

Visualización de Datos

Conceptos e introducción

Visualización de datos:

1. ¿Qué es y para que 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.

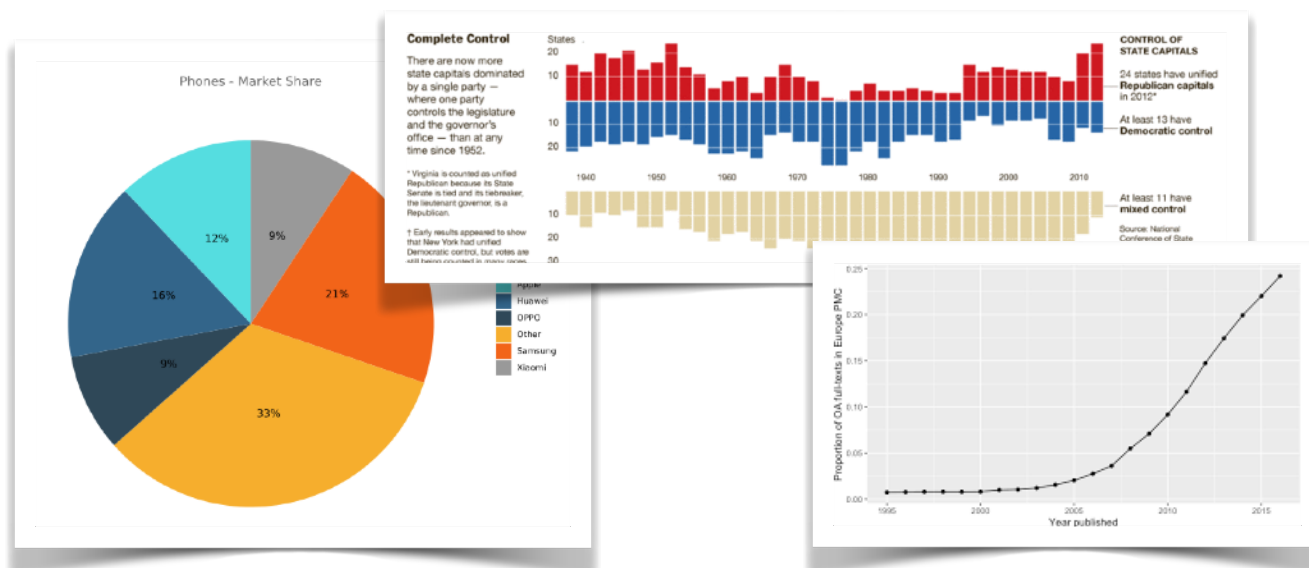
¿Qué es la visualización de datos?

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

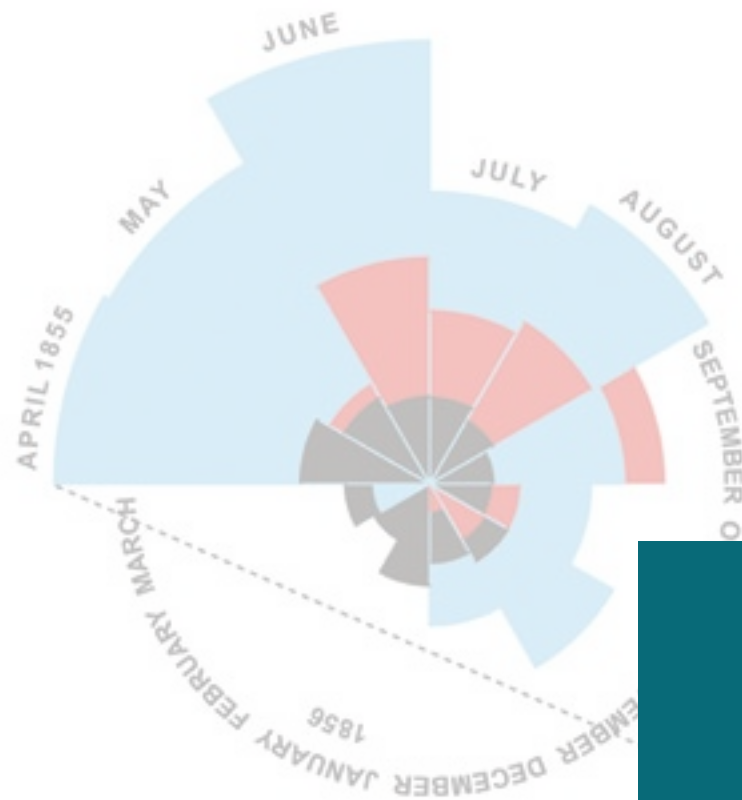
Porque vivimos en la **era de la información**, y requerimos métodos para analizar grandes cantidades de datos.



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

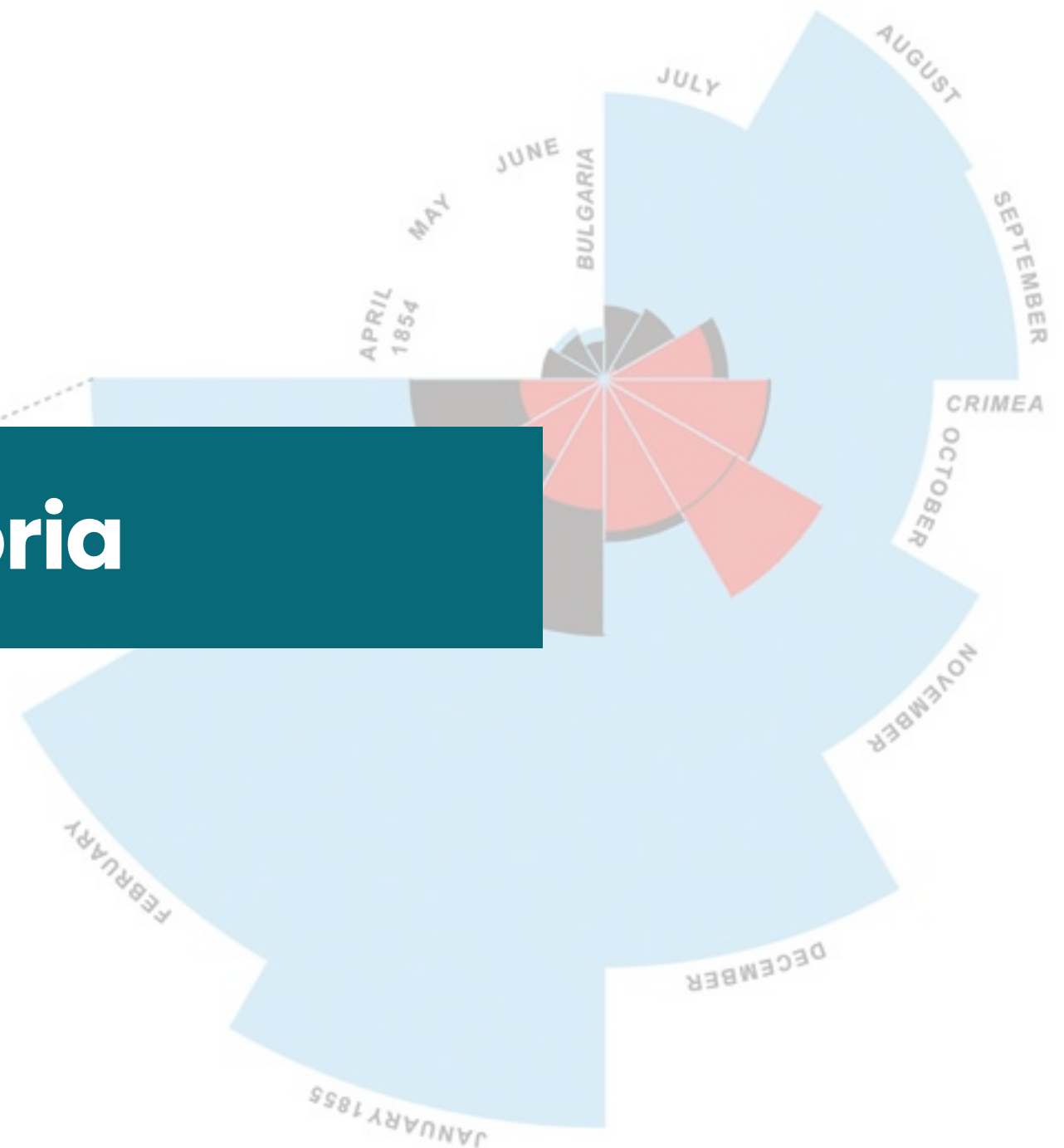
2.

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

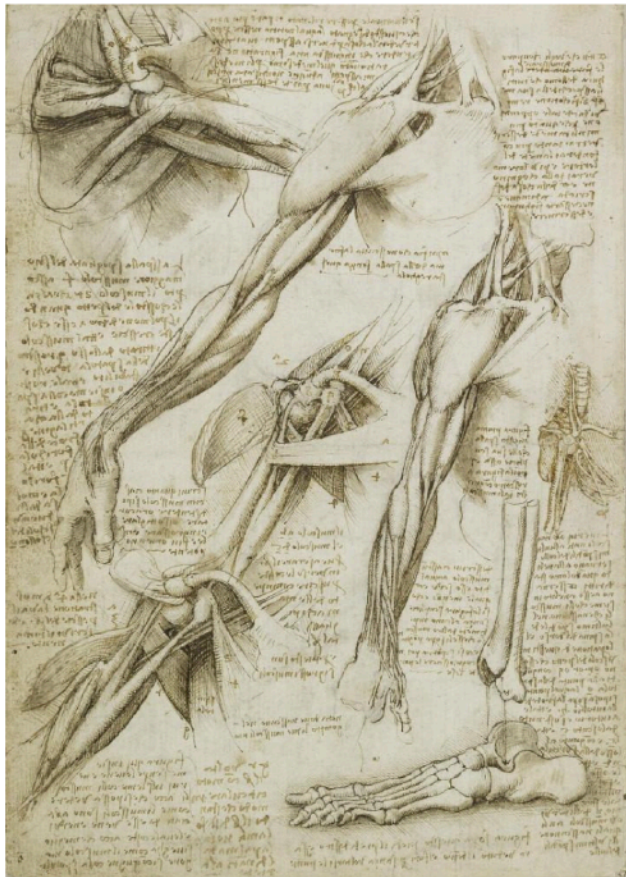
The black line across the red triangle in Nov' 1854 marks the boundary of the deaths from all other causes during the month

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

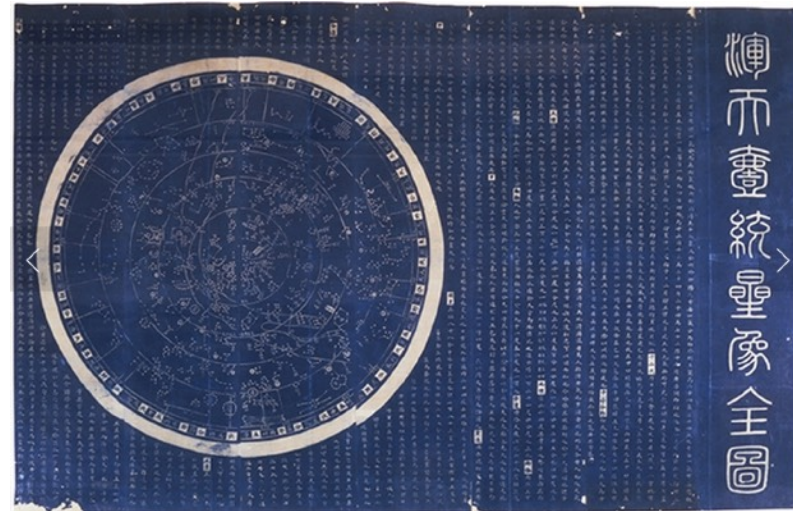
The entire areas may be compared by following the blue, the red & the black lines enclosing them. ©hugh-small.co.uk

Historia

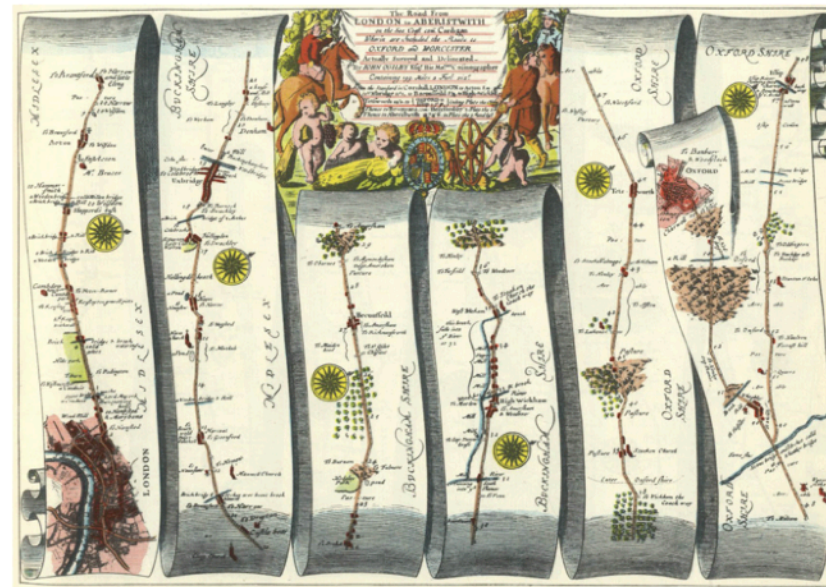
Las visualizaciones surgieron para **registrar datos**.



Da Vinci, 1510-1511



Carta estelar china



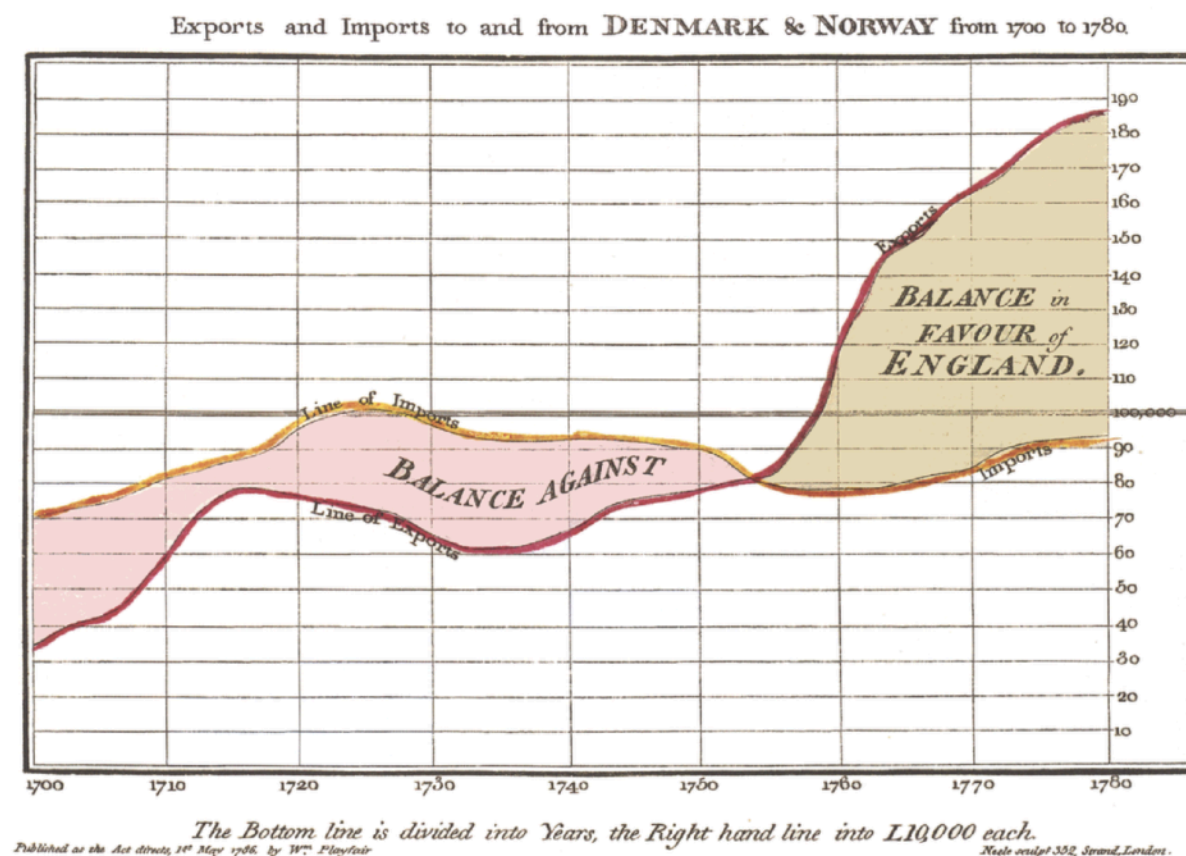
John Ogilby, el camino de
Londres a Aberistwith



Carta general de la
república mexicana

Historia

Las visualizaciones surgieron para **analizar datos**.



