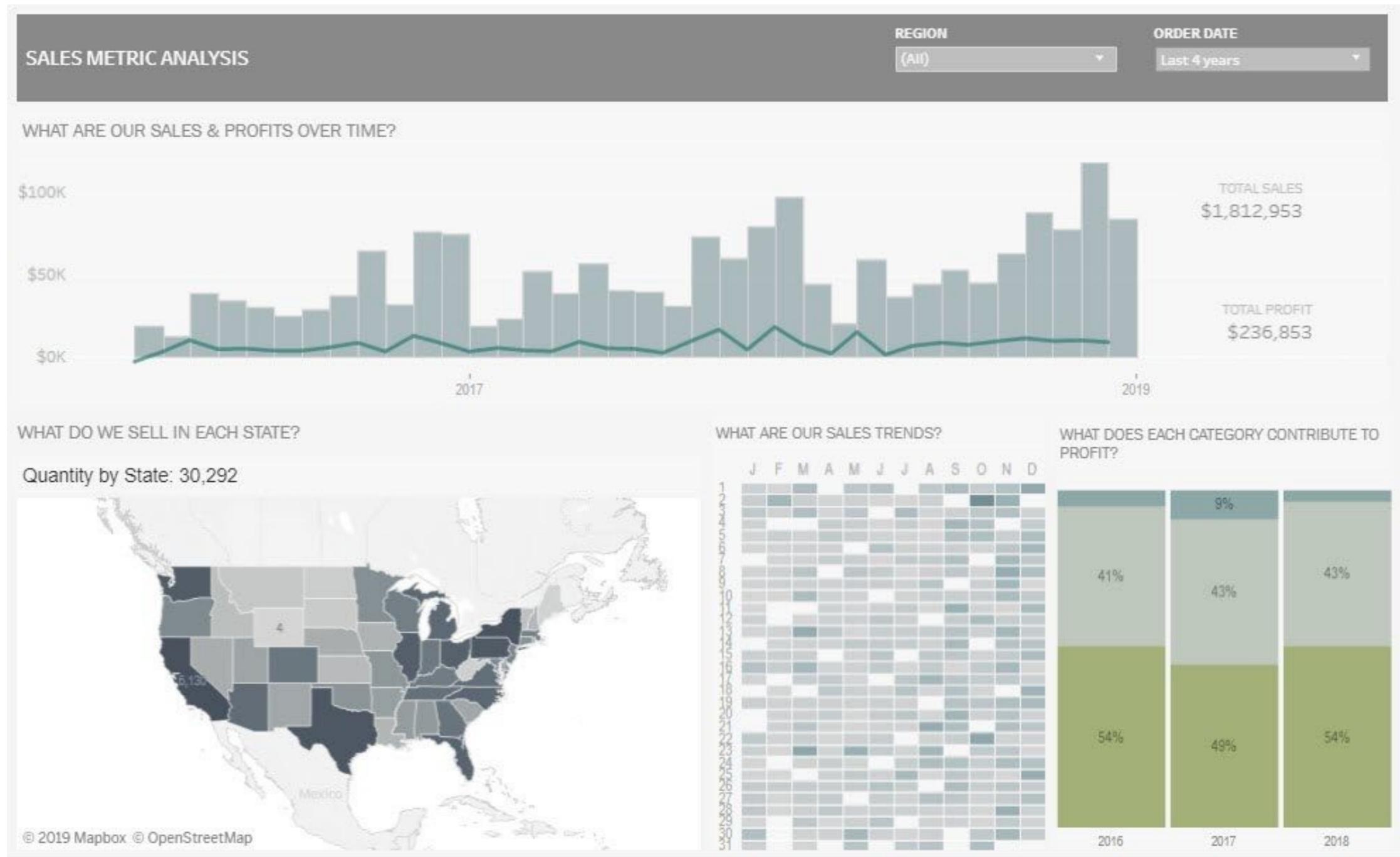
Preguntas de diseño

Preguntarse...

1. **1. ¿Quién es mi público?**
2. **2. ¿Qué preguntas tienen?**
3. **3. ¿Qué respuestas encuentra para ellos la visualización?**
4. **4. ¿Qué estoy tratando de comunicar?**
5. **5. ¿Qué otras preguntas inspirará mi visualización?
¿Qué conversaciones podría generar?**



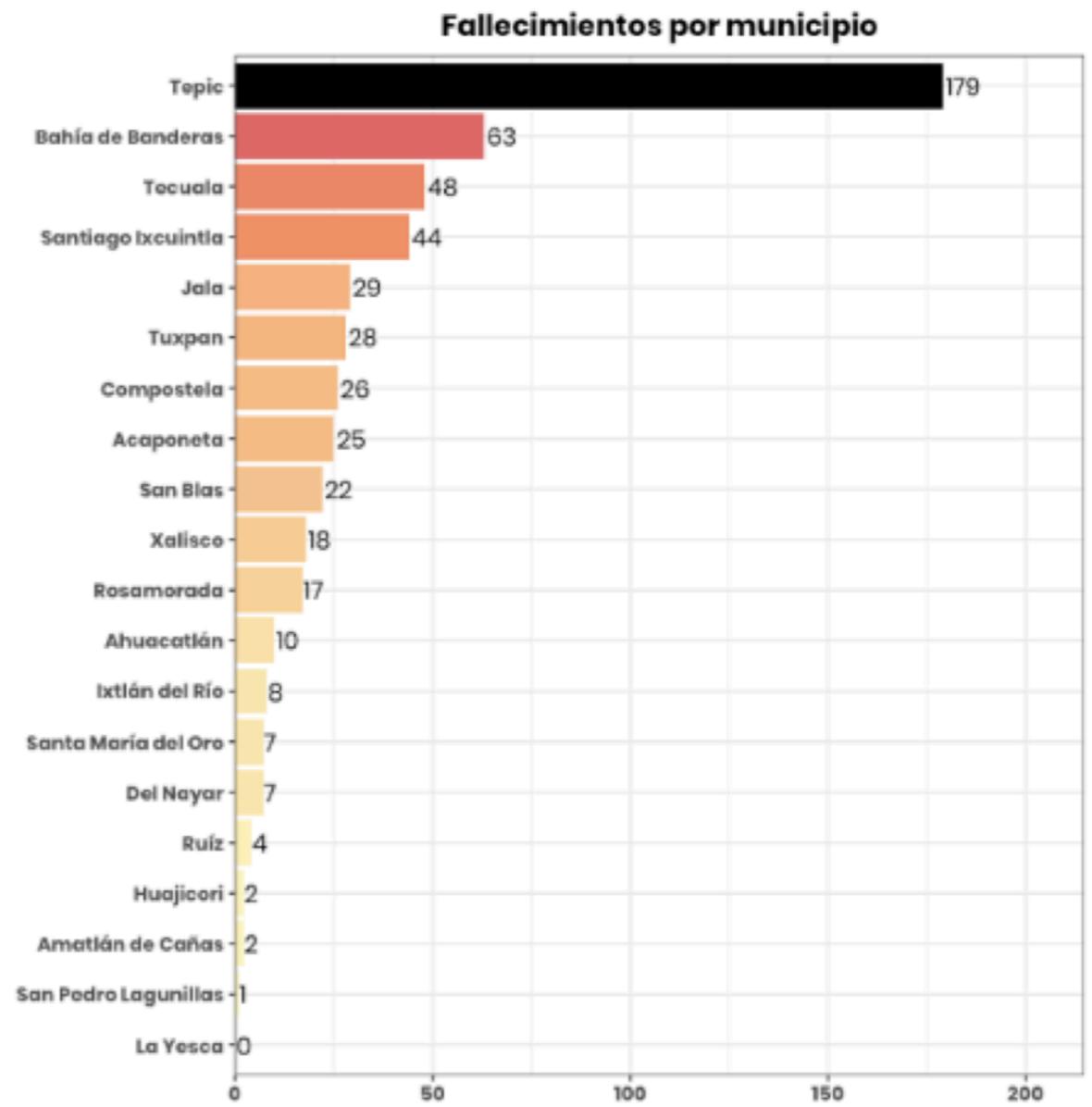
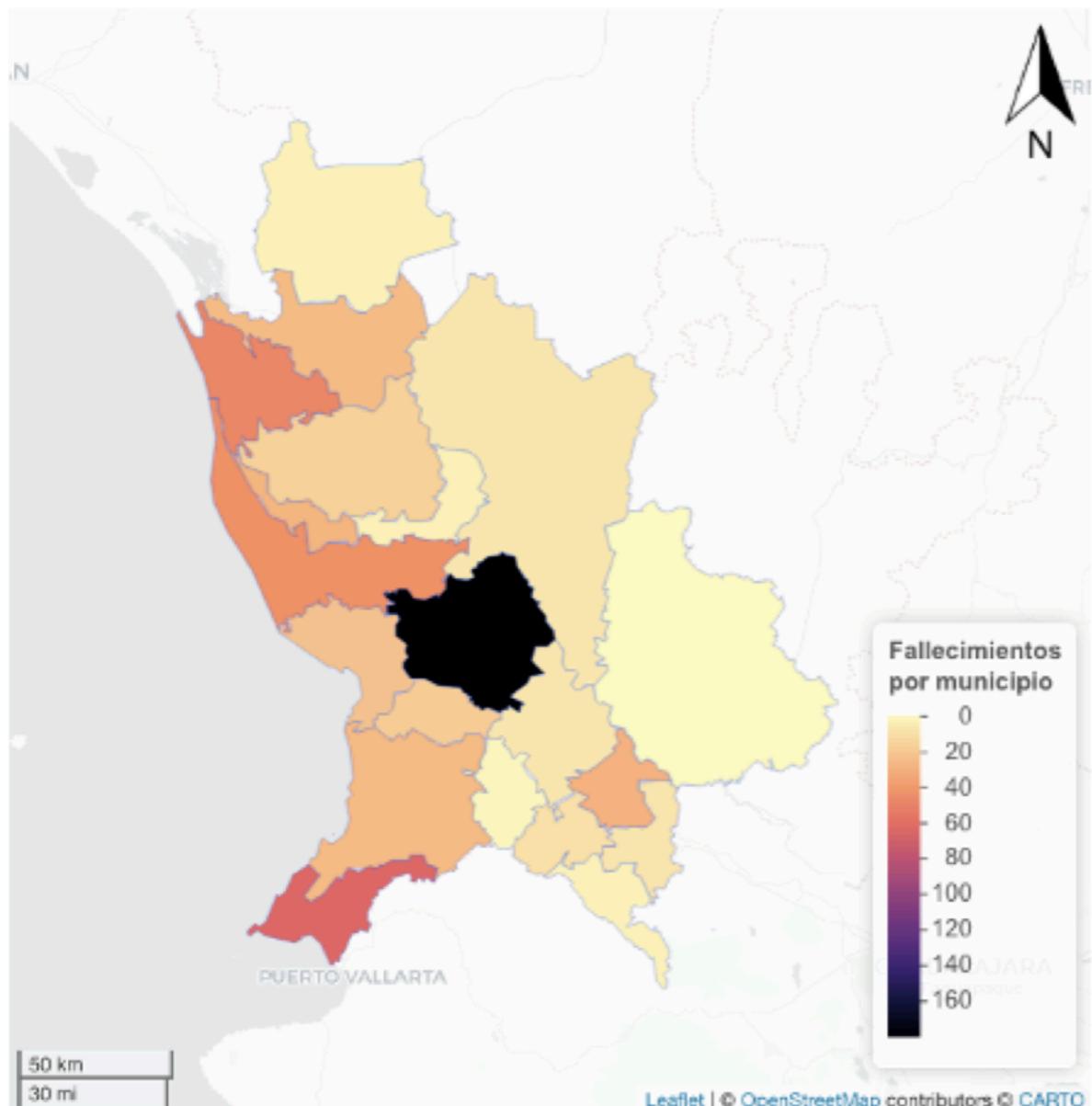
Preguntas de diseño