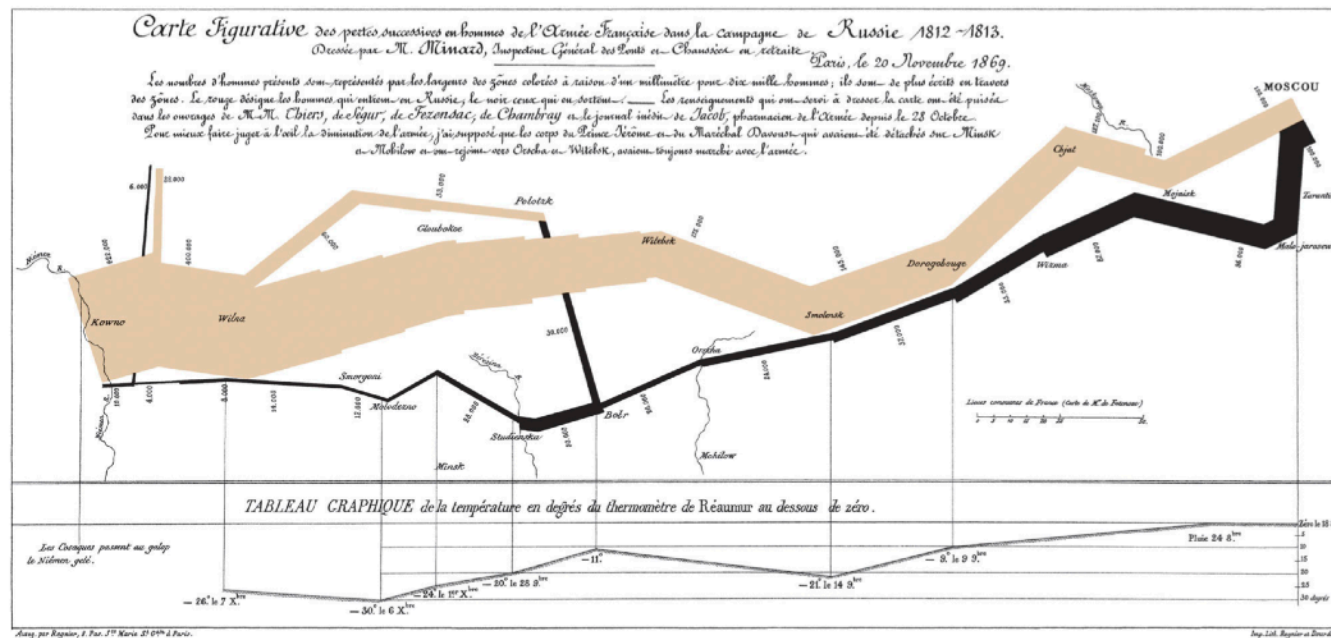
William PlayFair (1786)



John Snow (1854)

Historia

Las visualizaciones surgieron para **comunicar**.



Charles Joseph Minard
(1869)



Railways of London
Harry Beck, 1933

Historia

LA VISUALIZACIÓN DE DATOS PUEDE SALVAR VIDAS

 **Florence Nightingale**
Fue enfermera y aficionada de las matemáticas.

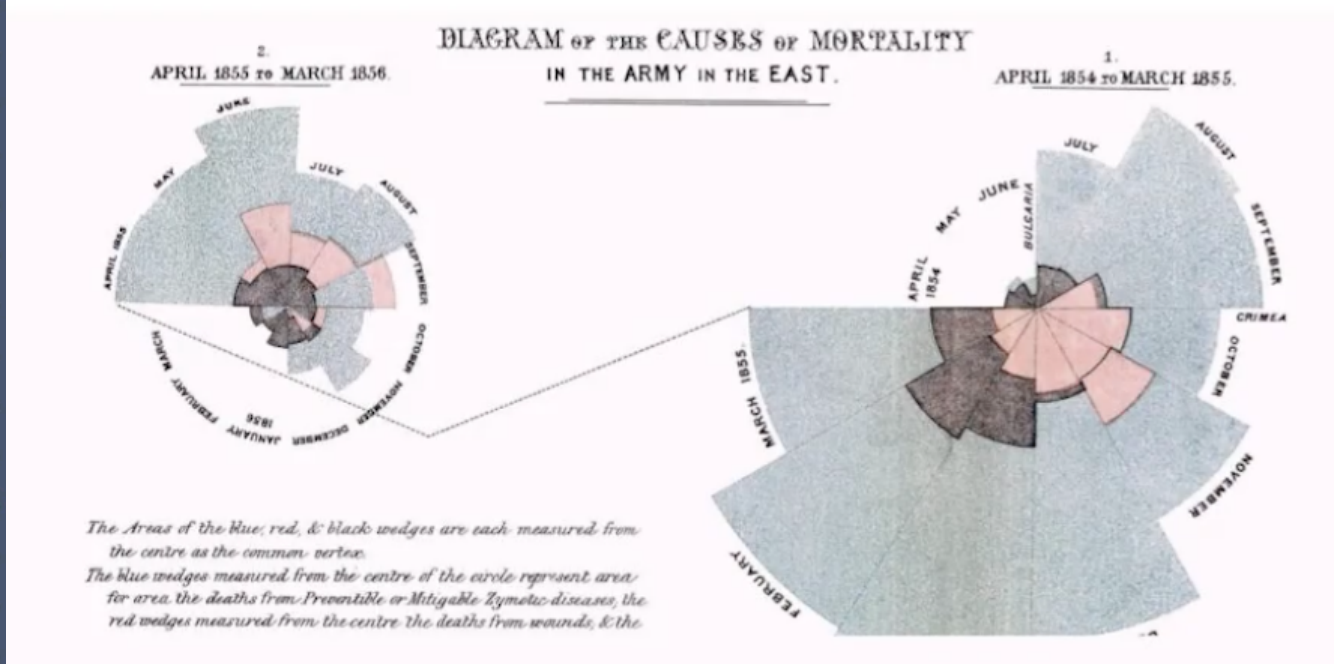
Durante la Guerra de Crimea **registró todas sus actividades** en un diario.

Demostró al gobierno británico que **la mayor causa de muerte no eran heridas de guerra, sino las infecciones.**

Hoy gracias a ella **sabemos la importancia de implementar buenos sistemas** de salud para salvar millones de vidas.

Comunica tus resultados con datos y storytelling en www.platzi.com/salvavidas

 **Platzi**



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:

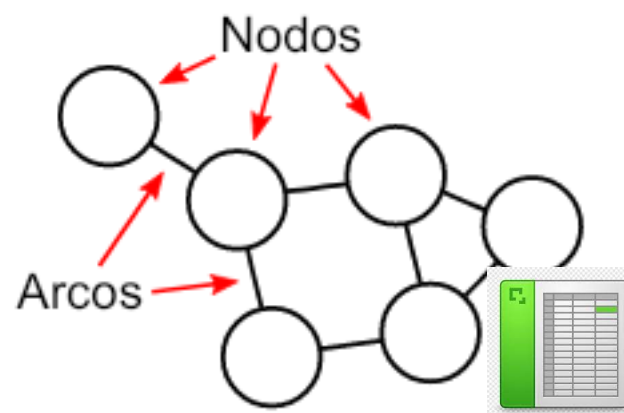
1. **Unidades** (entidades individuales discretas).
2. **Variables** (propiedades medibles u observables).
3. **Enlaces** (relación entre unidades/nodos).
4. **Posición** (ubicación en 2D o 3D).

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:

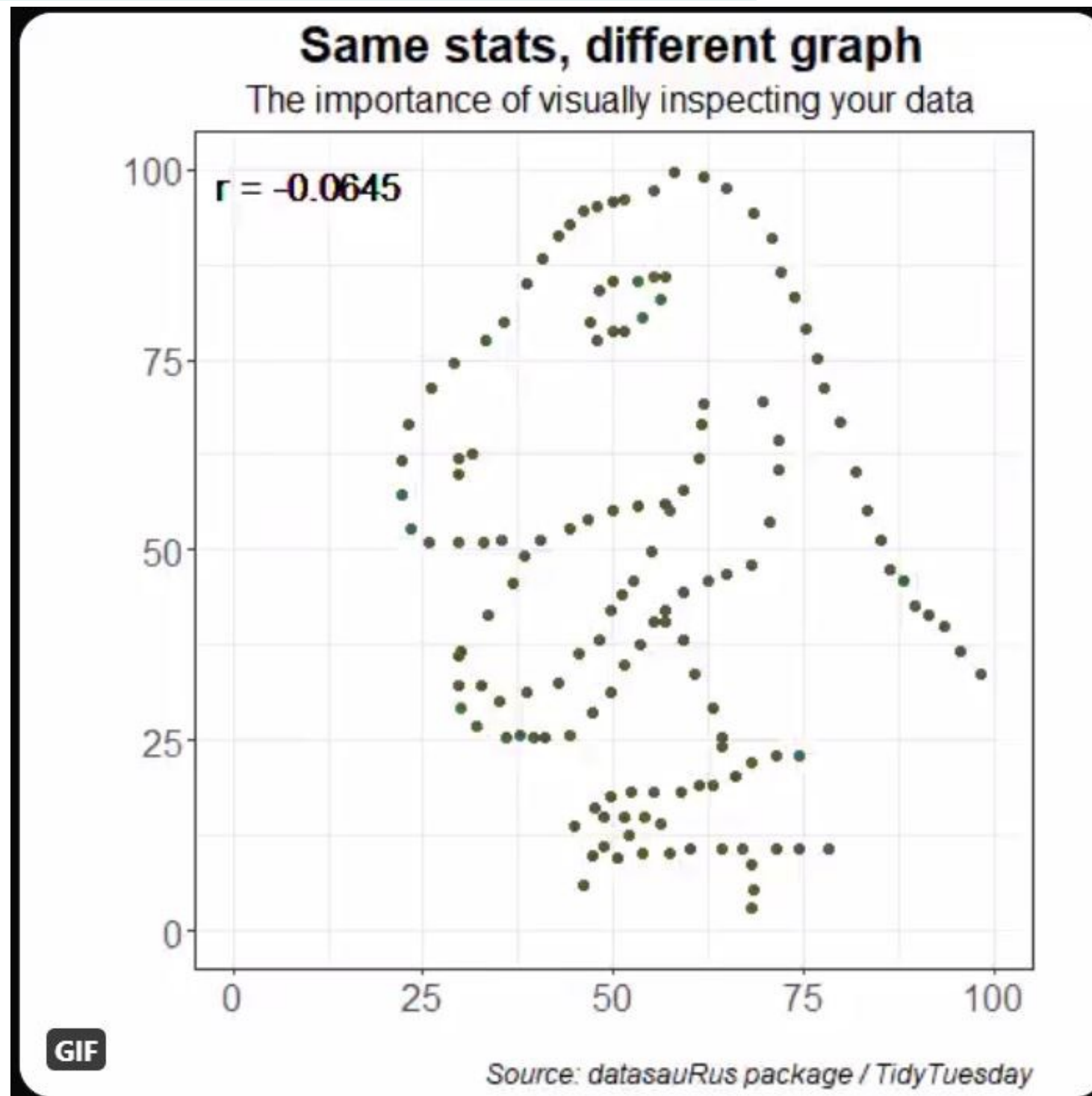
- 1) El **tipo de dato** que representan
- 2) El **número de dimensiones** que requieren

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



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

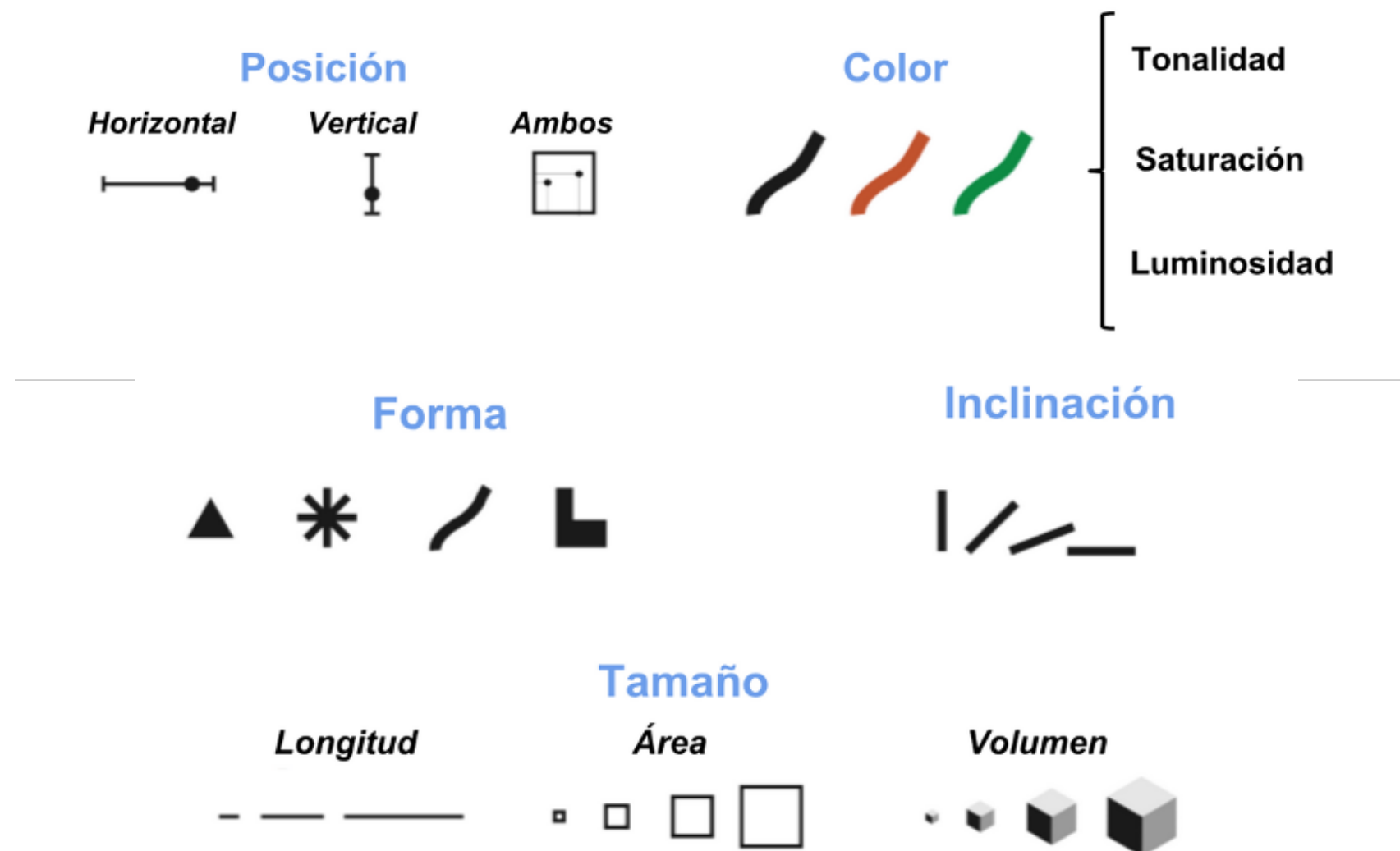
Marcas



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*..



Productos de la visualización

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



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

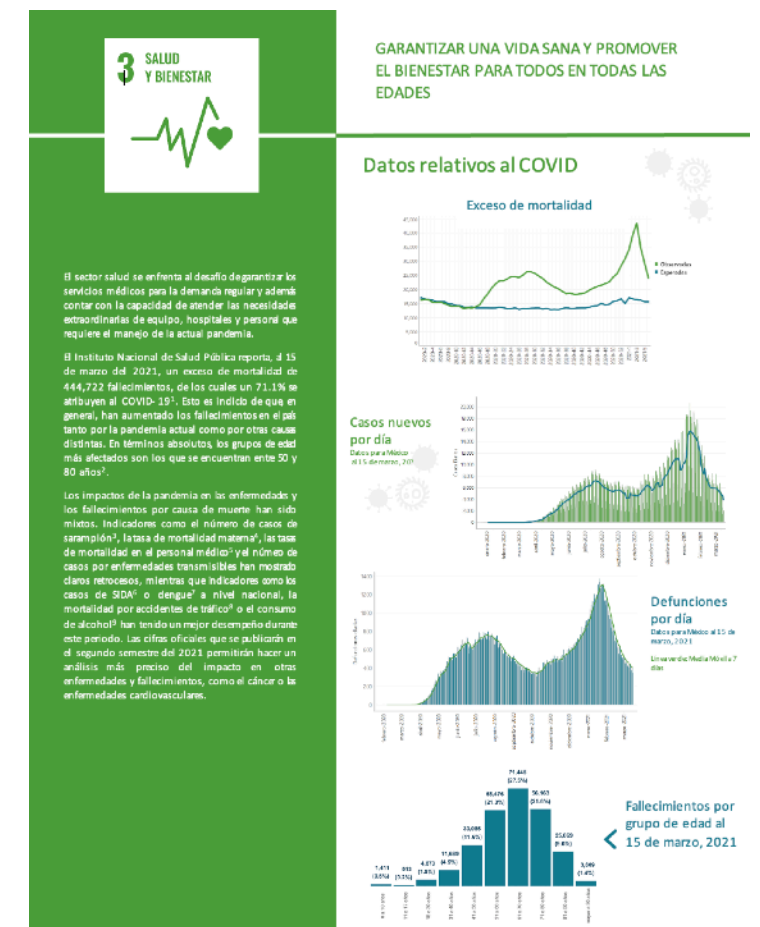
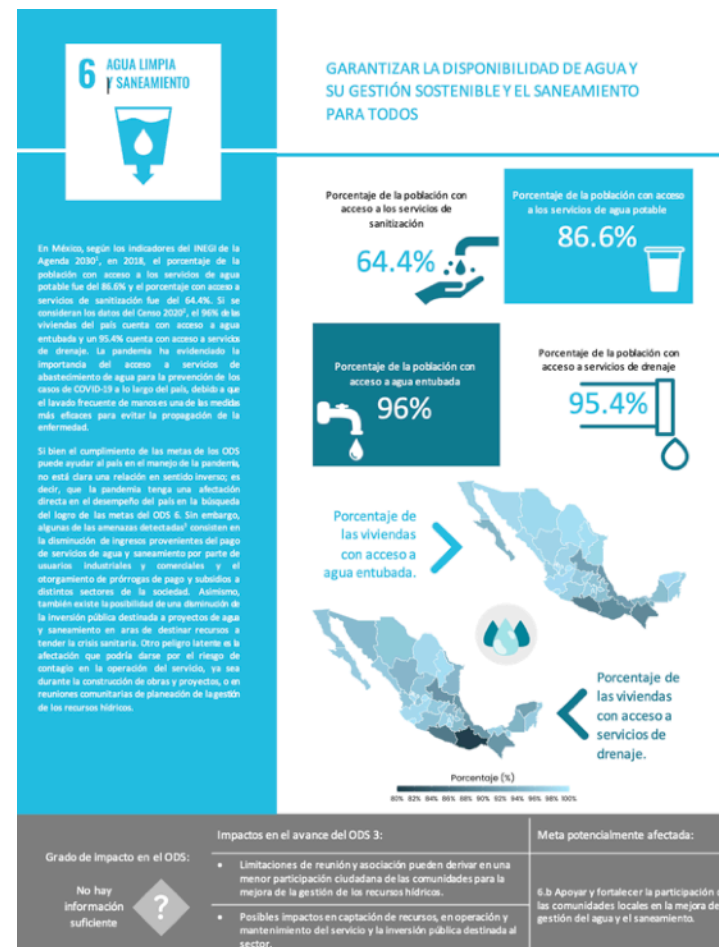
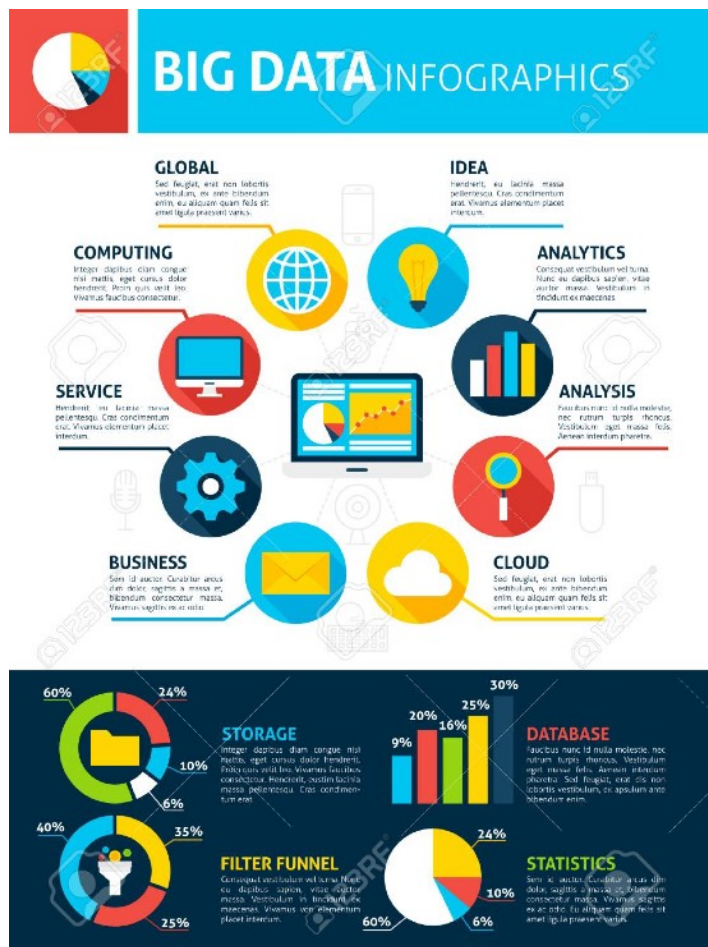
Gráfica interactiva



- 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 dijeron 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 fue "campeón" (2,468 veces), seguida de "gana" (2,101) y, por supuesto, "gol" (1,839). De hecho, éste se gritó fuerte: "gol" (dicho 1,839 veces), "gool" (20), "goolol" (43), e incluso "gooooool", así, con seis ceros, dicho más de 40 veces: sin olvidar el "golazo" (220) y sus variantes.

Cruz de las 200 palabras más comunes.
Elaboración en RStudio: Jesús Díaz

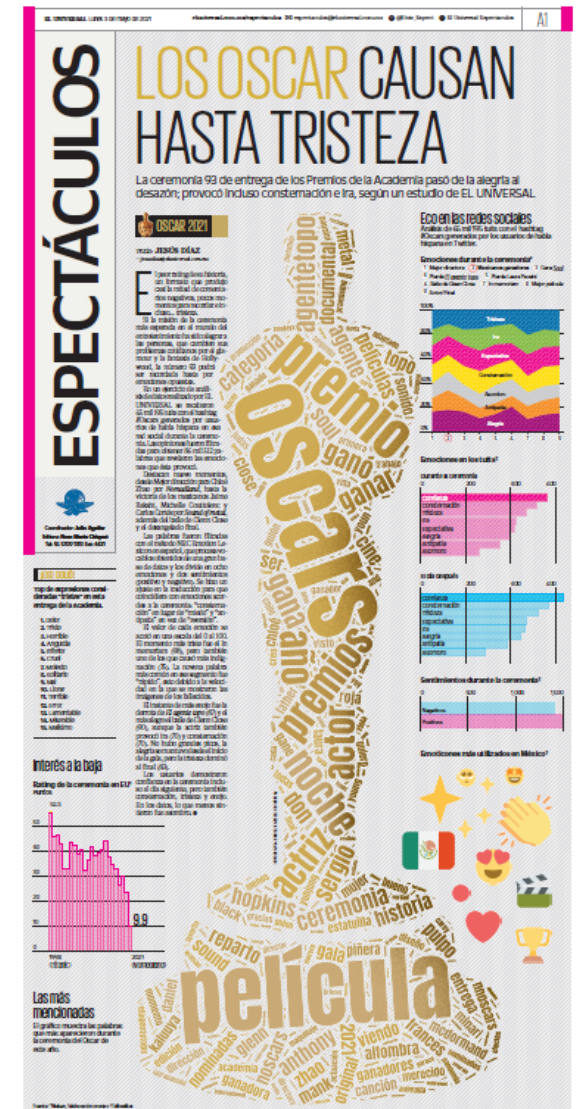
"Cruzazuleada" y "cruzazulear" representan un estigma, casi una maldición para La Máquina, por ello no fueron tan mencionadas, apenas 146 y 123 veces, respectivamente.

La palabra "maldición" más común (350 veces), y otras no propias de fútbol como "fantasmas" (247) y "miedo" (236).

La palabra "maldición" más común (350 veces), y otras no propias de fútbol como "fantasmas" (247) y "miedo" (236).

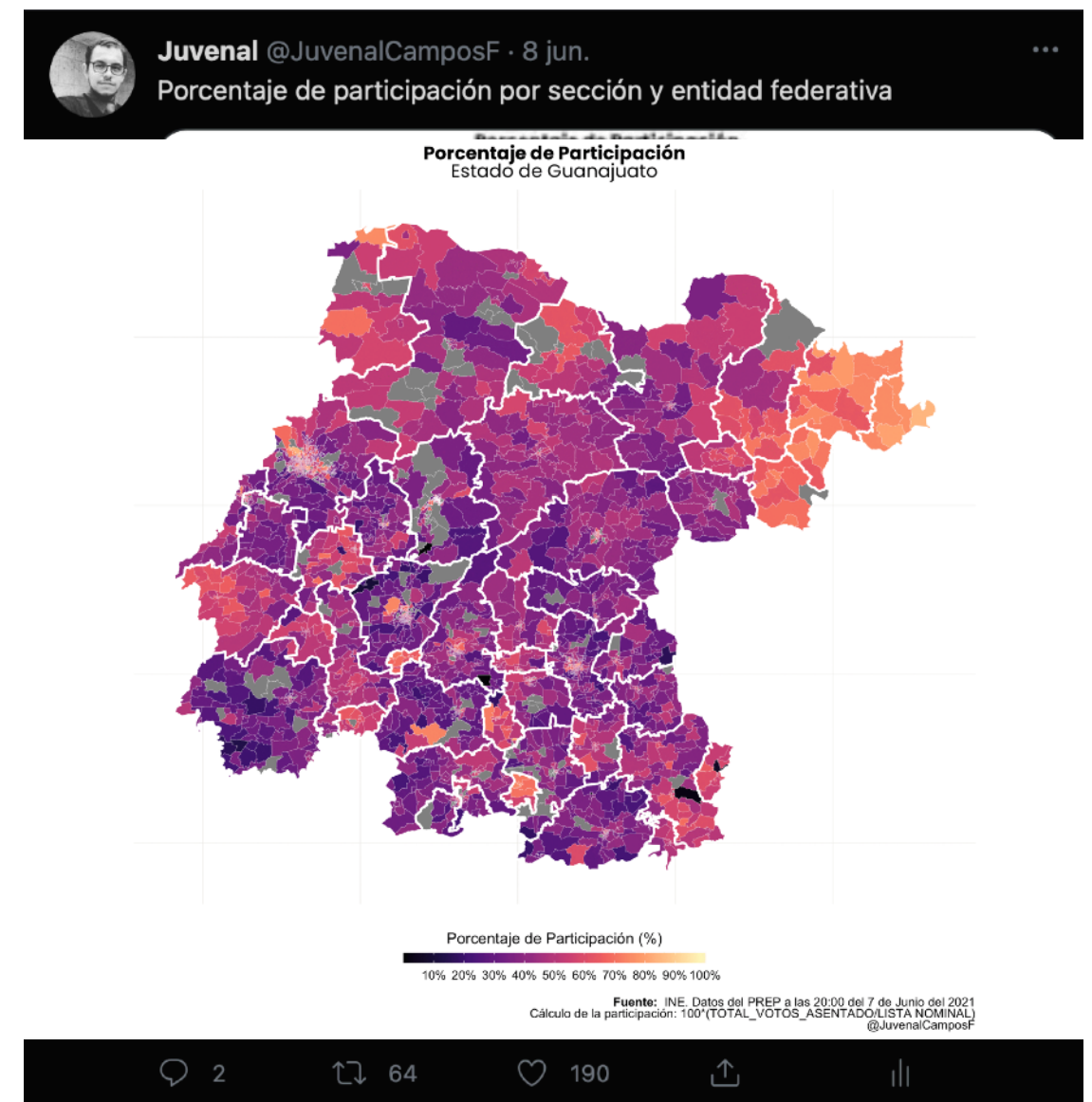
Hubo tiempo para bromear y pasarla bien, el emoji más usado fue el "Cara llorando de risa" (259 veces), además de que se escribieron carcajadas: "jajaja" (377 ocasiones), "jajajaja" (292), "jaja" (190), "jajajajaja" (129).

Lo mejor: la esperanza, "esta es la buena" destacó 780. Y si, fue la buena. ■



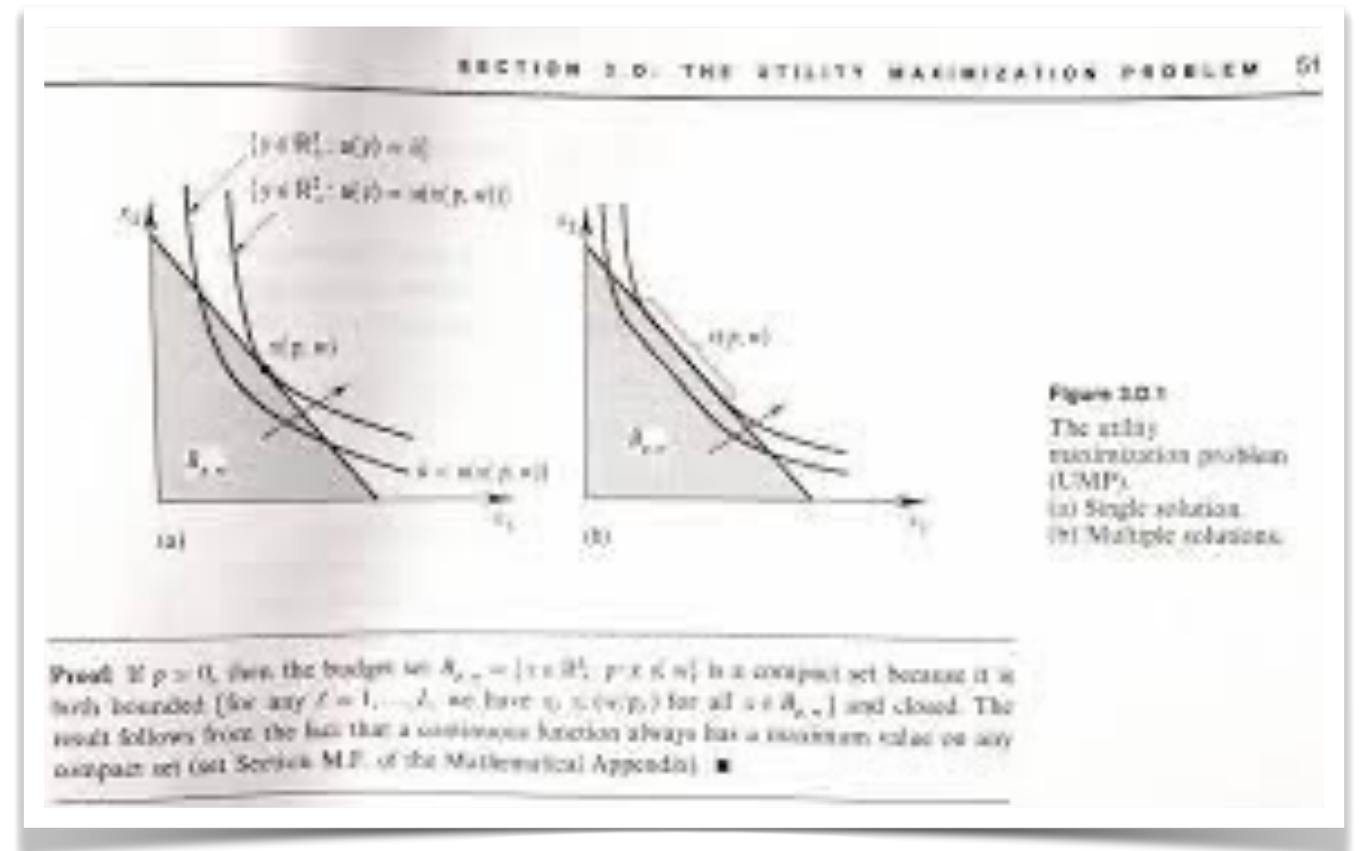
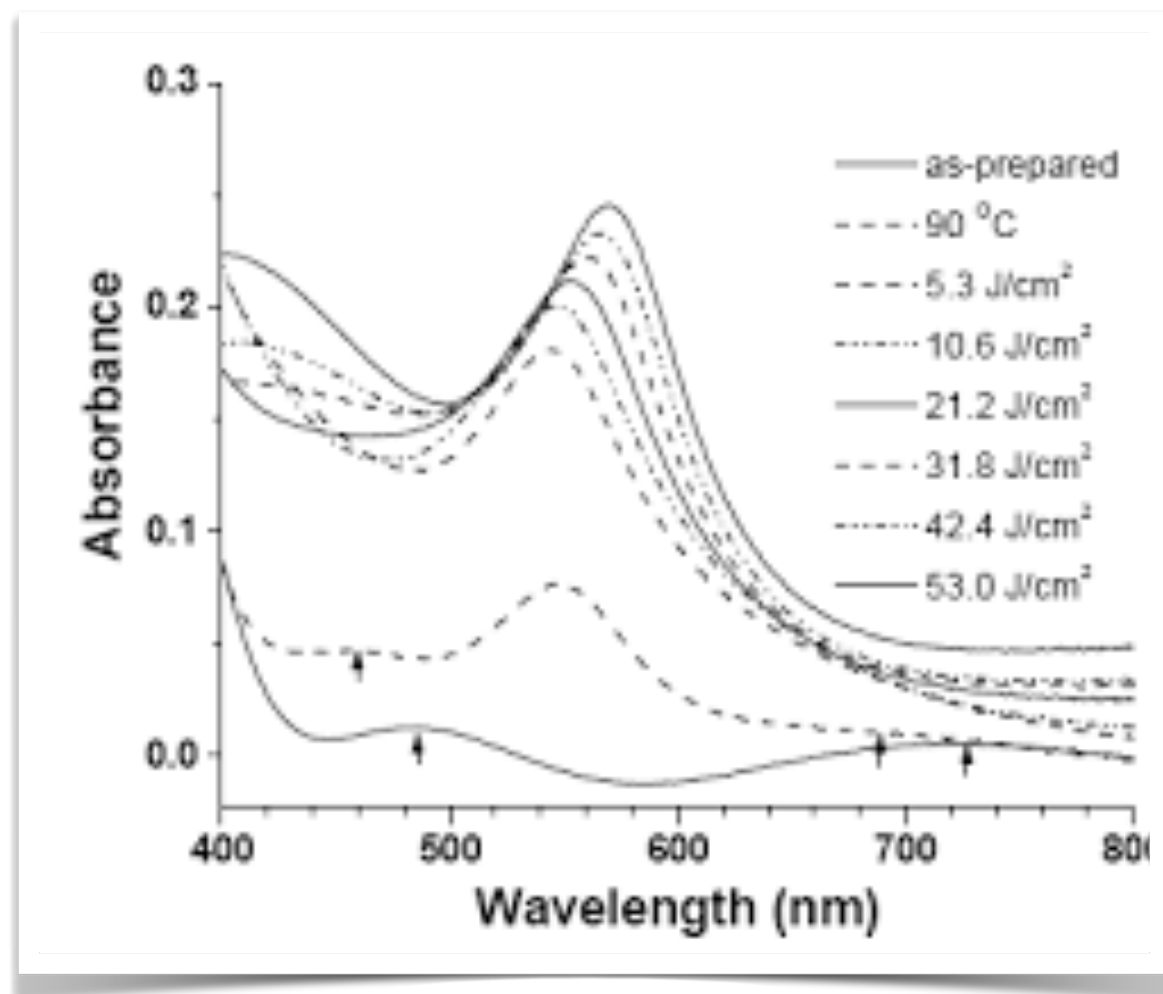
Gráfica sola (@RRSS)