<https://www.tableau.com/es-mx/learn/articles/data-visualization-tips>

Escogiendo la mejor gráfica

¿Cuál es **mejor**? ¿Qué **variable** y **mensaje** da cada una?



Fuente: Elaboración propia con datos de la Secretaría de Salud, 24 de Agosto del 2020.

Escogiendo la mejor gráfica

Las gráficas se eligen en función de **nuestras preguntas de diseño**, de **nuestros datos** y del **mensaje** que queremos transmitir.



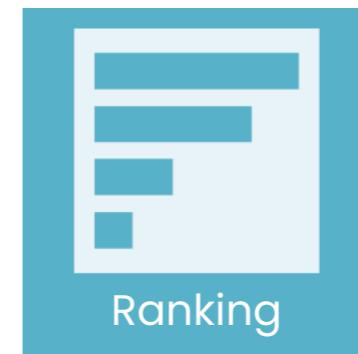
Desviación



Correlación



Cambio



Ranking



Distribución



Parte de un todo



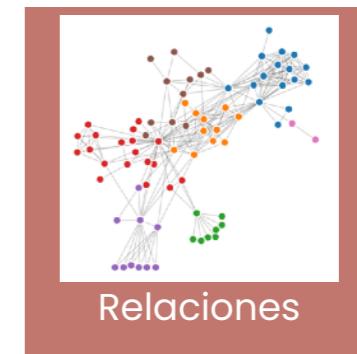
Magnitud



Espacio



Flujo



Relaciones

<http://ft-interactive.github.io/visual-vocabulary/>



from Data to Viz

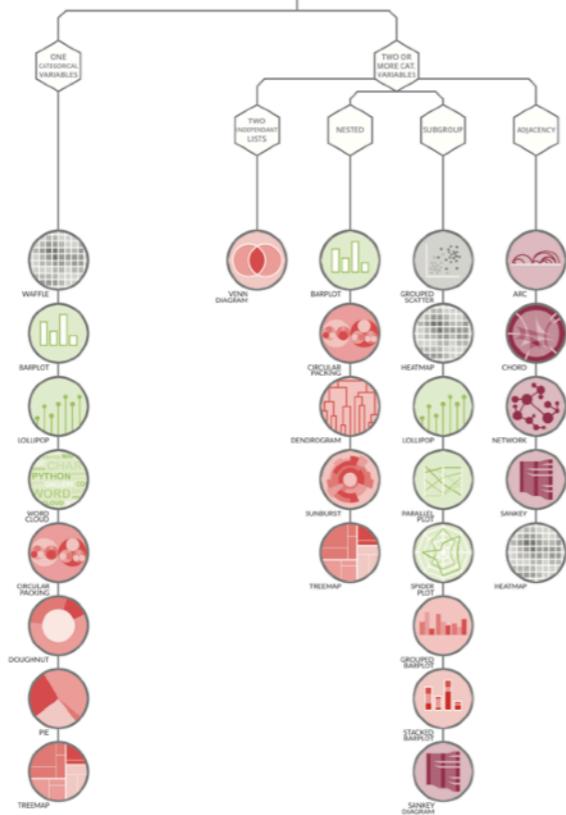
'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps :

- ① Identify what type of data you have.
- ② Go to the corresponding decision tree and follow it down to a set of possible charts.
- ③ Choose the chart from the set that will suit your data and your needs best.

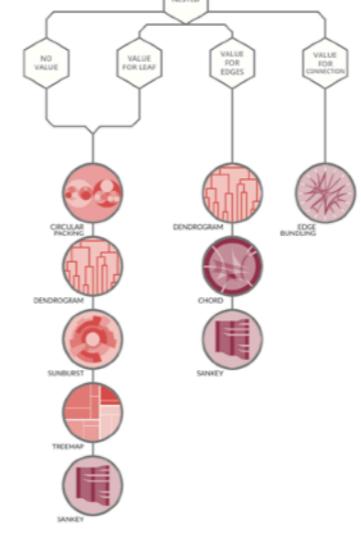
Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

data-to-viz.com

CATEGORIC



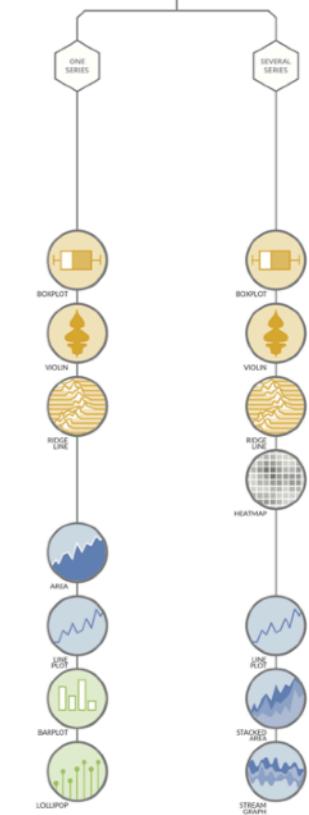
RELATIONAL



MAP



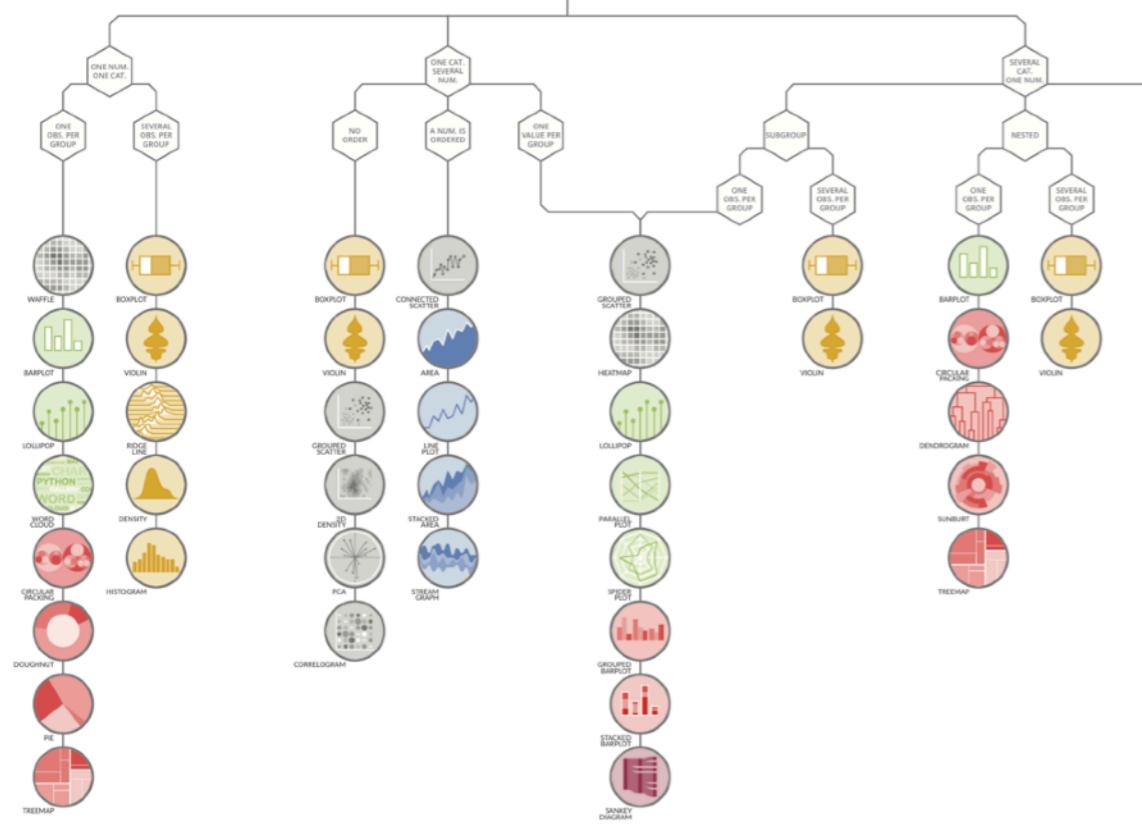
TIME SERIES



WHAT DO YOU WANT TO SHOW ?