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



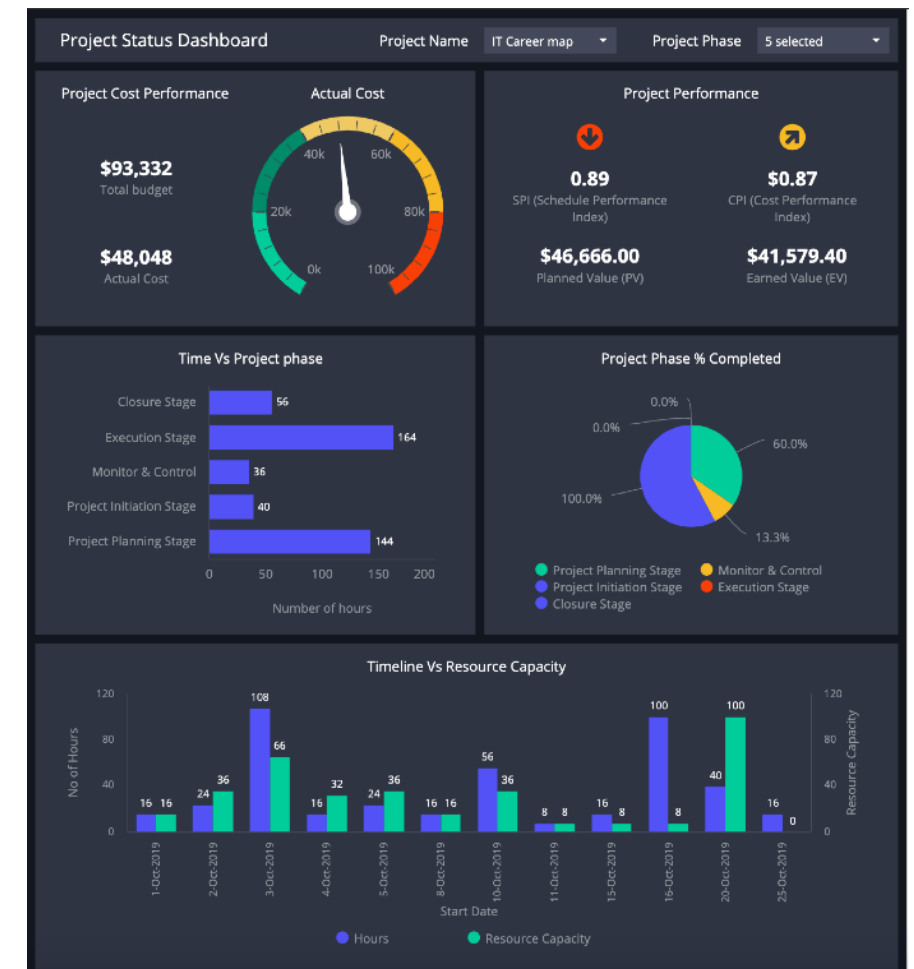
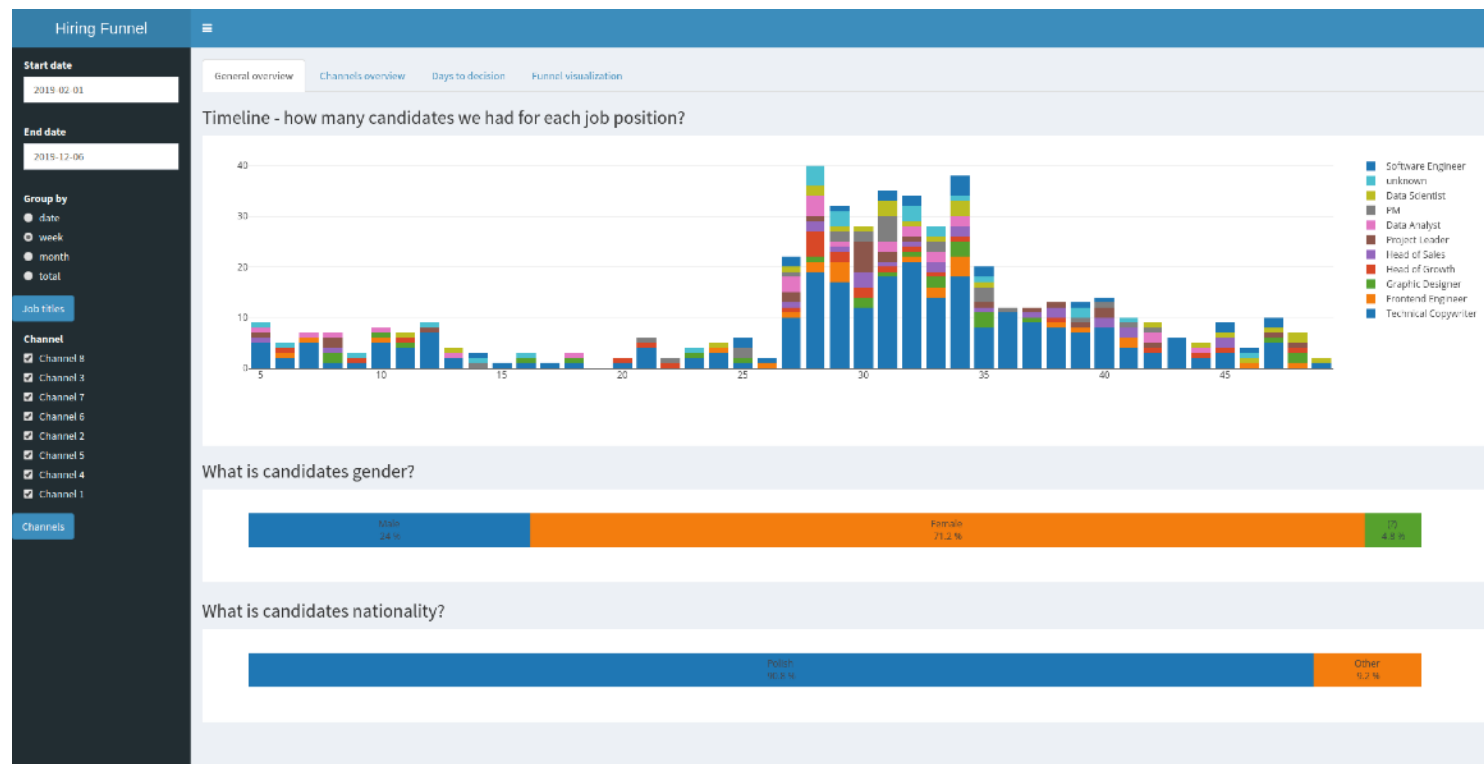
En artículos científicos

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



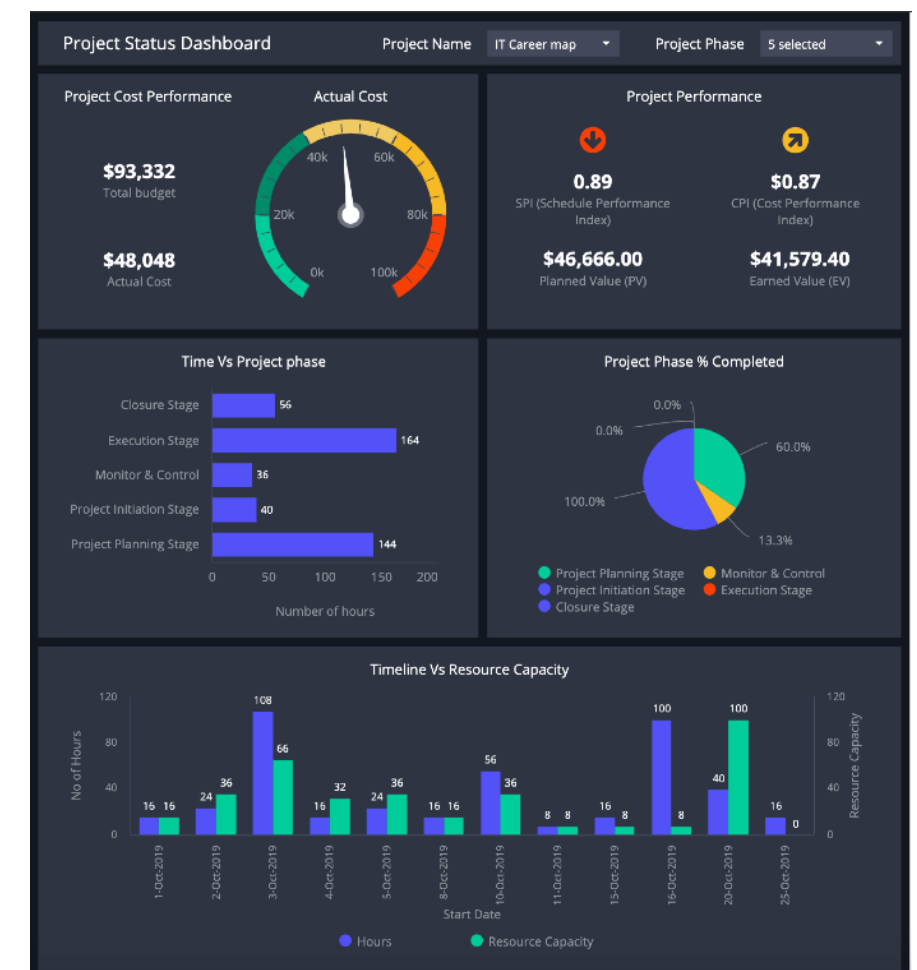
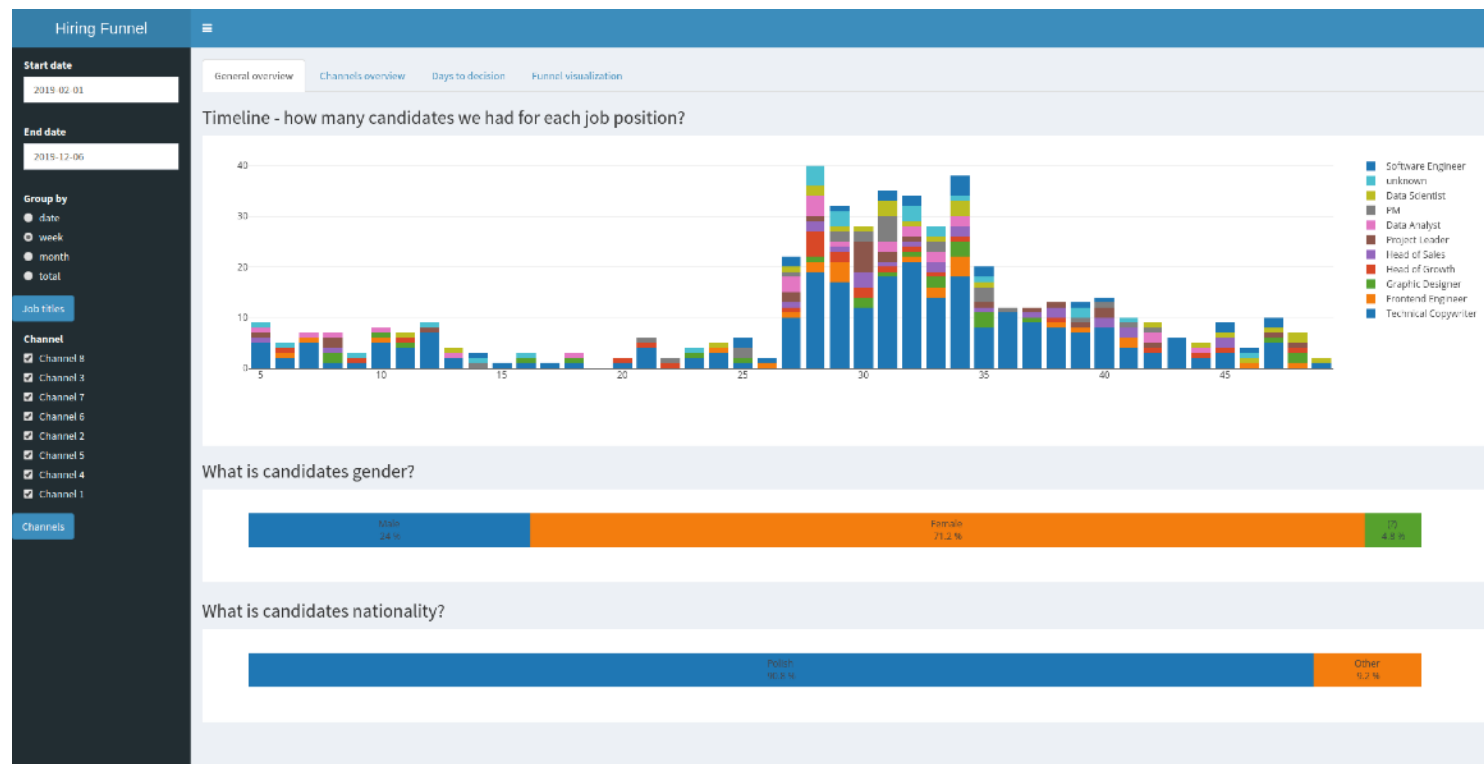
En dashboards

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

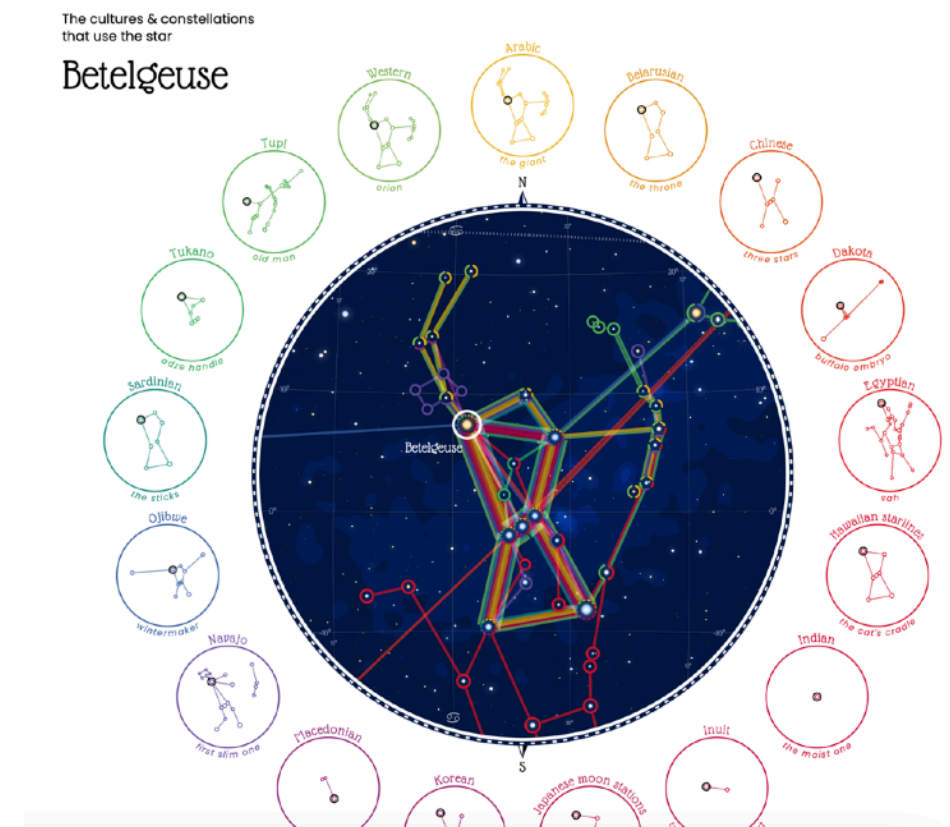
El reto de diseñar las visualizaciones en un dashboard es que **se diseñan varias visualizaciones** que deben mostrar información relevante, deben **trabajar en conjunto** y deben **adaptarse a cambios en los datos**.



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**.



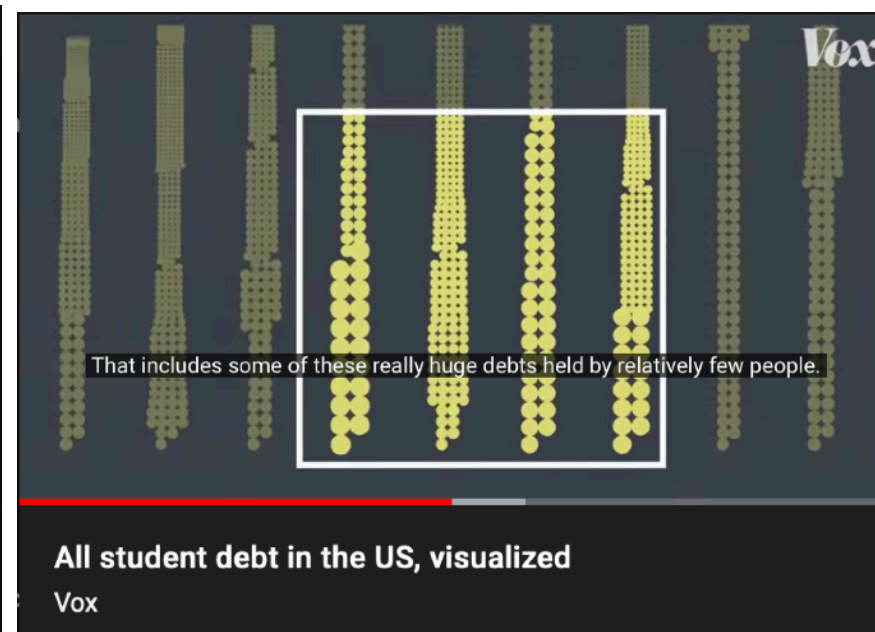
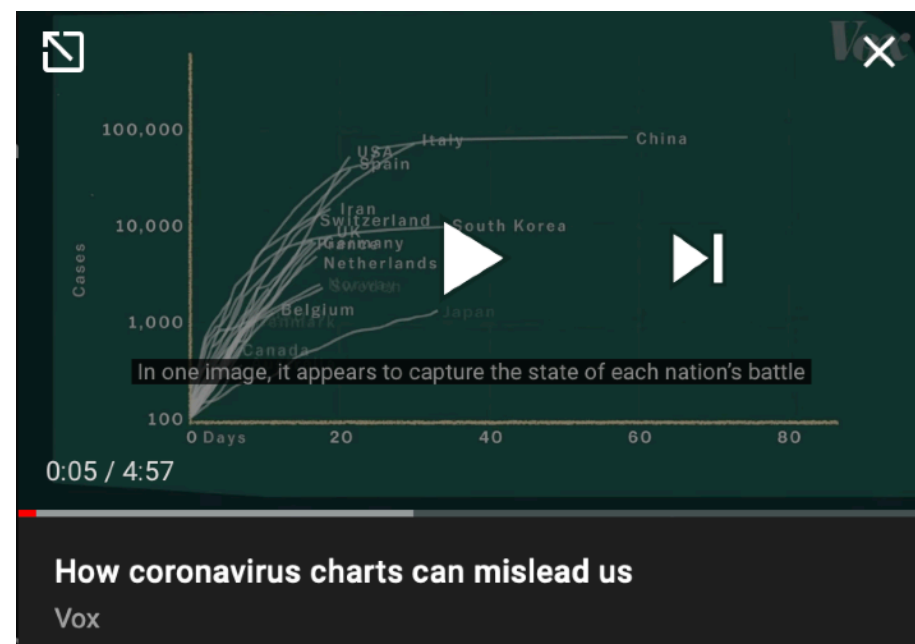
<https://whydocatsanddogs.com>



<https://figuresinthesky.visualcinnamon.com>

Video

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



07_ShinyApp_Municipal - RStudio

funciones.R x styles.css x app.R x server.R x

Go to file/function Addins

Run App

abstenci Next Prev All Replace 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 ----
13
```

Creación

38:92 # Aplicacion

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"

> |
```

Environment History Connections Bui

Import Dataset List

R Global Environment

Data

entid...	32 obs. of 6...
mapa	2468 obs. of...
mapa_...	2468 obs. of...
mapa_...	11 obs. of 2...
muni_...	11 obs. of 1...
tema_...	List of 13

Values

Files Plots Packages Help Viewer

¿Cómo crear visualizaciones de datos?

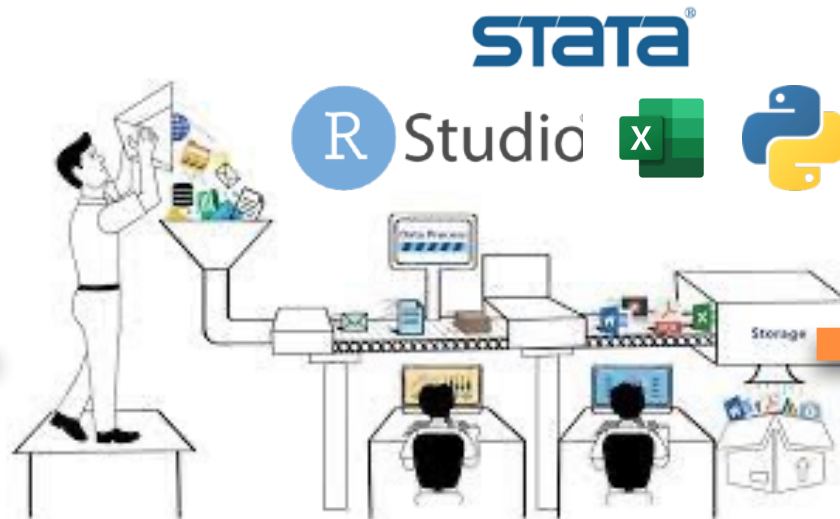
Proceso:

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?



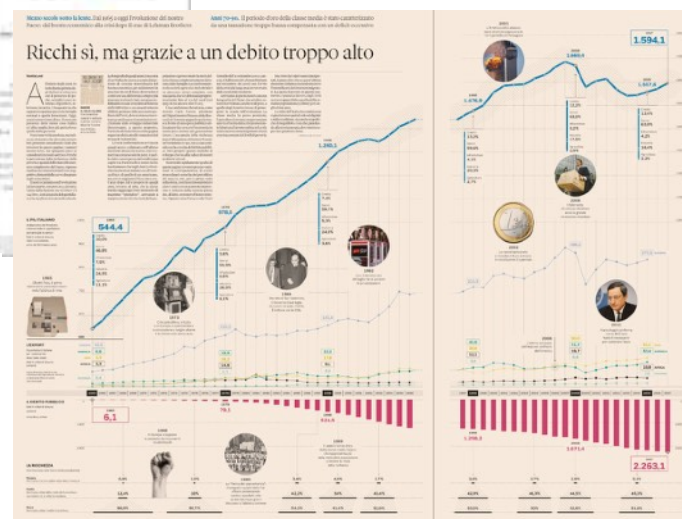
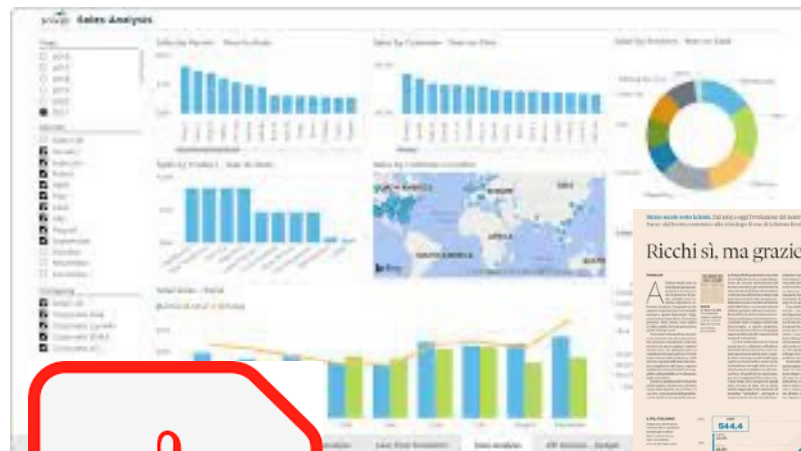
Obtención de los datos



Procesamiento



Diseño y creación



Publicación



Post-producción

Herramientas de trabajo

Análisis, procesamiento y visualización



Creación de Dashboards



Tecnologías web



Librerías de visualización



Herramientas de post-producción y edición





Diseño

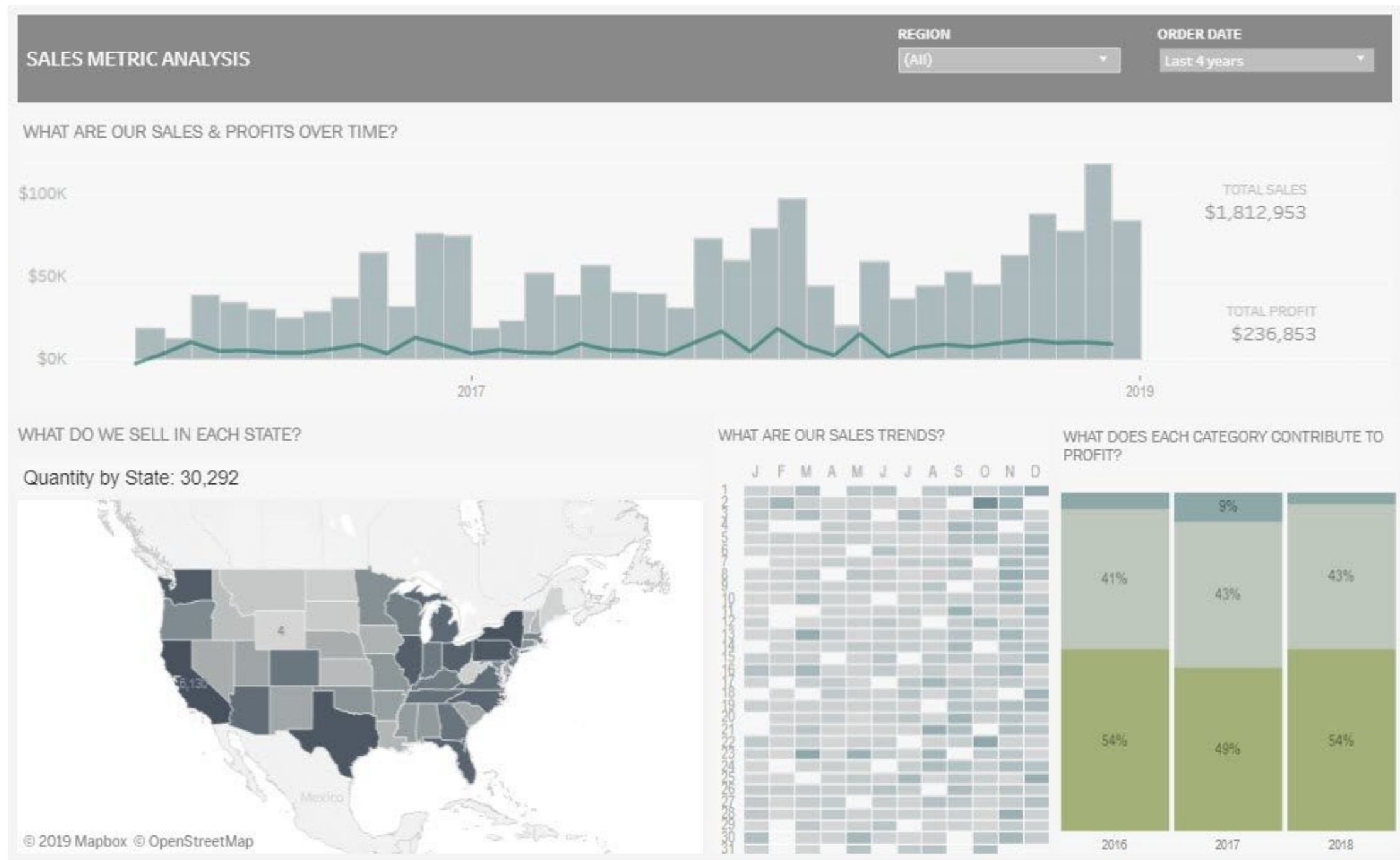
Preguntas de diseño

Preguntarse...

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



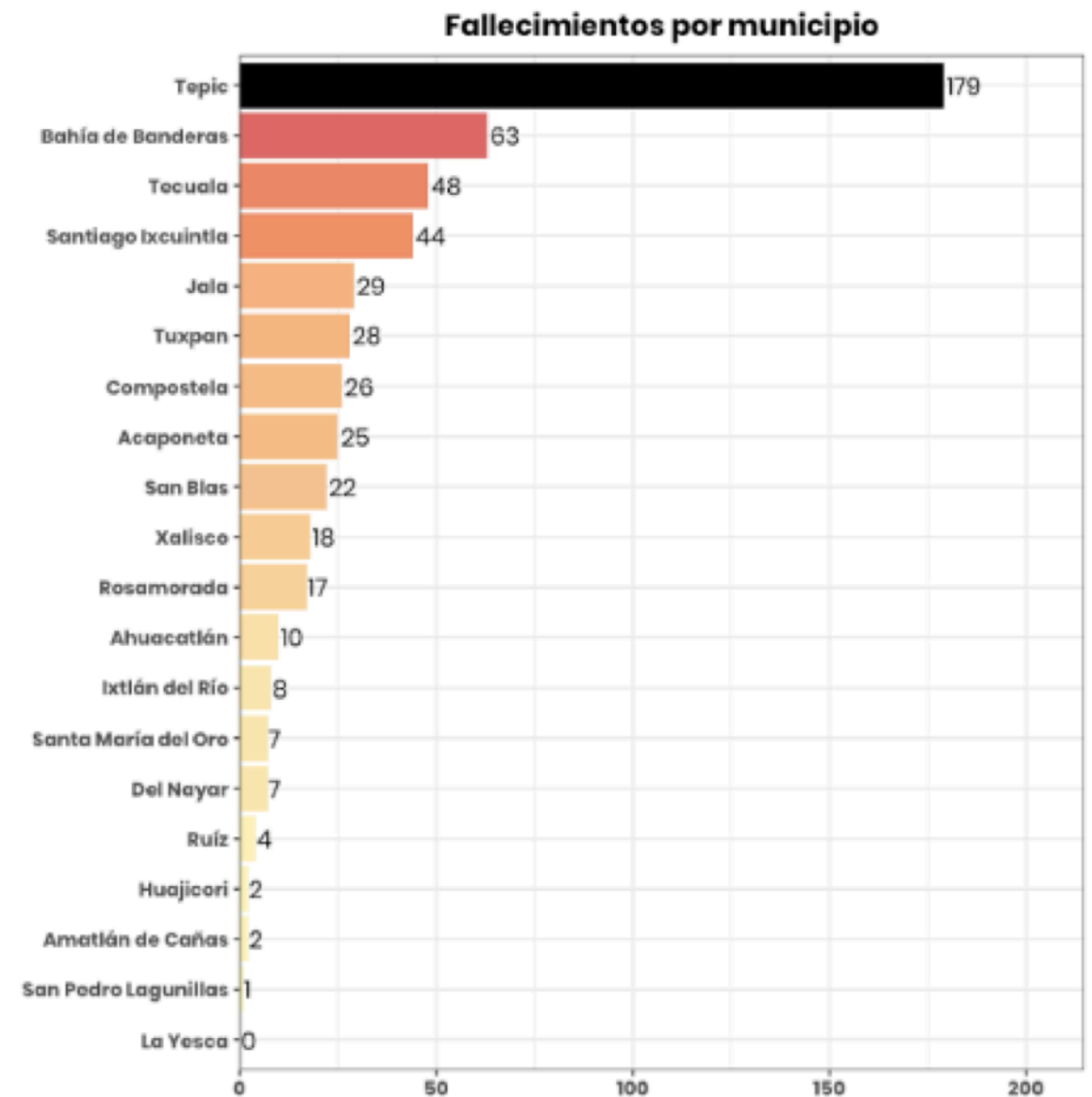
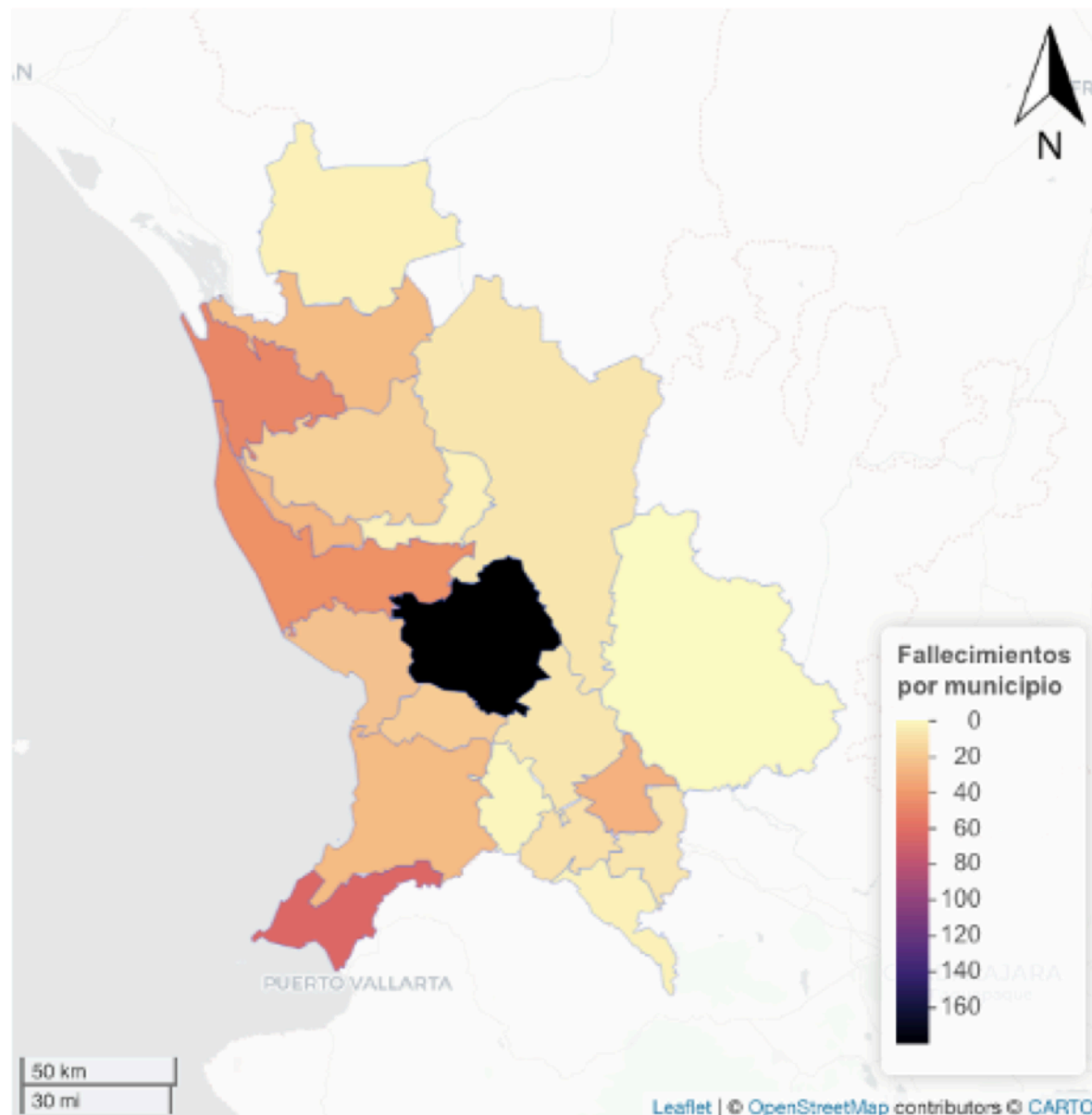
Preguntas de diseño