- | | |
|-------------------|-------------|
| ● Distribution | ● Evolution |
| ● Correlation | ● Maps |
| ● Ranking | ● Flow |
| ● Part of a whole | |

CATEGORIC AND NUMERIC



Uso de colores

Regla 60-30-10



Armonización de colores



Complementary



Analogous



Triadic



Split-Complementary



Rectangle (tetradic)



Square

Uso de colores neutros

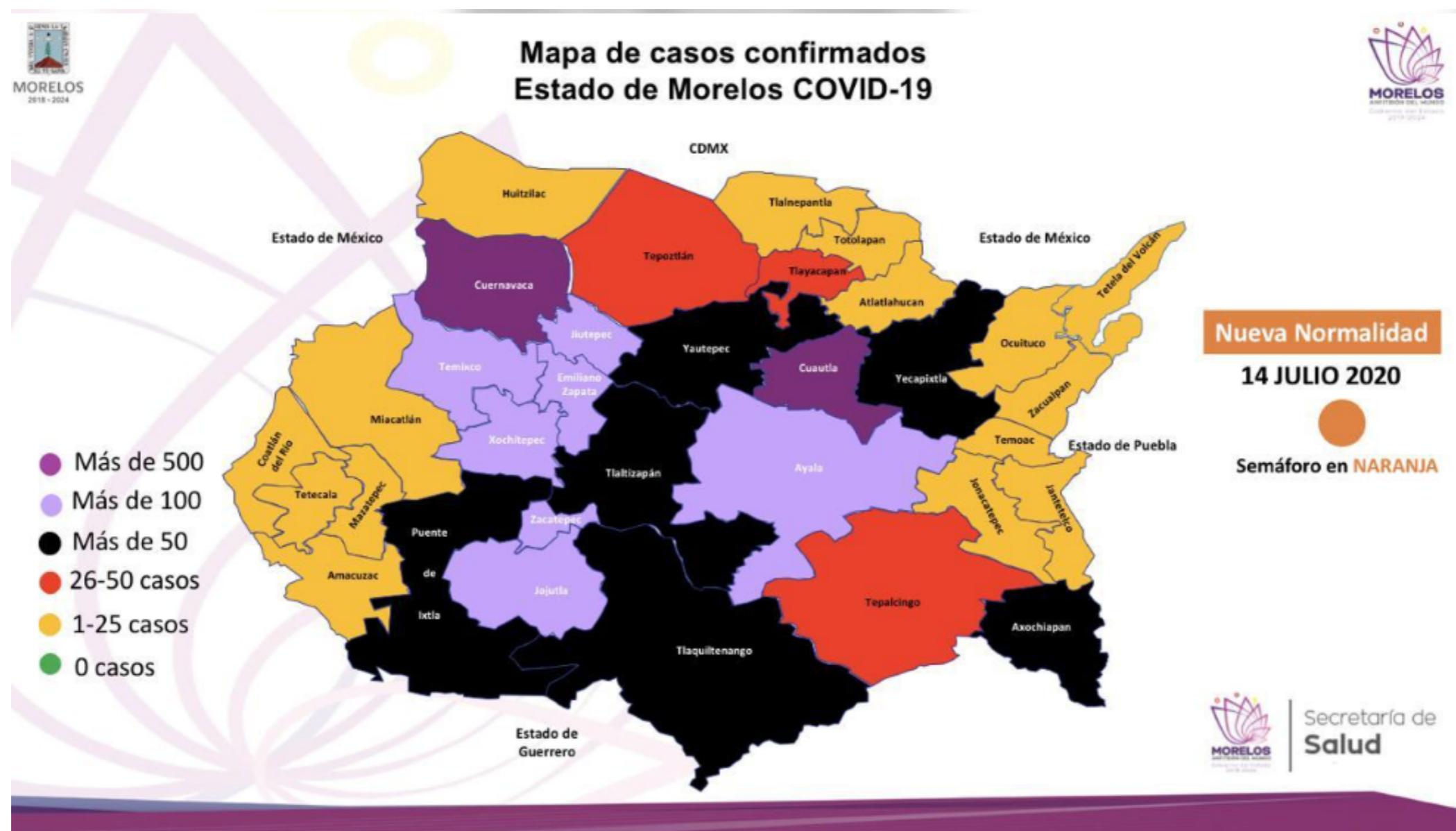


Menos es más



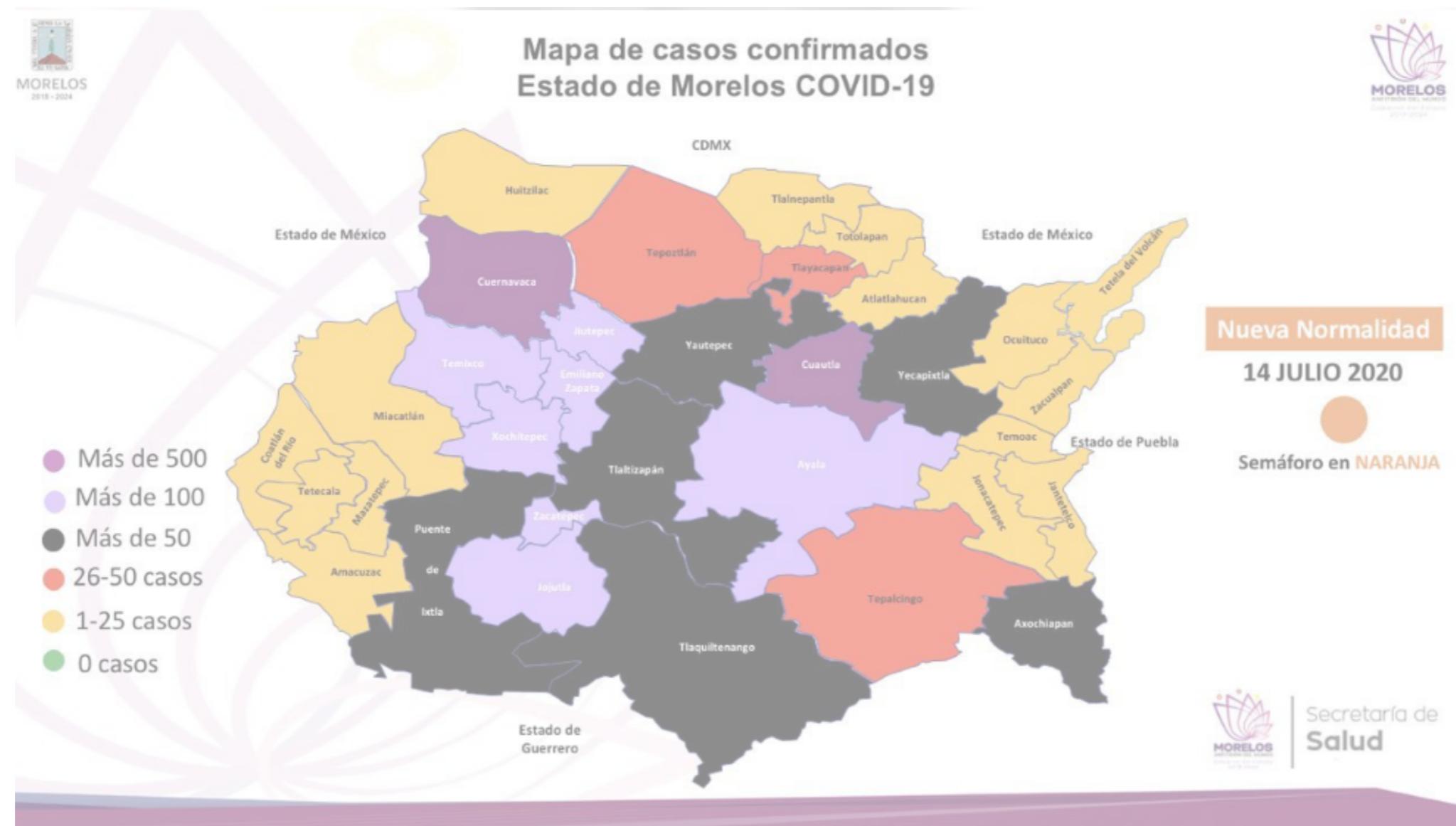
Uso de colores

¿Qué está **mal** en este uso de colores?



Uso de colores

¿Qué está **mal** en este uso de colores?



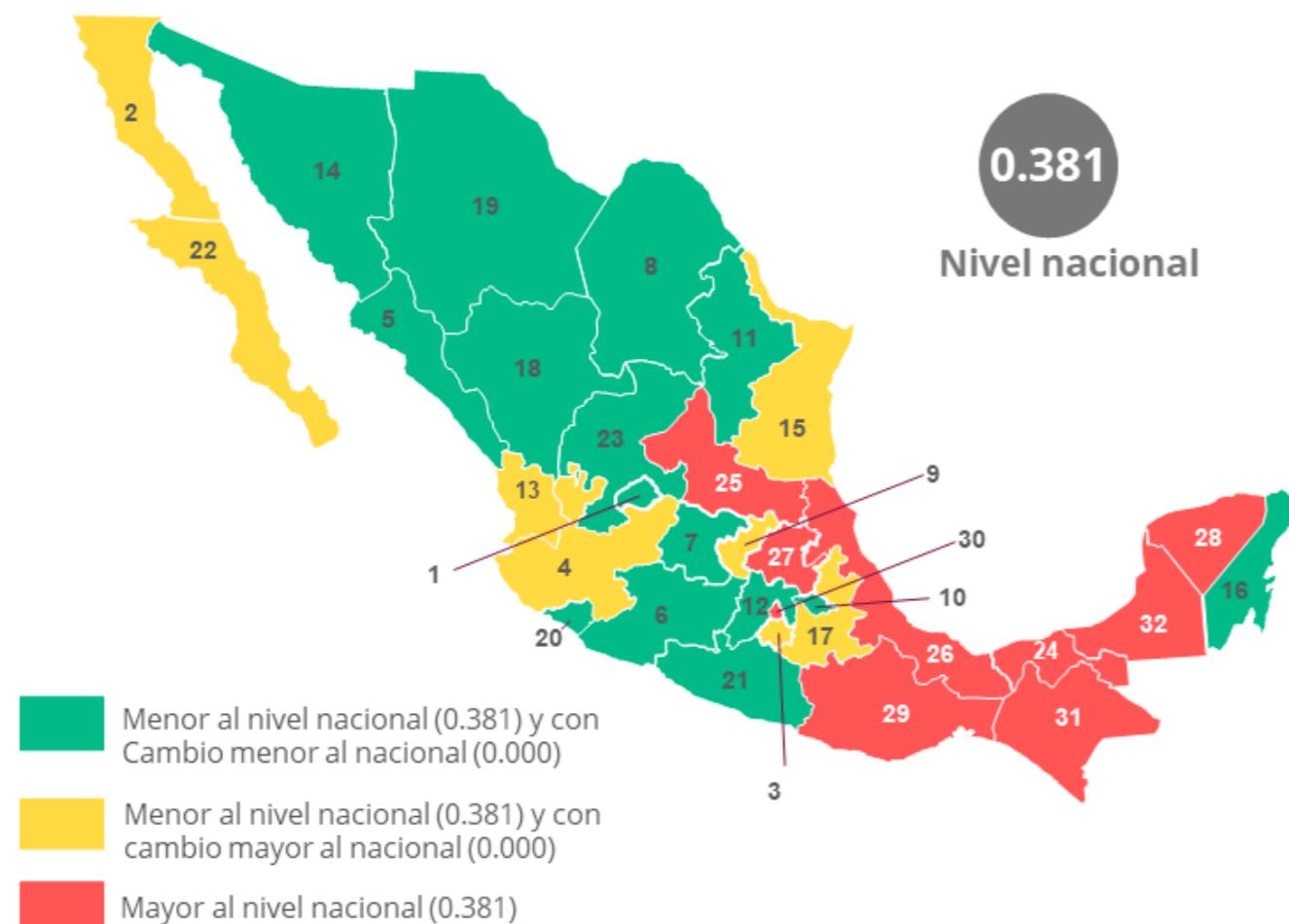
Siempre tener en cuenta el **significado cultural** de los colores.

Uso de colores

Las paletas “semáforo” igualmente traen una carga **contextual**.

Desigualdad laboral

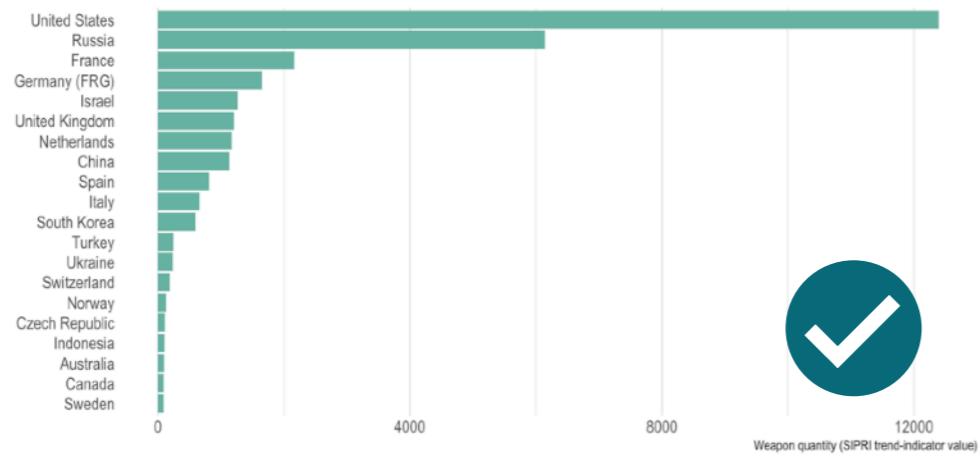
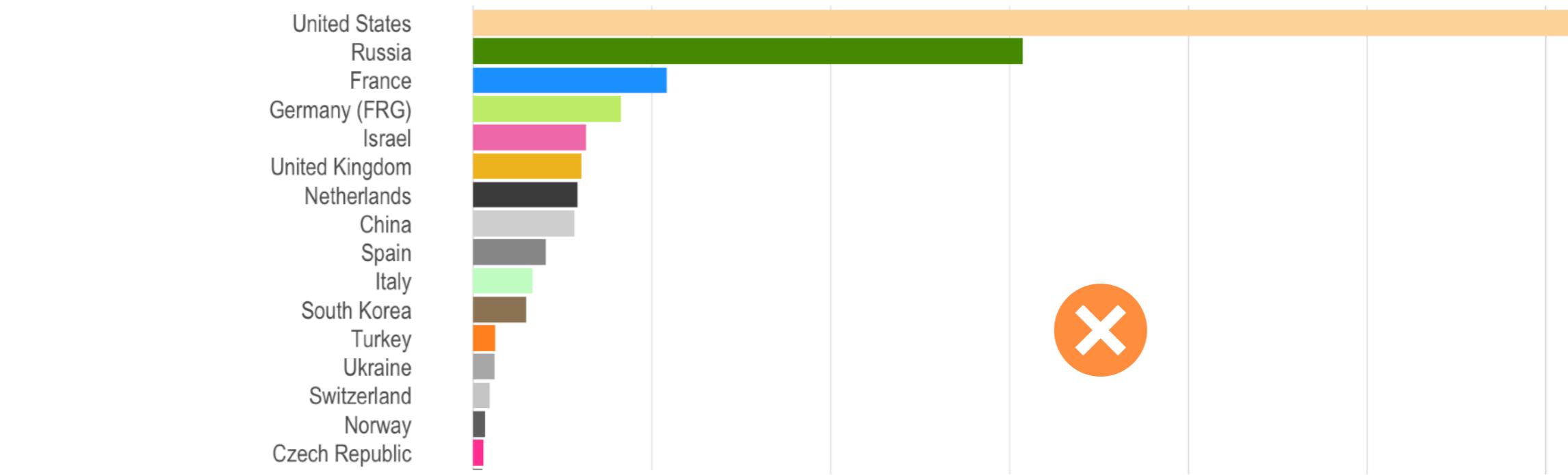
1^{er} trimestre 2021



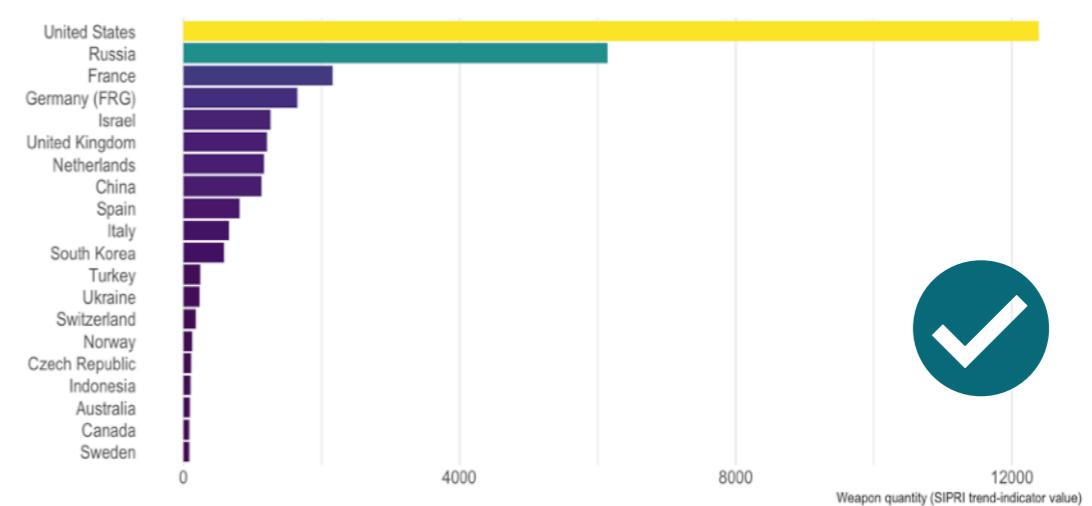
ELABORADO POR MÉXICO, ¿CÓMO VAMOS? CON DATOS DEL INEGI.

Uso de colores

Si tus colores no representan nada, mejor no los uses.



40

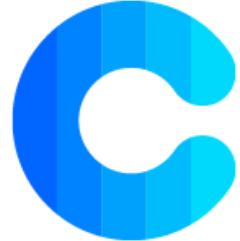


40

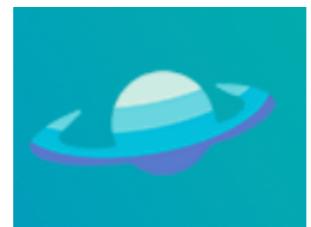
https://www.data-to-viz.com/caveat/color_com_nothing.html