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

Escogiendo la mejor gráfica

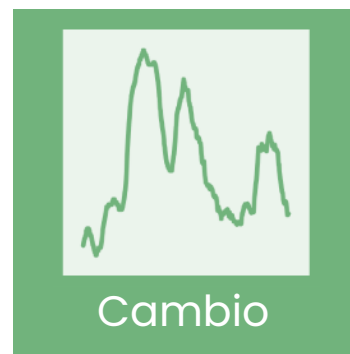
¿Cuál es **mejor**? ¿Qué **variable** y **mensaje** da cada una?



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.



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



from Data to Viz

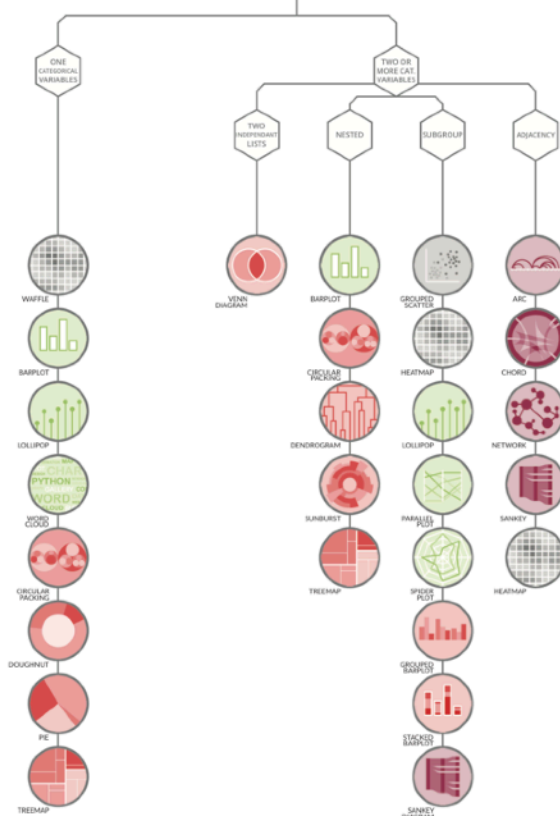
'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps:

- 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.

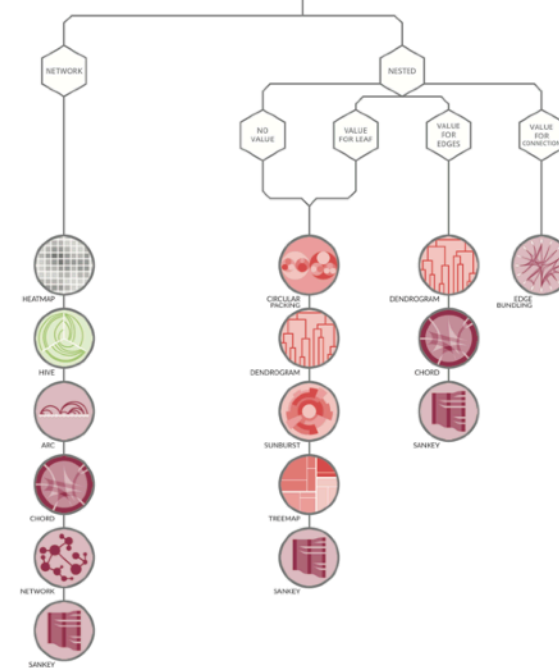
Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

[data-to-viz.com](https://www.data-to-viz.com)

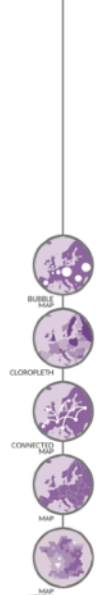
CATEGORIC



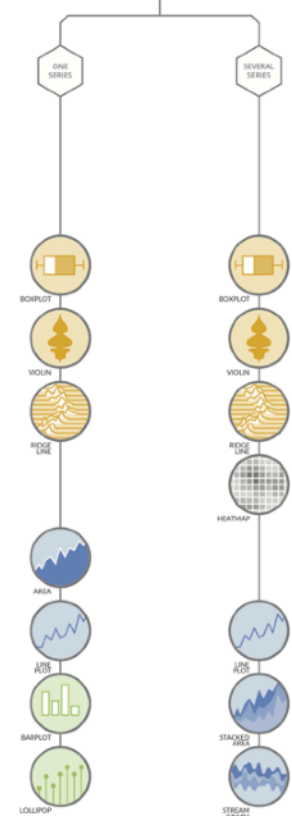
RELATIONAL



MAP



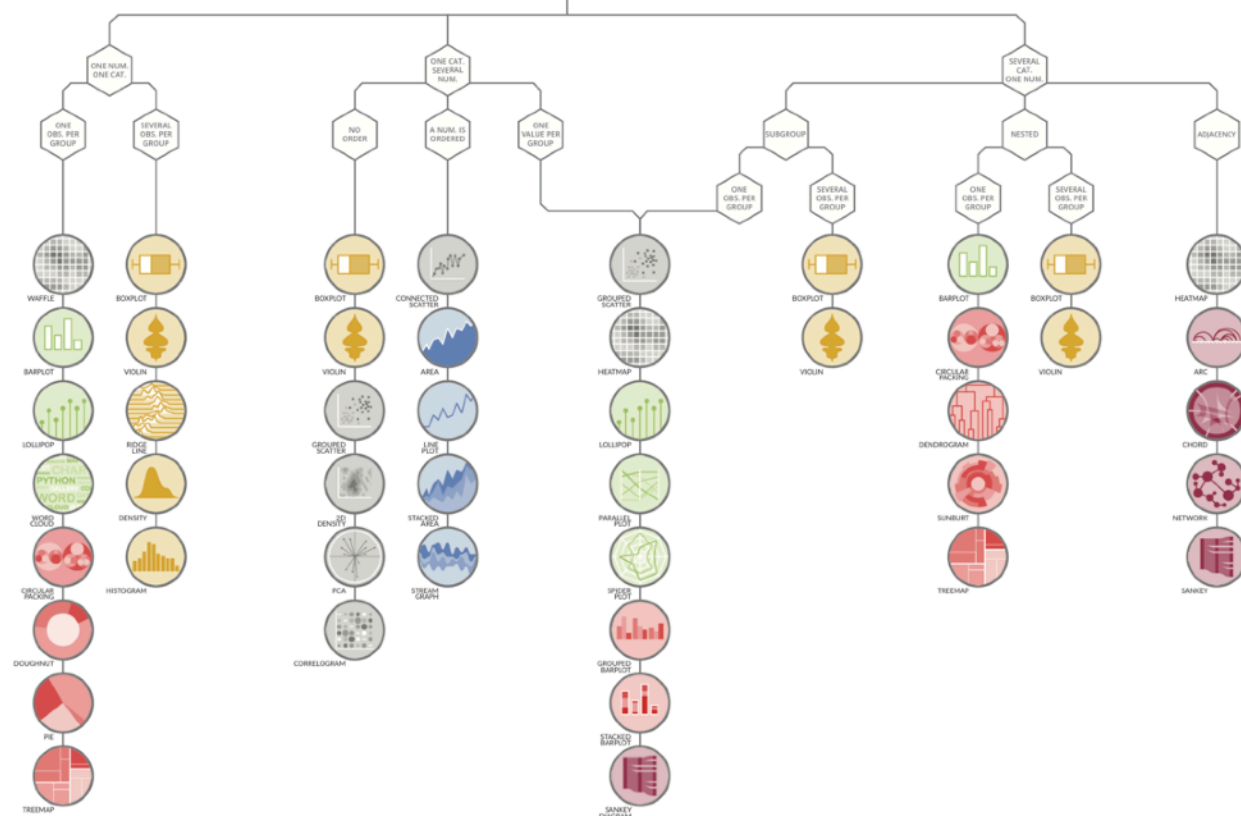
TIME SERIES



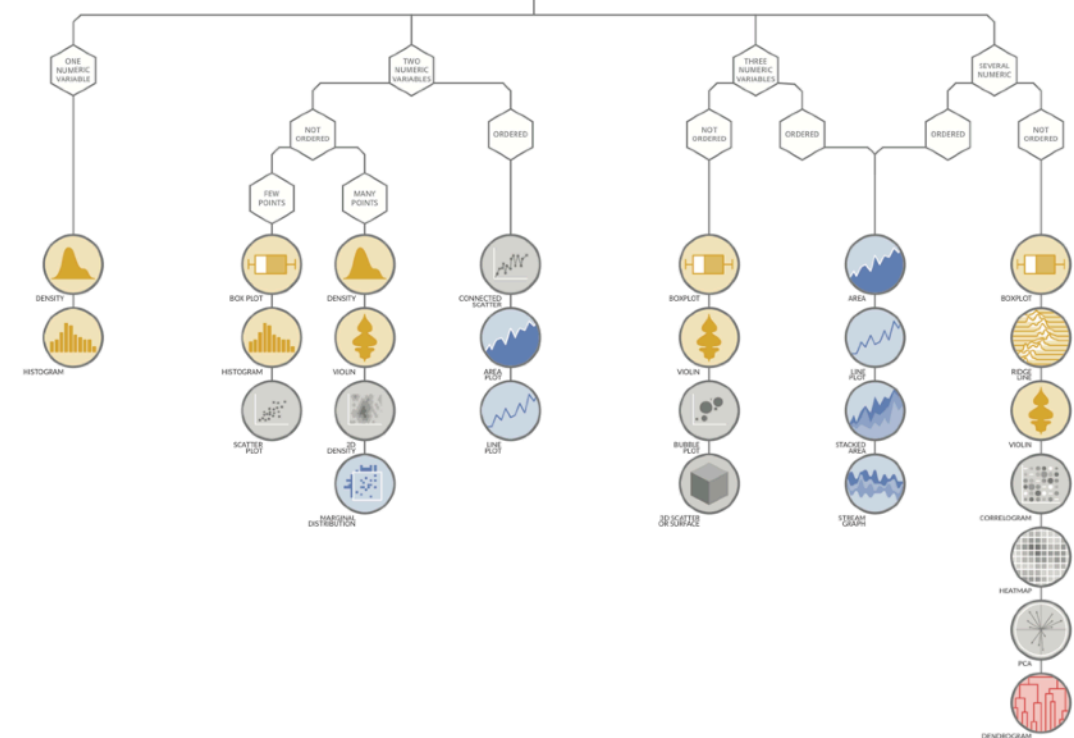
WHAT DO YOU WANT TO SHOW ?

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

CATEGORIC AND NUMERIC



NUMERIC



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

Uso de colores

Regla 60-30-10



Uso de colores neutros



Armonización de colores



Complementary



Analogous



Triadic



Split-Complementary



Rectangle (tetradic)