Uso de colores

Usa generadores de paletas de colores



<https://colors.co/>



<https://mycolor.space>



<https://color.adobe.com/>



Google Color Picker

<https://htmlcolors.com/google-color-picker>

Genera tus propias reglas de uso

Spot Palette



#046A78

#4E9EAC

#BFFBF9

#FE8E3C

Generic Gradient



#046A78

#008A88

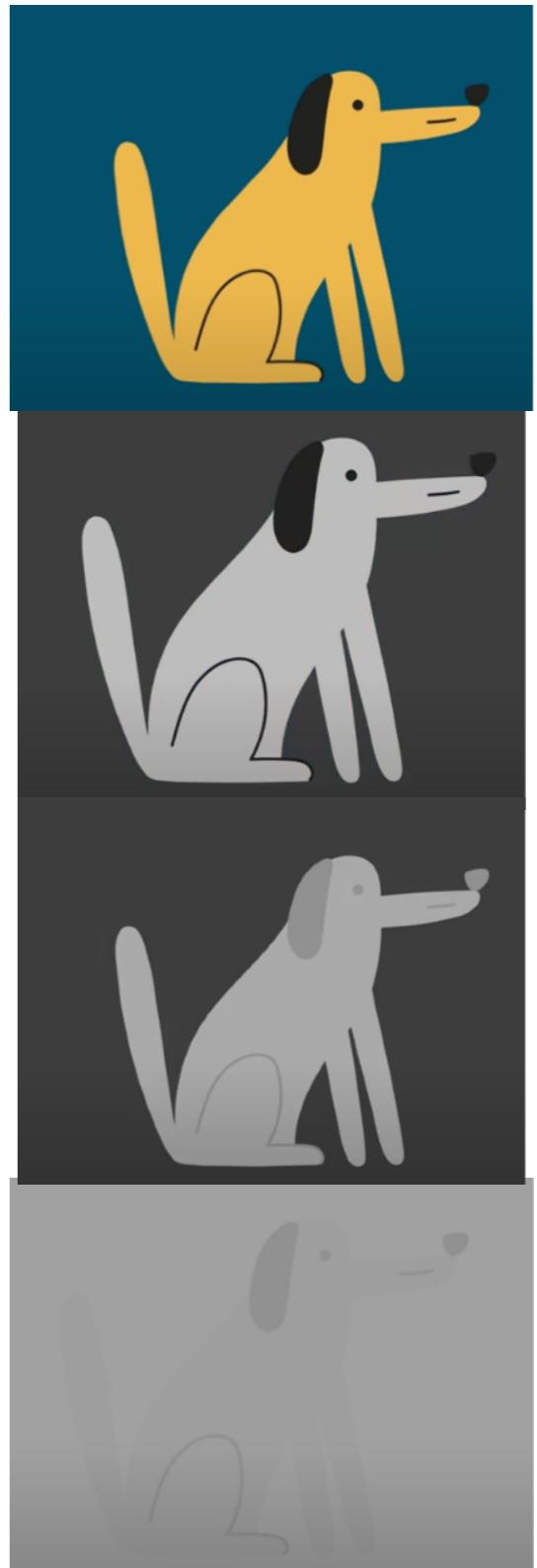
#2BAA8C

#69C885

#ADE279

#F9F871

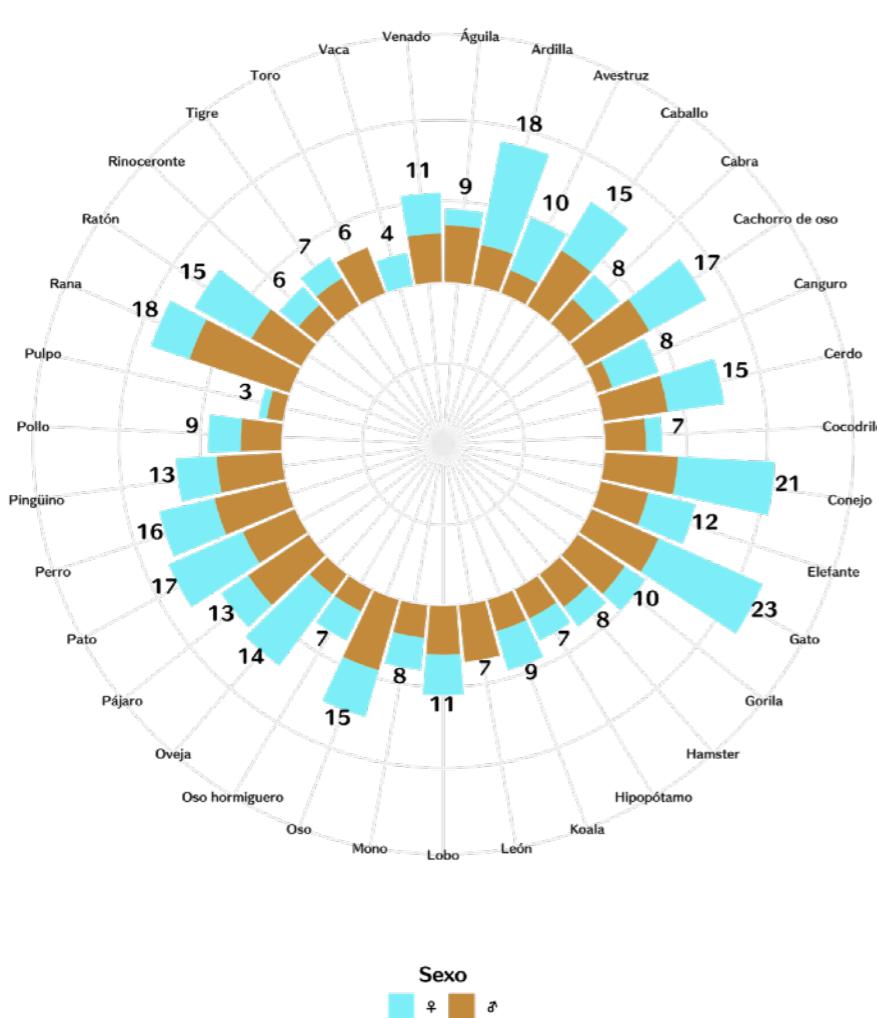
Uso de colores contrastantes



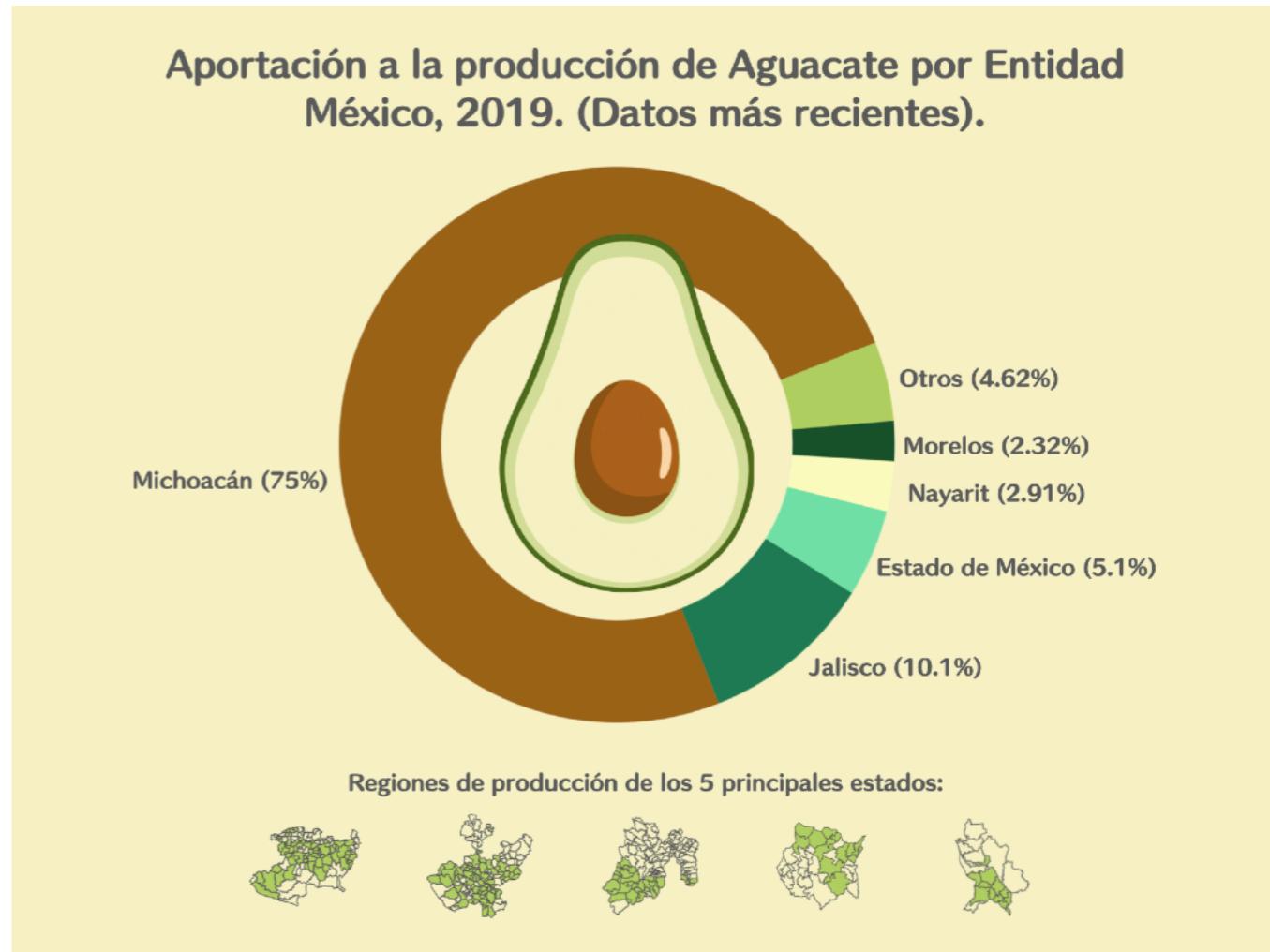
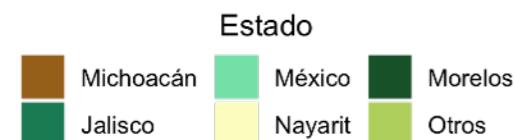
Uso de pistas contextuales