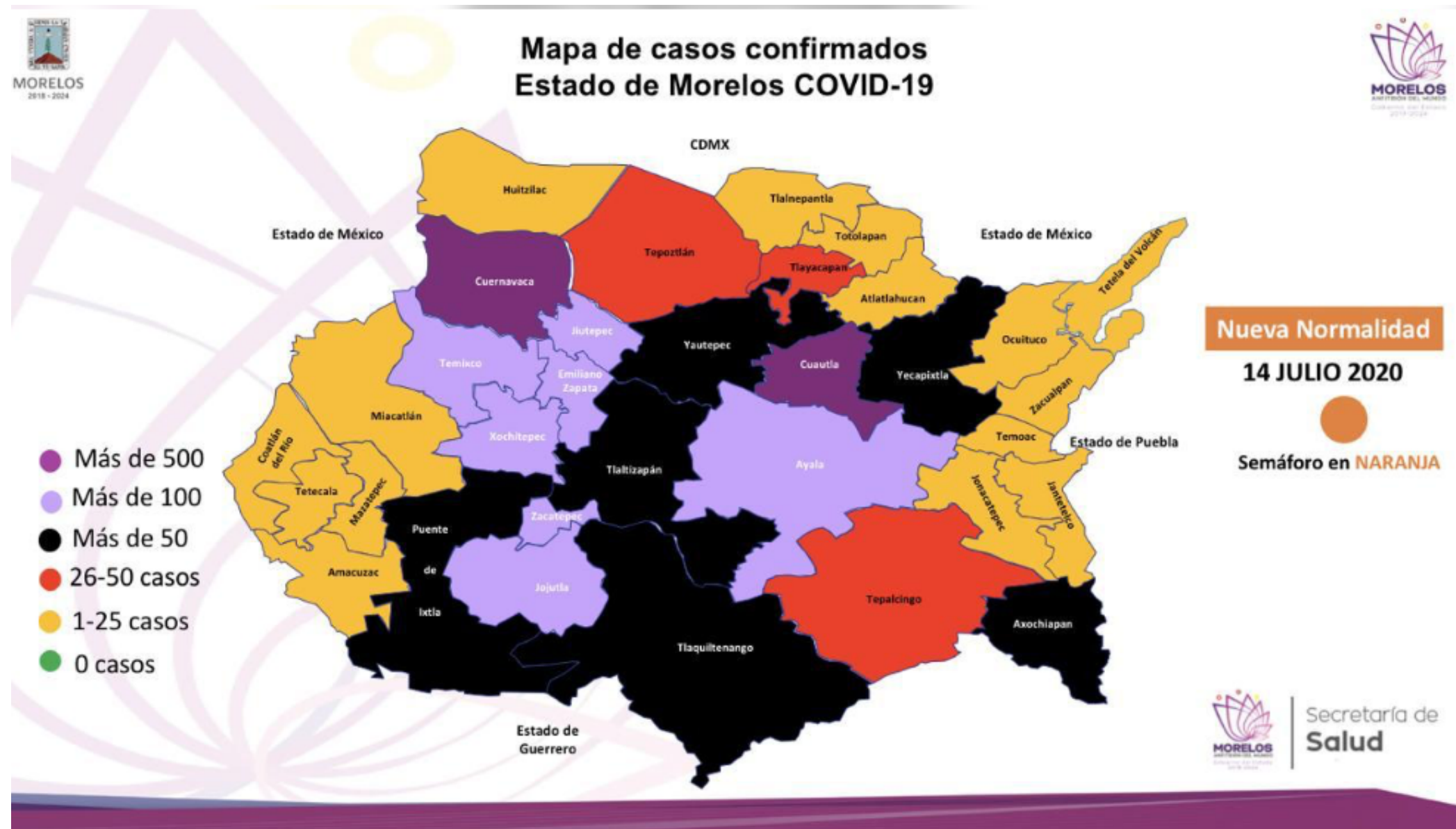
Square

Menos es más



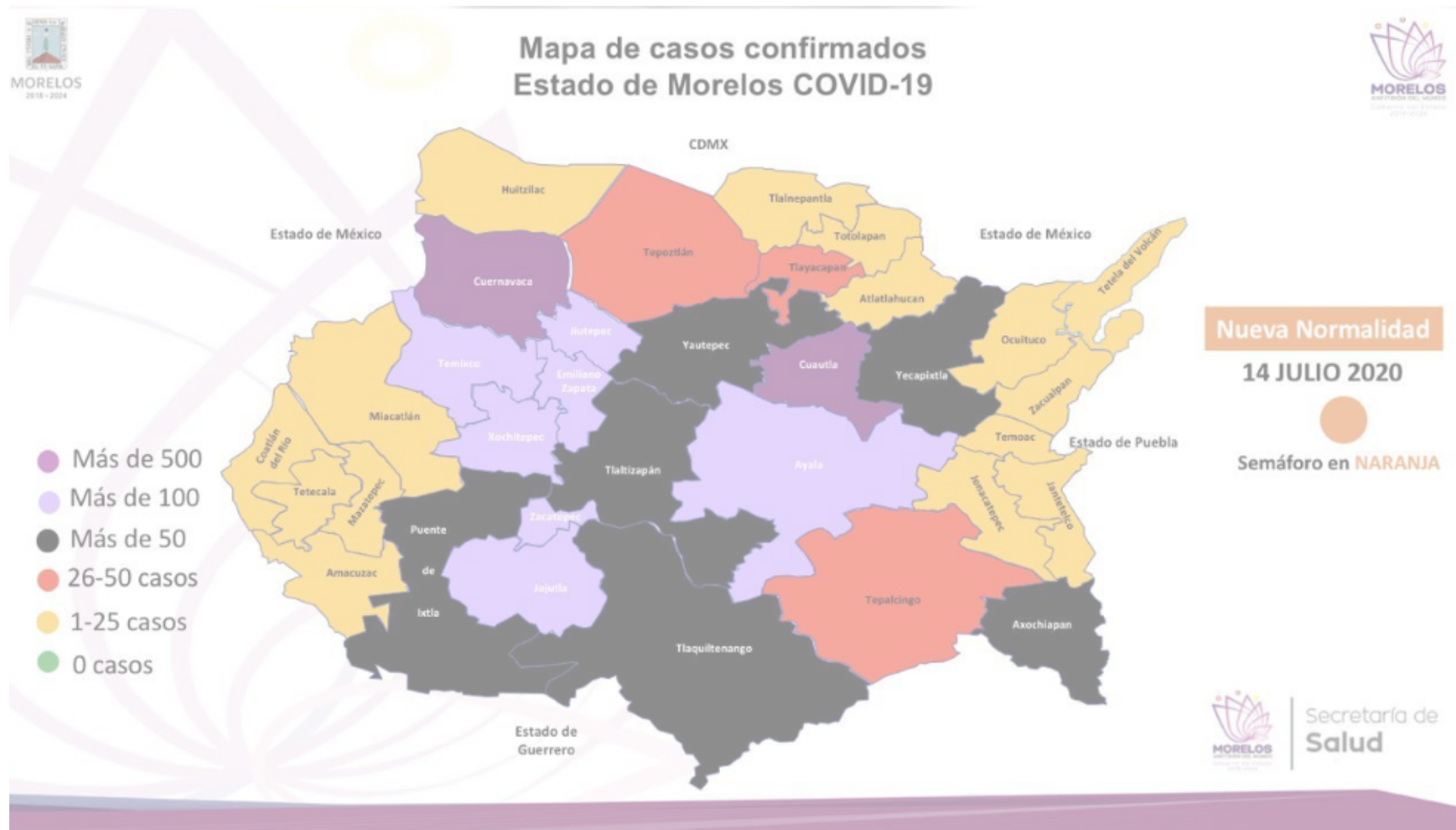
Uso de colores

¿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.

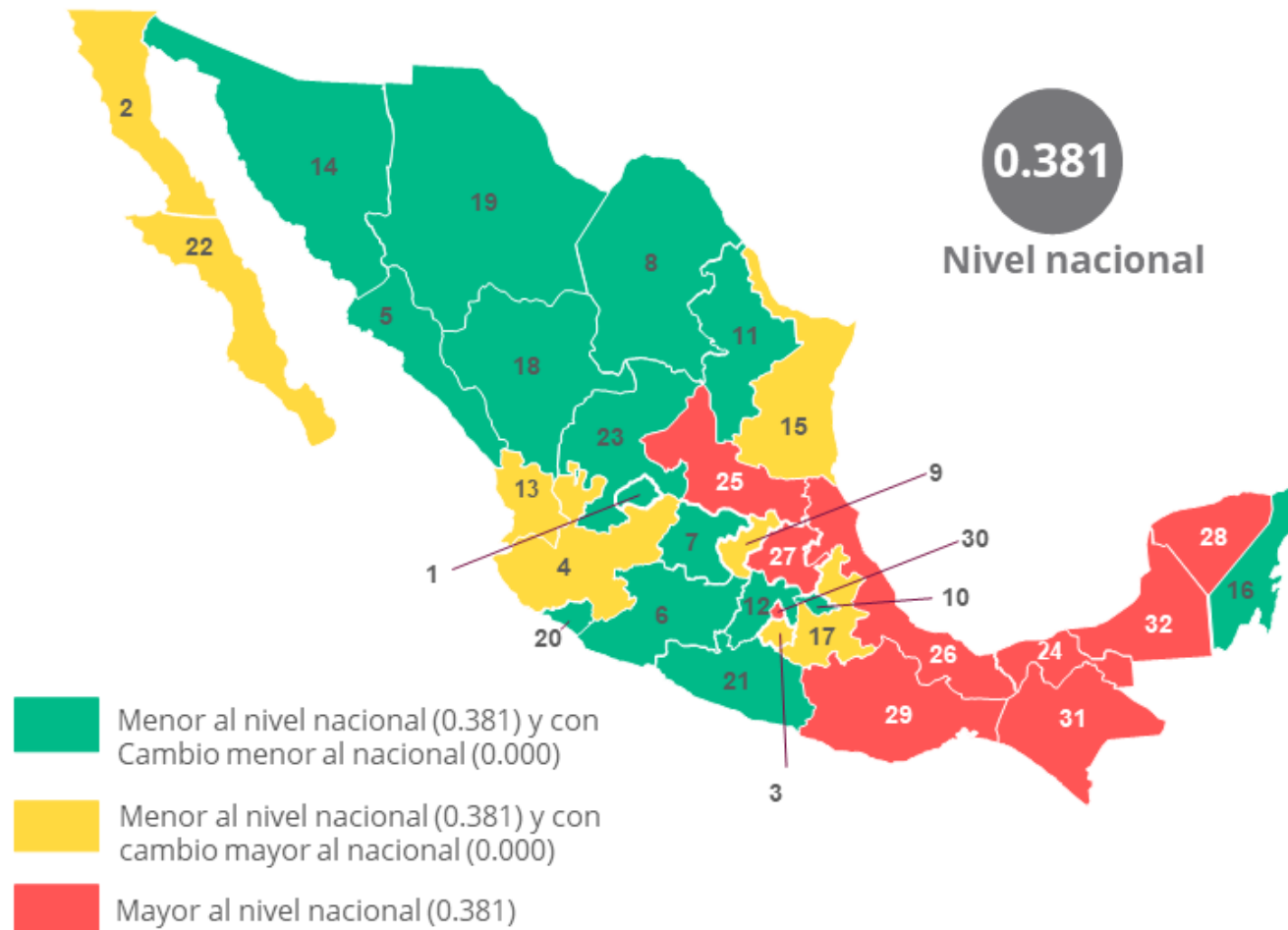
Uso de colores

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



Desigualdad laboral

1^{er} trimestre 2021



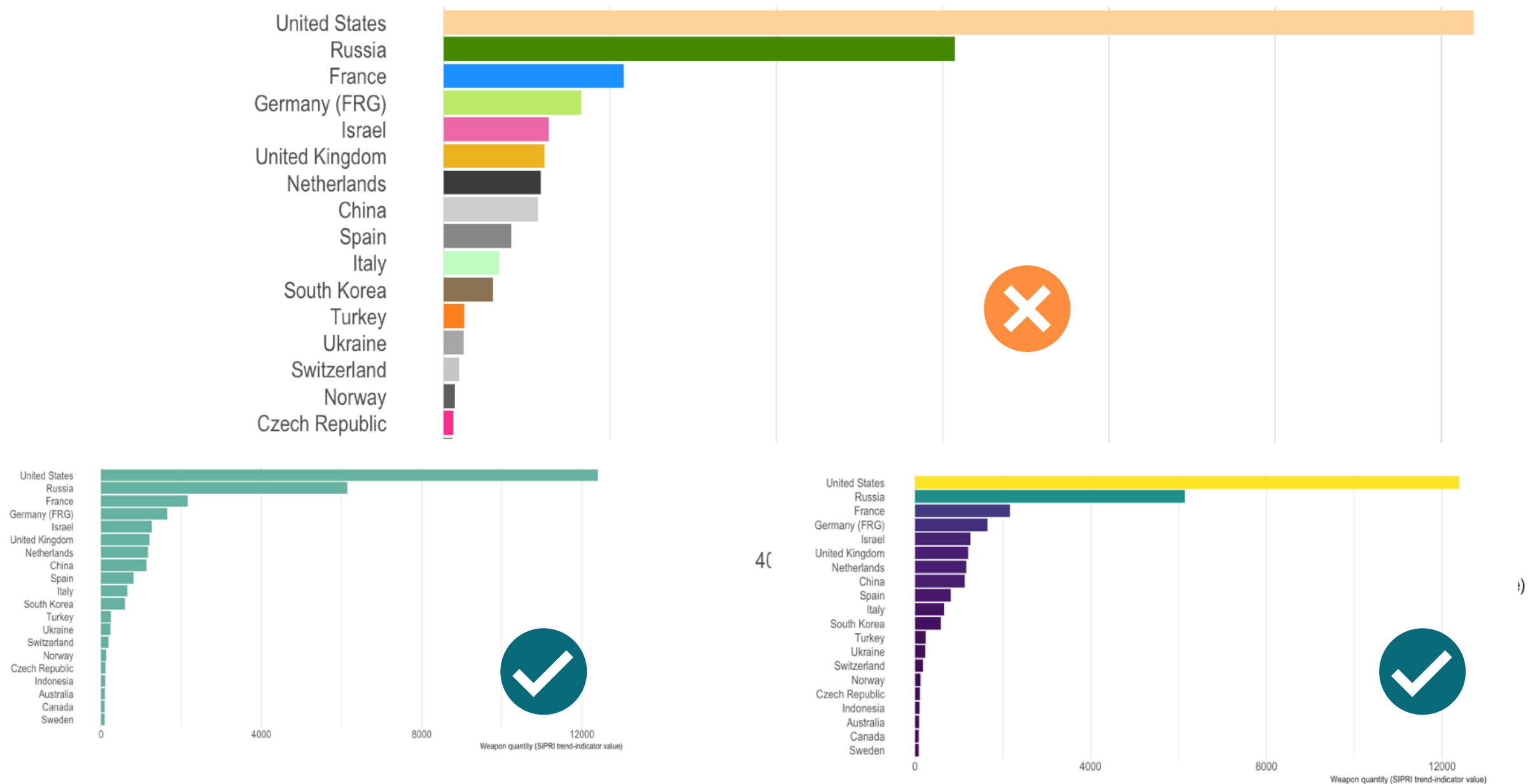
	Estado	Coefficiente de Gini	
1	Aguascalientes	0.299	
2	Baja California	0.301	
3	Morelos	0.311	
4	Jalisco	0.316	
5	Sinaloa	0.329	
6	Michoacán	0.334	
7	Guanajuato	0.337	
8	Coahuila	0.339	
9	Querétaro	0.349	
10	Tlaxcala	0.351	
11	Nuevo León	0.351	
12	Edo. Méx.	0.357	
13	Nayarit	0.357	
14	Sonora	0.358	
15	Tamaulipas	0.359	
16	Quintana Roo	0.360	

	Estado	Coefficiente de Gini	
17	Puebla	0.364	
18	Durango	0.364	
19	Chihuahua	0.365	
20	Colima	0.367	
21	Guerrero	0.369	
22	Baja California Sur	0.369	
23	Zacatecas	0.375	
24	Tabasco	0.405	
25	San Luis Potosí	0.405	
26	Veracruz	0.410	
27	Hidalgo	0.415	
28	Yucatán	0.419	
29	Oaxaca	0.420	
30	CDMX	0.439	
31	Chiapas	0.446	
32	Campeche	0.452	