Vecinos de Animal Crossing por Especie y Sexo

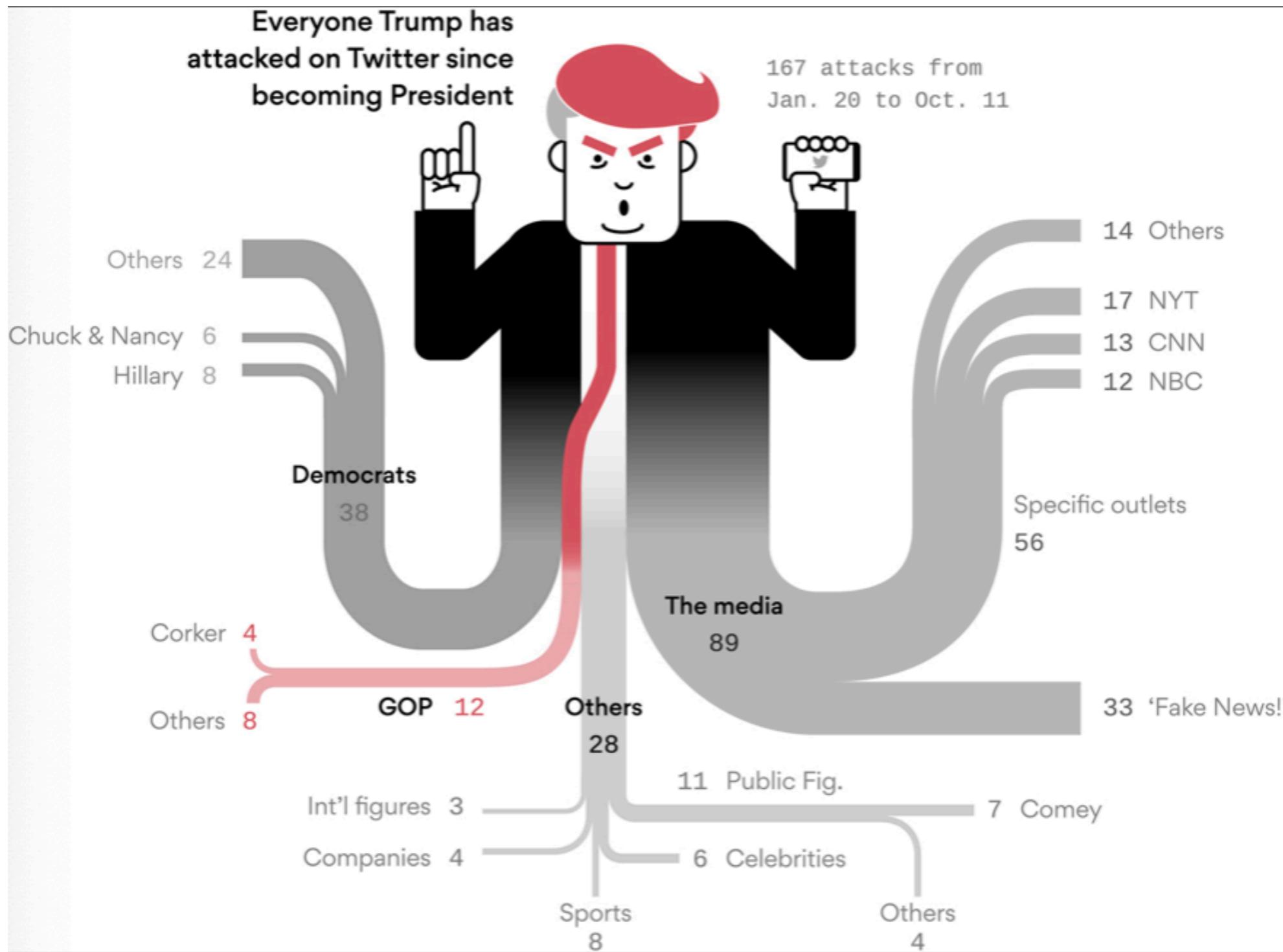
Animal Crossing Villagers by gender and species



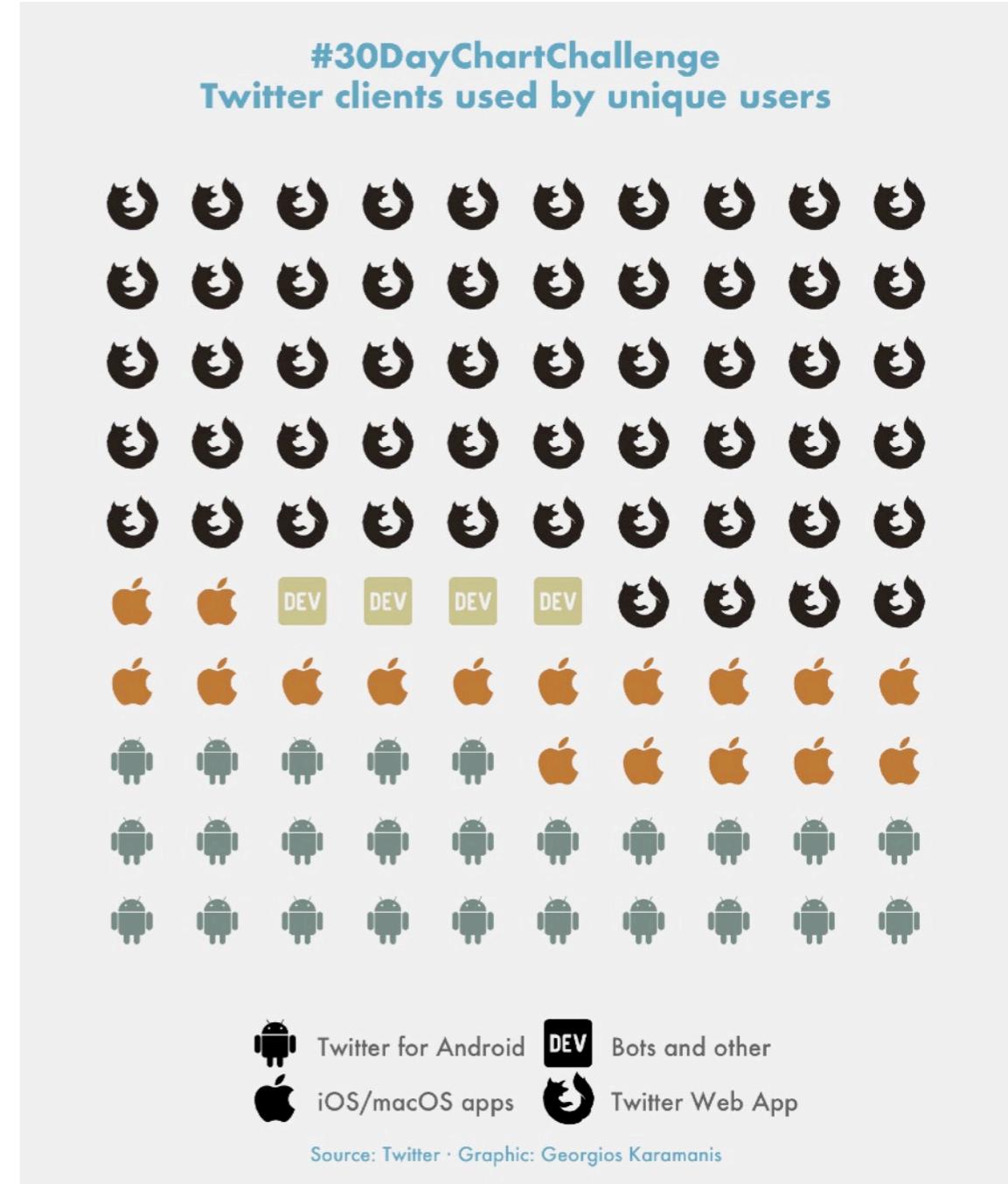
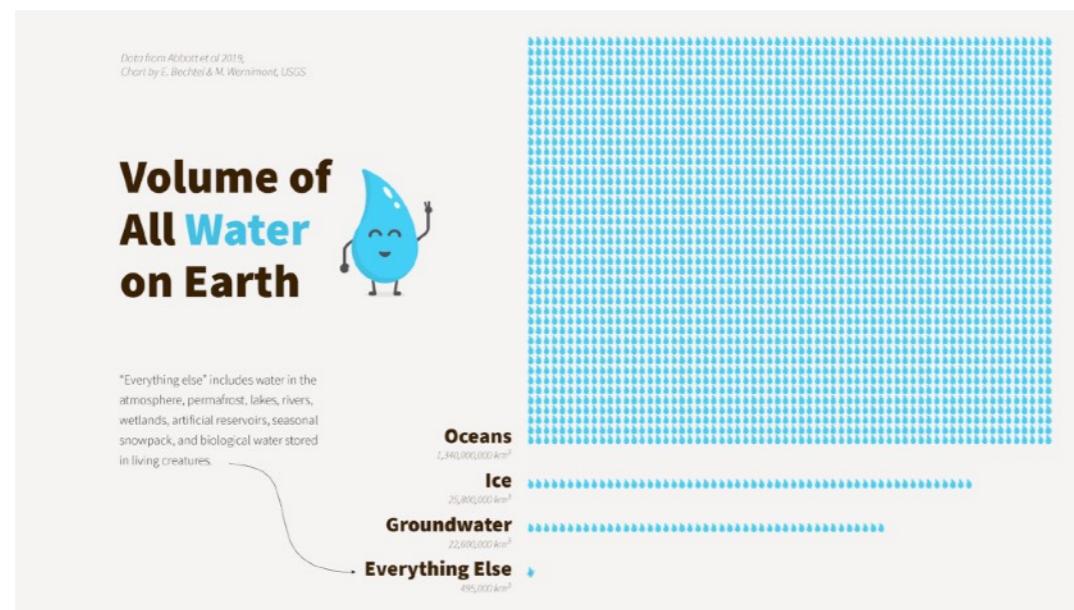
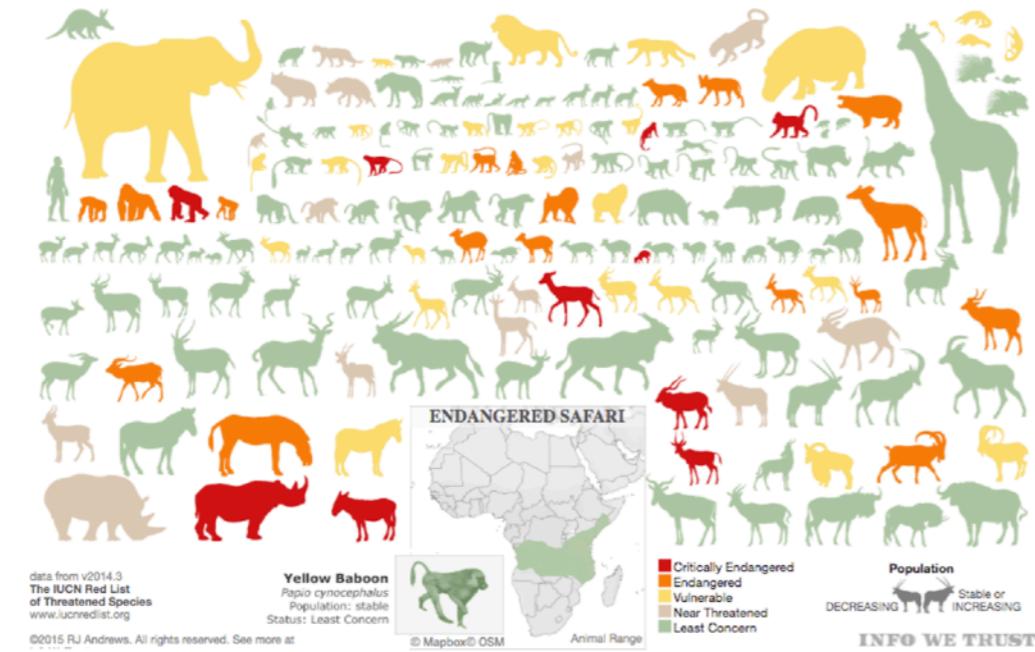
Uso de pistas contextuales



Uso de pistas contextuales

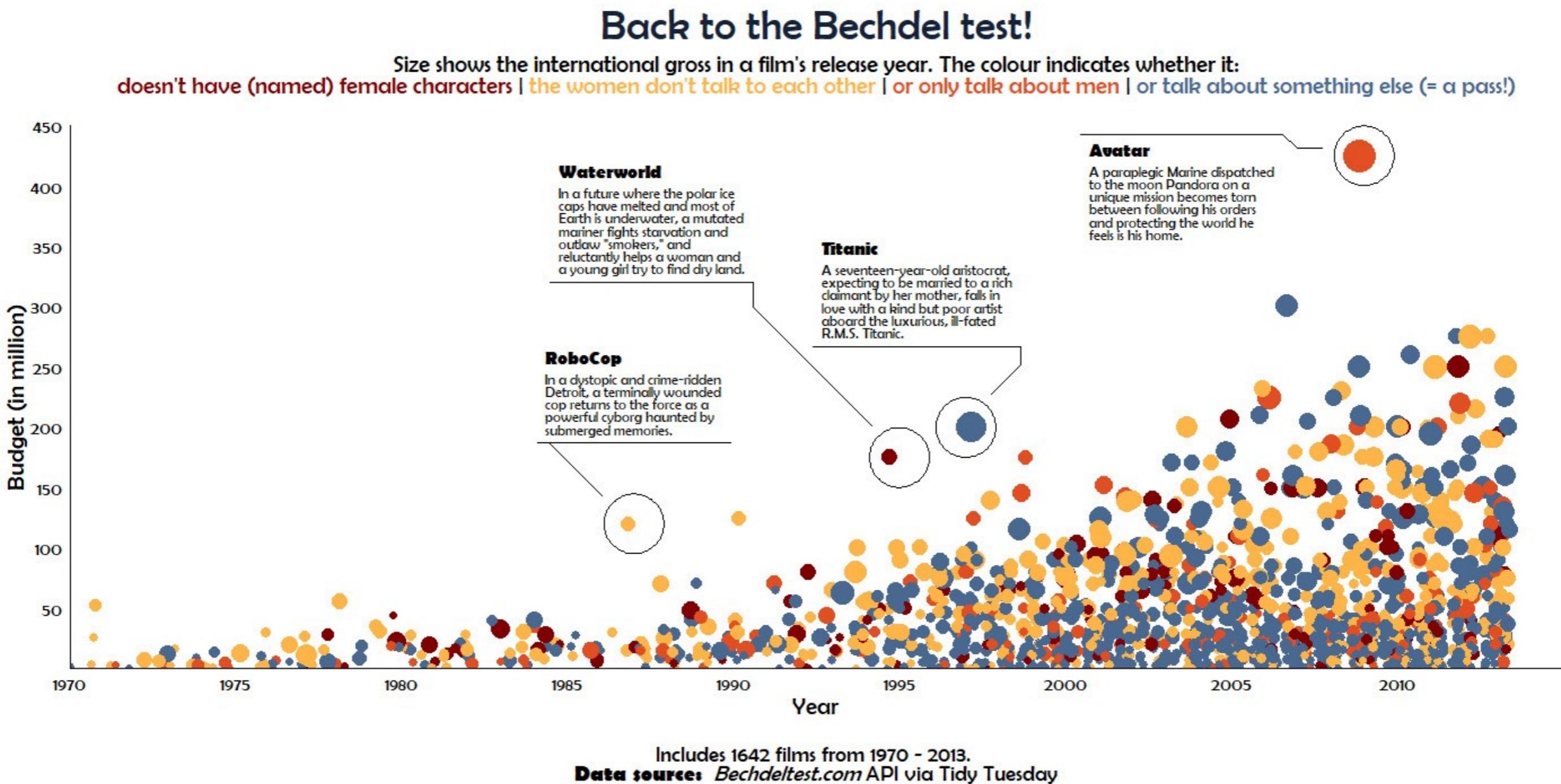


Uso de pistas contextuales

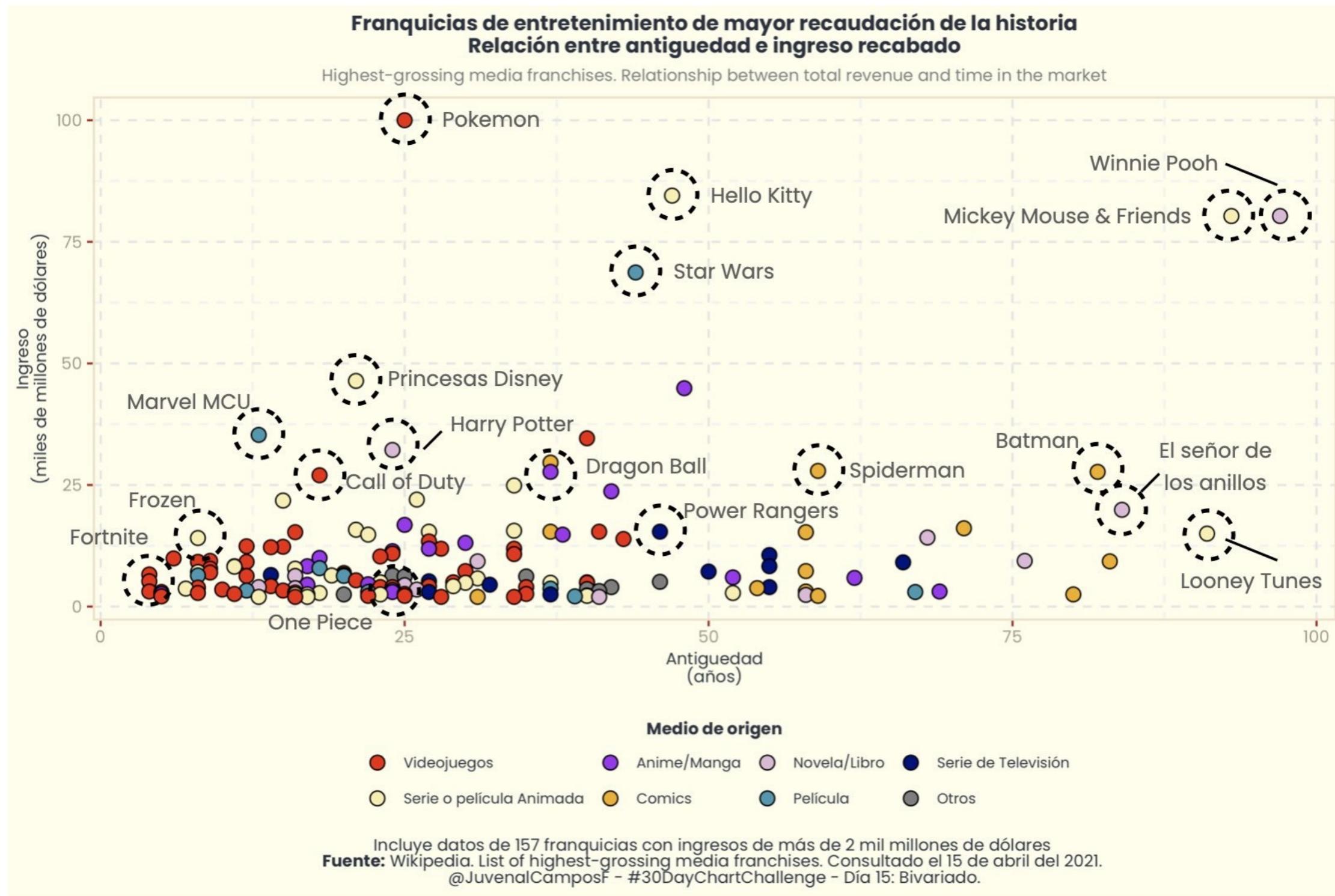


Uso de texto de manera estratégica

Usamos texto para indicar detalles importantes, sin **distráer**, sin poner **demasiado** y con el **contraste** adecuado.



Uso de texto de manera estratégica



Otros trucos

1. **Ordena tus datos.** Aprovecha la marca de posición.
2. Si **cortas el eje Y**, justifica el por qué.
3. **No uses demasiadas líneas en una gráfica de líneas.** No quieres una gráfica de espagueti.
4. El ojo humano es **malo leyendo ángulos**. Checa alternativas a las gráficas de pastel.



Order your data



To cut or not to cut?



The spaghetti chart



Pie chart

<https://www.data-to-viz.com/caveats.html>

Otros trucos

1. Juega siempre con **el tamaño de bin** de los histogramas.
Puede llevar a resultados diferentes.
2. **No llenes tus gráficas** de distribuciones. O de ruido en general.
3. Si tu gráfica tiene **demasiados puntos**, usa técnicas para hacerla más legible
4. Evita la **paleta arcoíris** en variables numéricas.



Play with histogram bin size



Too many distributions.



Overplotting



The rainbow color palette

<https://www.data-to-viz.com/caveats.html>

Otros trucos

1. Salvo en raras situaciones, **el 3D debe ser evitado**, ya que distorsiona la realidad.
2. Si usas varias gráficas para mostrar un punto, trata que exista **consistencia** entre ellas.
3. Preocúpate si tu leyenda tiene **más de 14 categorías**.
4. **Evita relaciones de aspecto extremas**. Hacen a las gráficas difíciles de leer.



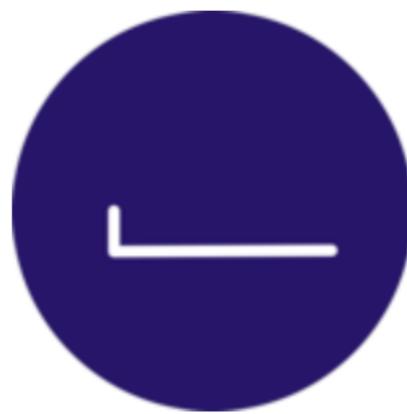
The issue with 3D



Consistency between charts



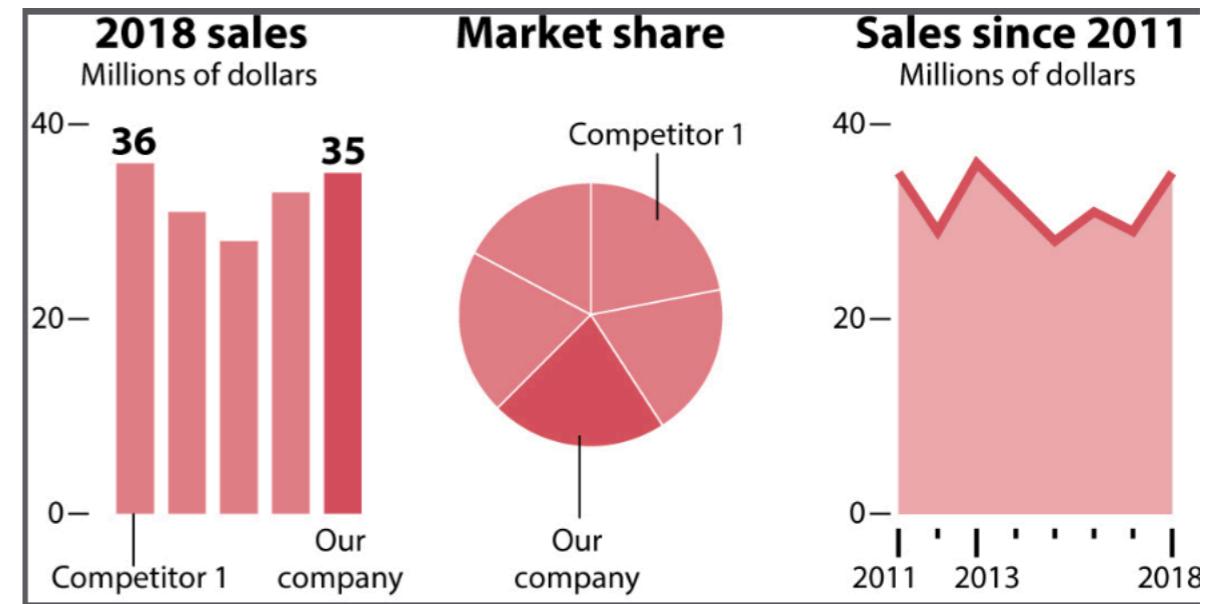
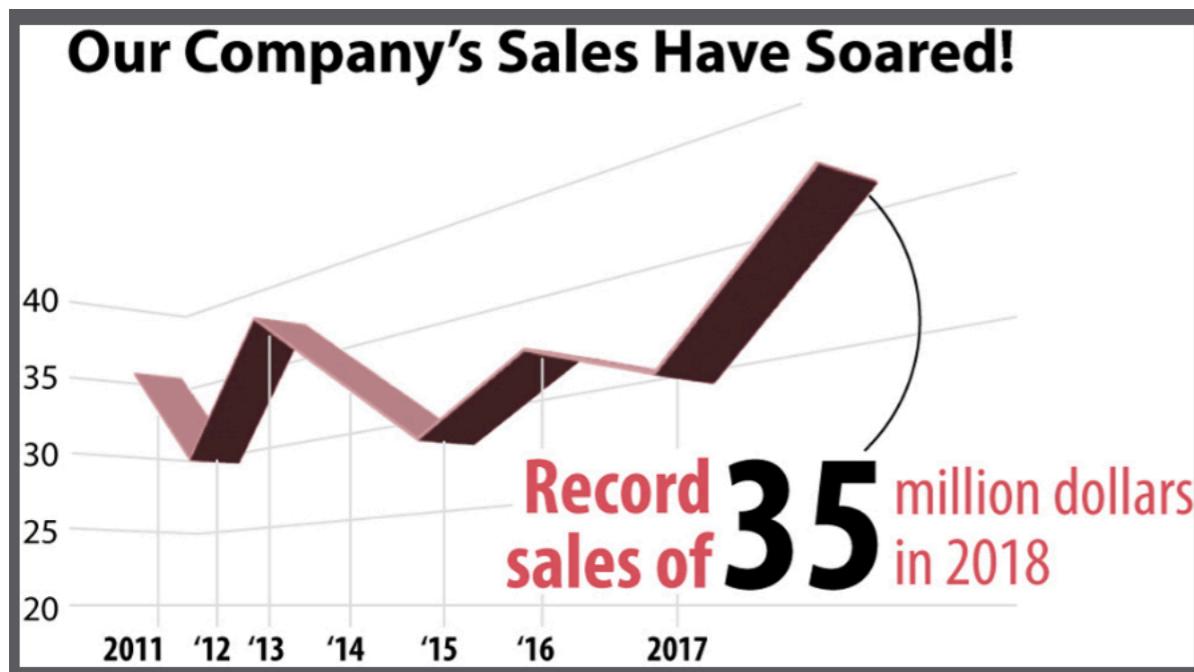
Mind your legend



Mind the aspect ratio

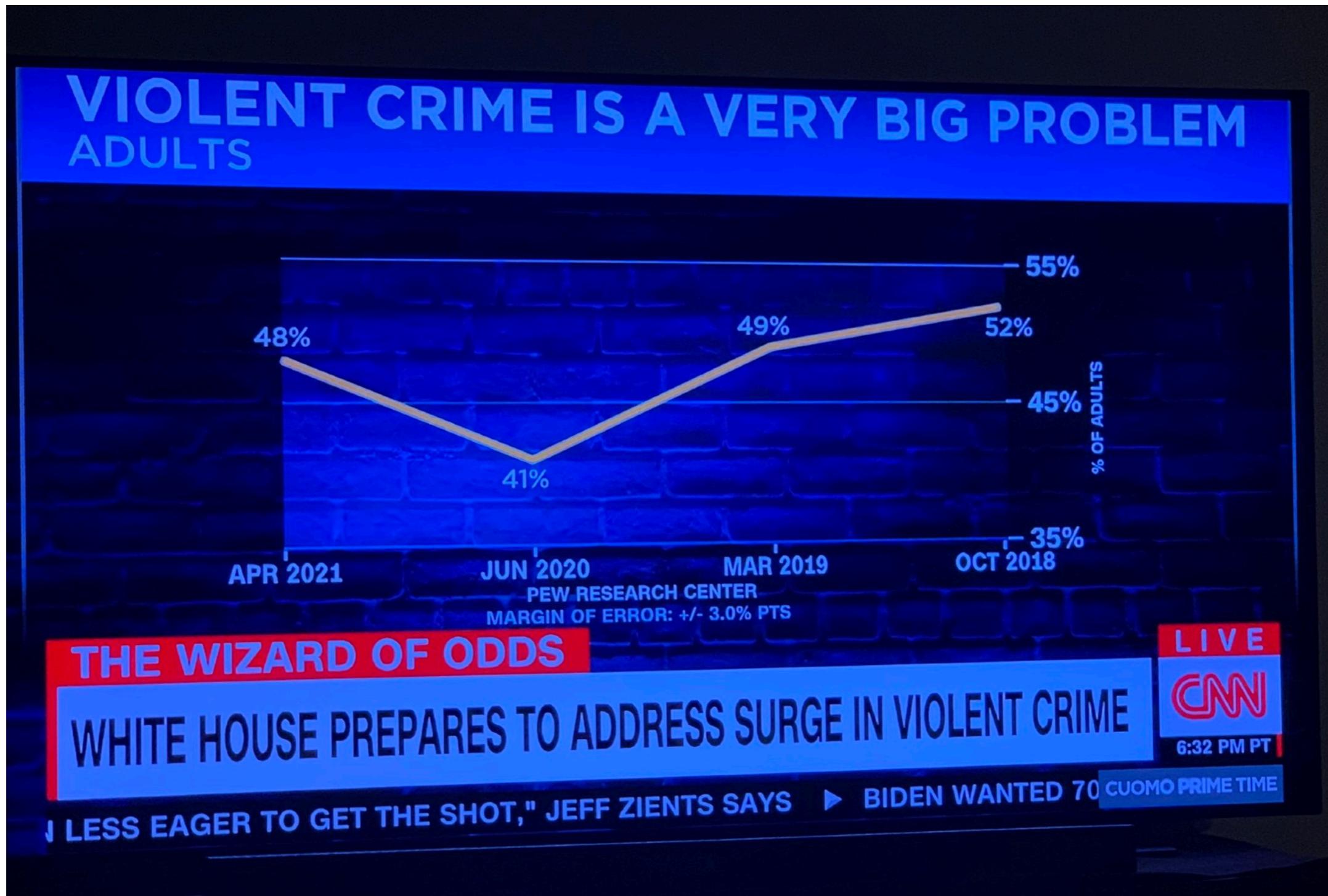
<https://www.data-to-viz.com/caveats.html>

El problema del 3D



Alberto Cairo. How Charts Lie. 2019.

Respetá las convenciones de visualización



Alberto Cairo. How Charts Lie. 2019.

Otros trucos

Más en <https://www.data-to-viz.com/caveats.html>

<https://www.data-to-viz.com/caveats.html>

En conclusión

Hacer visualizaciones es menos complejo que todo lo que acabo de mostrar.

- Es un proceso que **se va interiorizando** con el tiempo.
- Hay que **usar las herramientas adecuadas** para el medio en donde las queremos publicar.
- Hay que **darle su respectiva importancia al diseño** de nuestras visualizaciones y **justificar su propósito**.
- Hay que **evitar errores comunes** para transmitir el mensaje que queremos y no engañar a la gente.
- Hay que **probar varias alternativas** y ponerlas a prueba.
- Hay que estar dispuesto a estar **aprendiendo** constantemente nuevas técnicas y métodos.

<https://www.data-to-viz.com/caveats.html>



Gracias :3