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

Uso de colores

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



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

Uso de colores

Usa generadores de paletas de colores



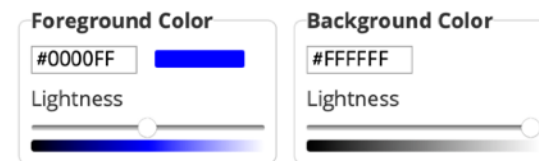
<https://colors.co/>



<https://mycolor.space>

Contrast Checker

[Home](#) > [Resources](#) > Contrast Checker



<https://webaim.org/resources/contrastchecker/>



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



Google Color Picker

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

Genera tus propias reglas de uso

Spot Palette



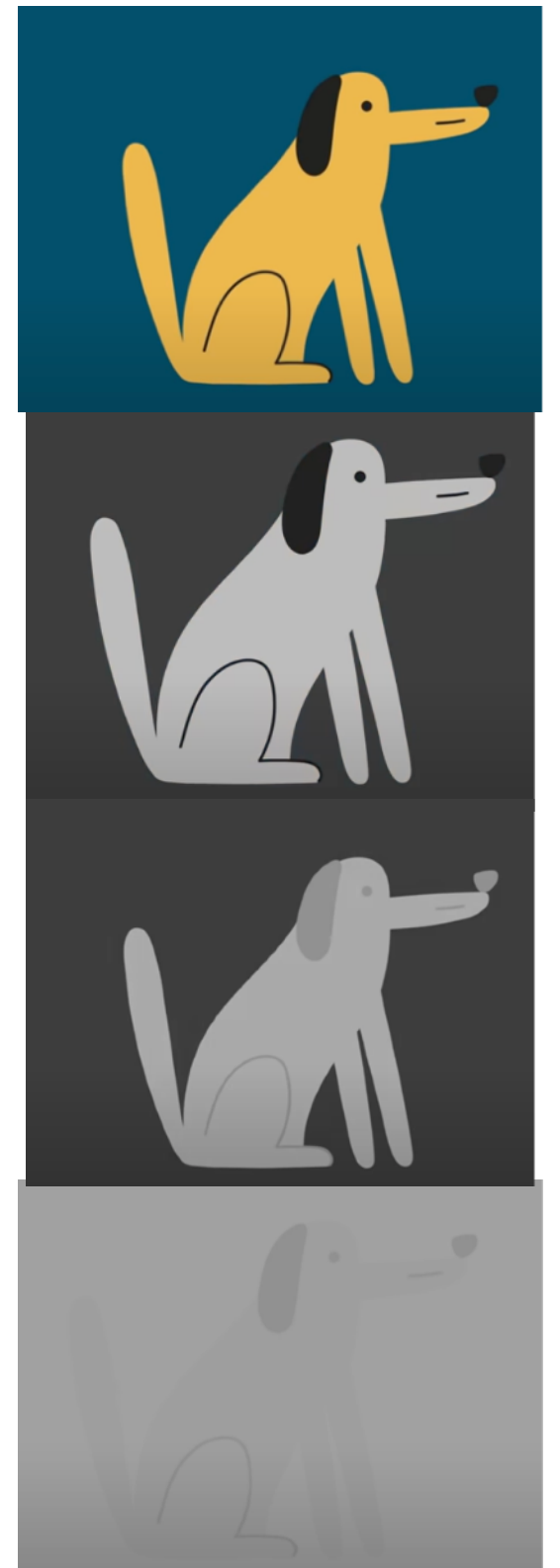
#046A78 #4E9EAC #BFFBFF #FE8E3C

Generic Gradient

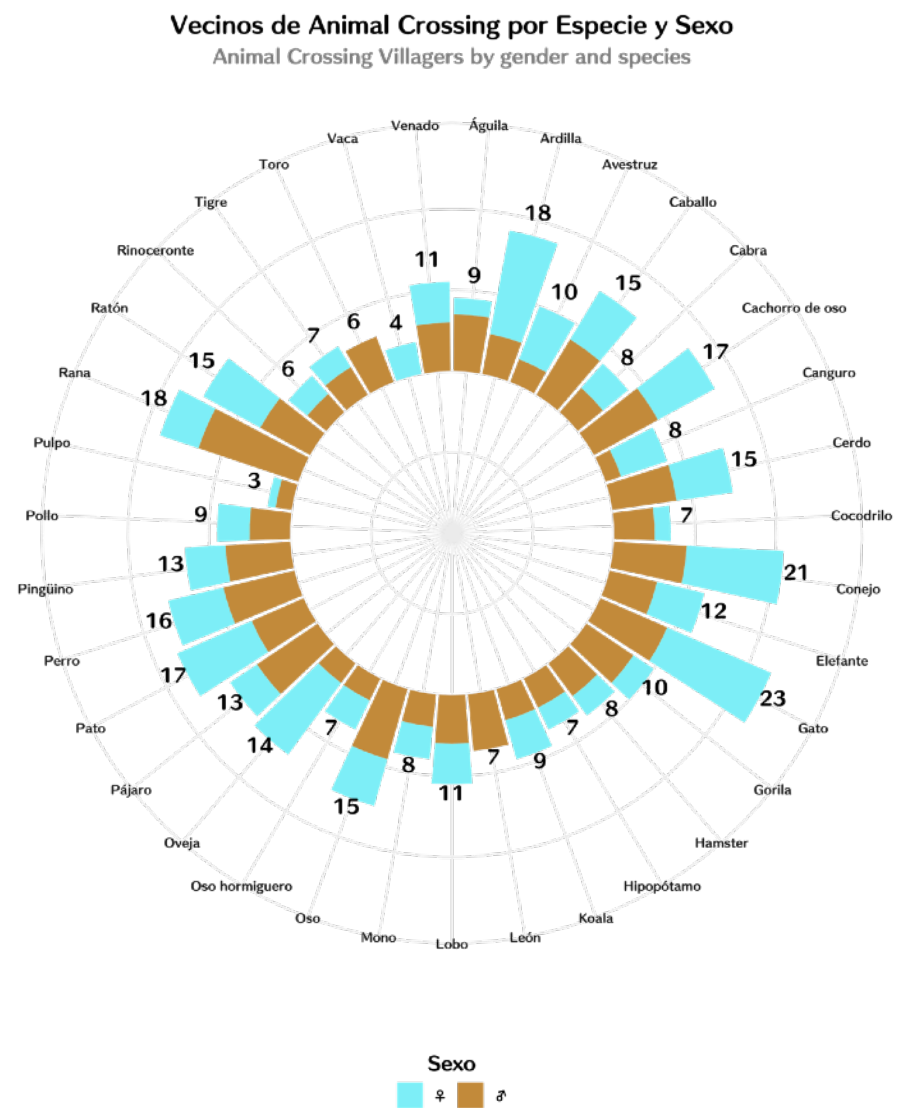


#046A78 #008A88 #2BAA8C #69C885 #ADE279 #F9F871

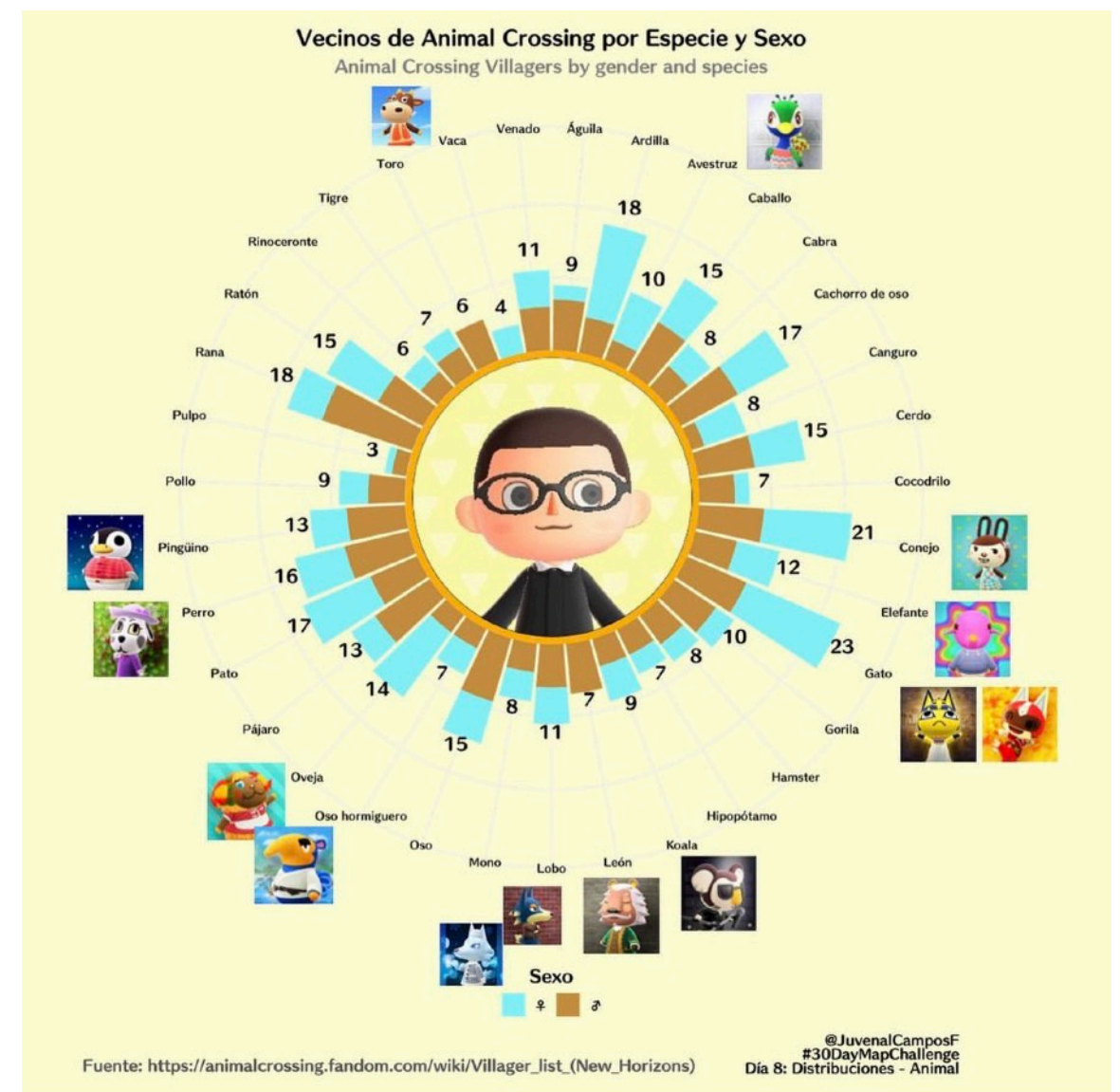
Uso de colores contrastantes



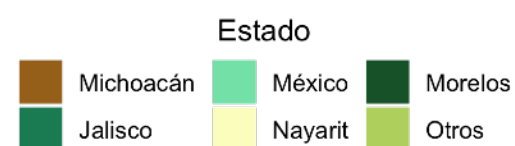
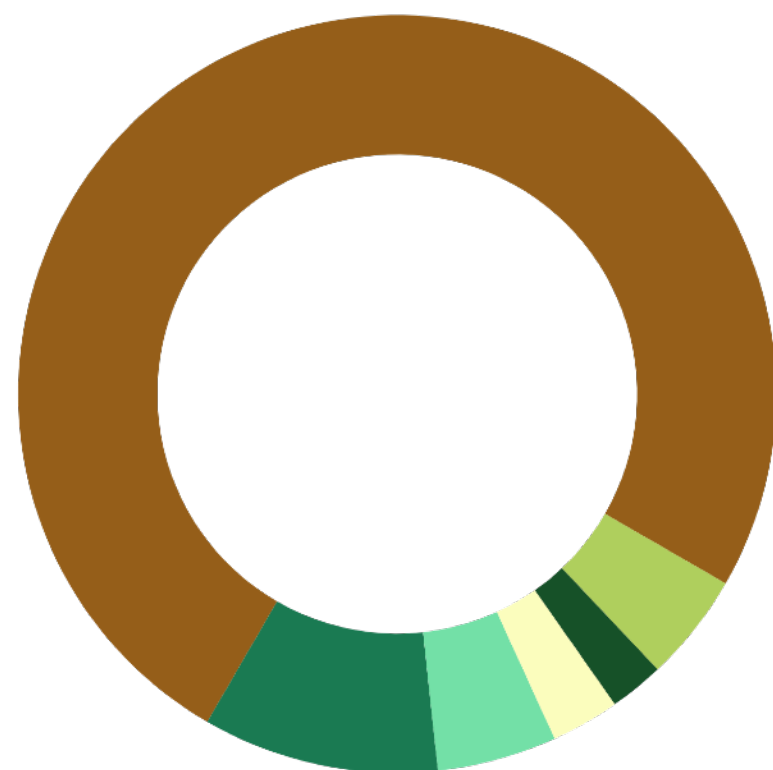
Uso de pistas contextuales



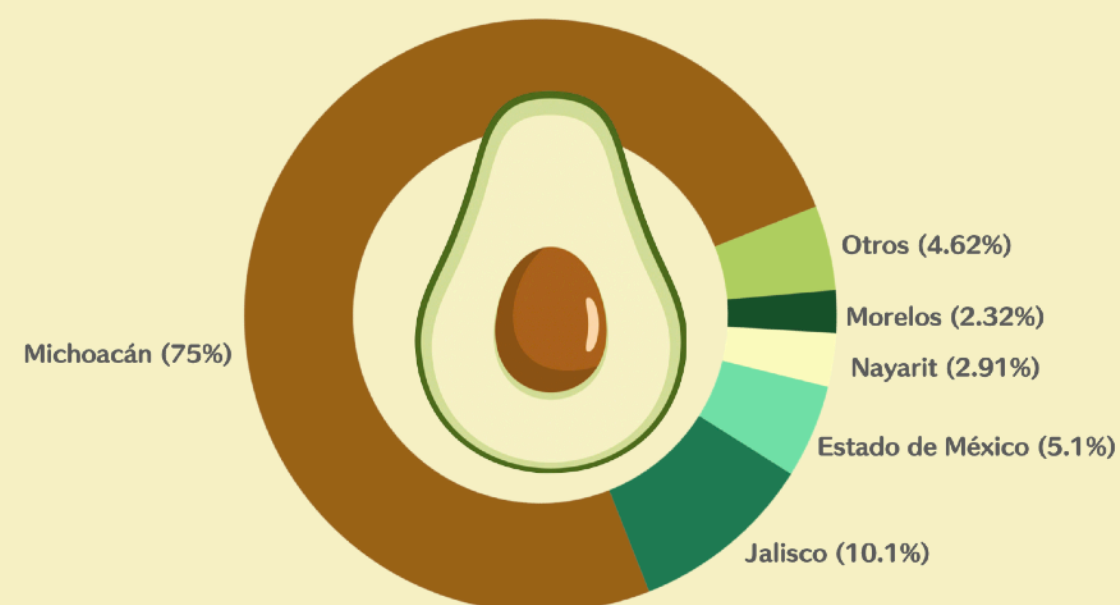
@JuvenalCamposF
#30DayMapChallenge
Día 8: Distribuciones - Animal



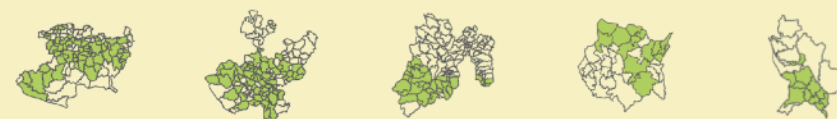
Uso de pistas contextuales



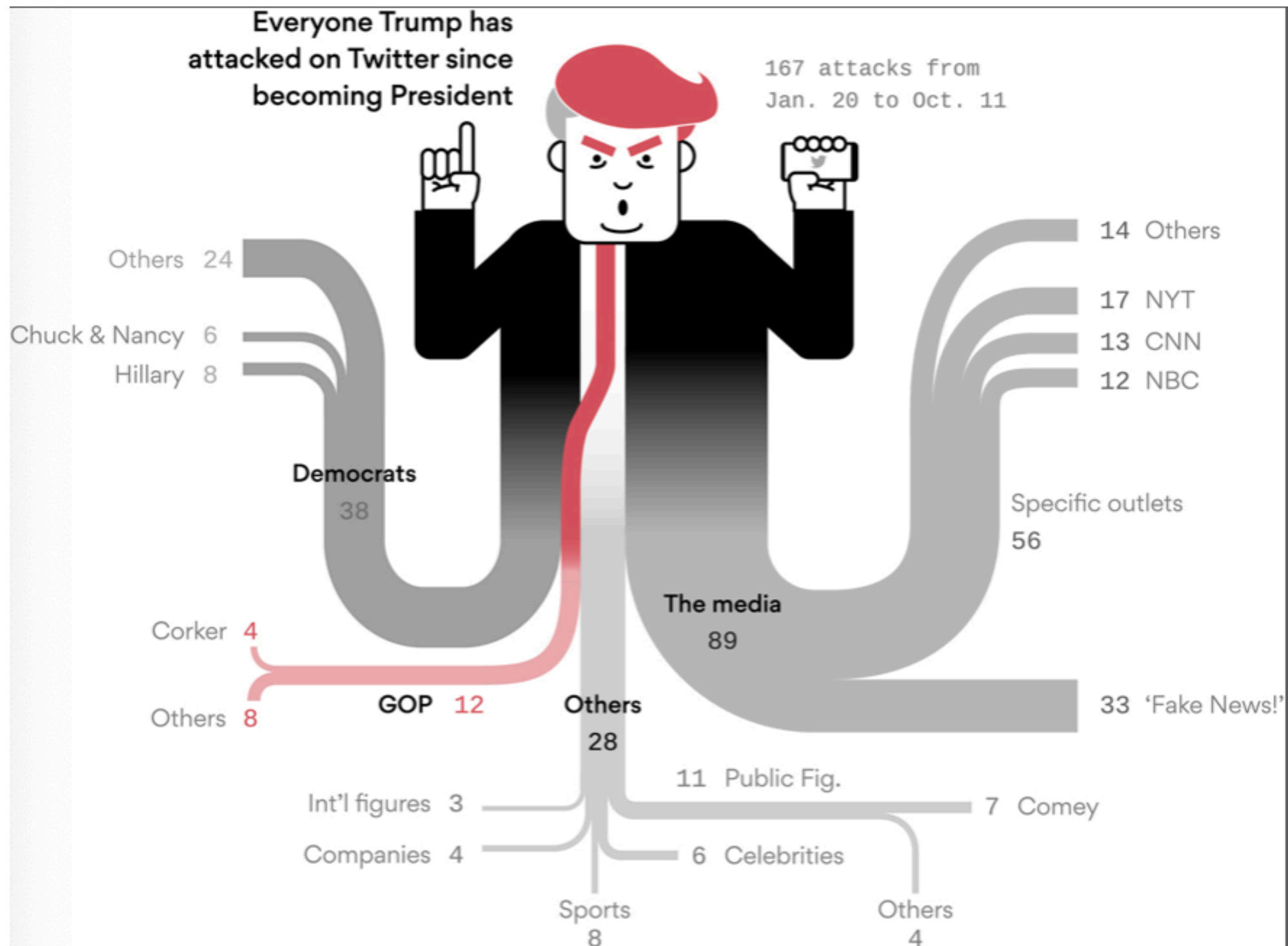
Aportación a la producción de Aguacate por Entidad México, 2019. (Datos más recientes).



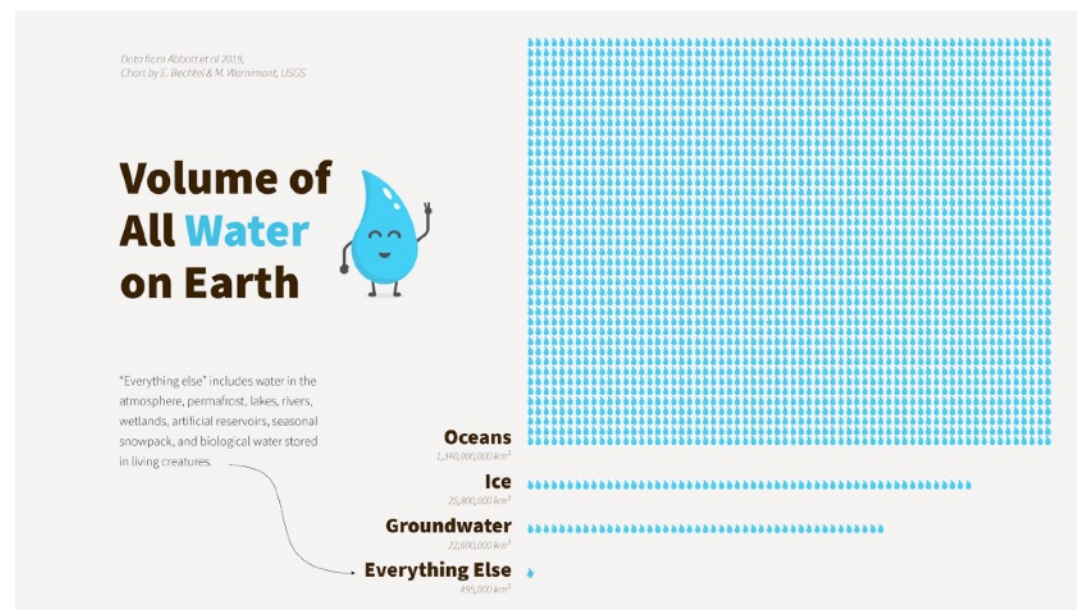
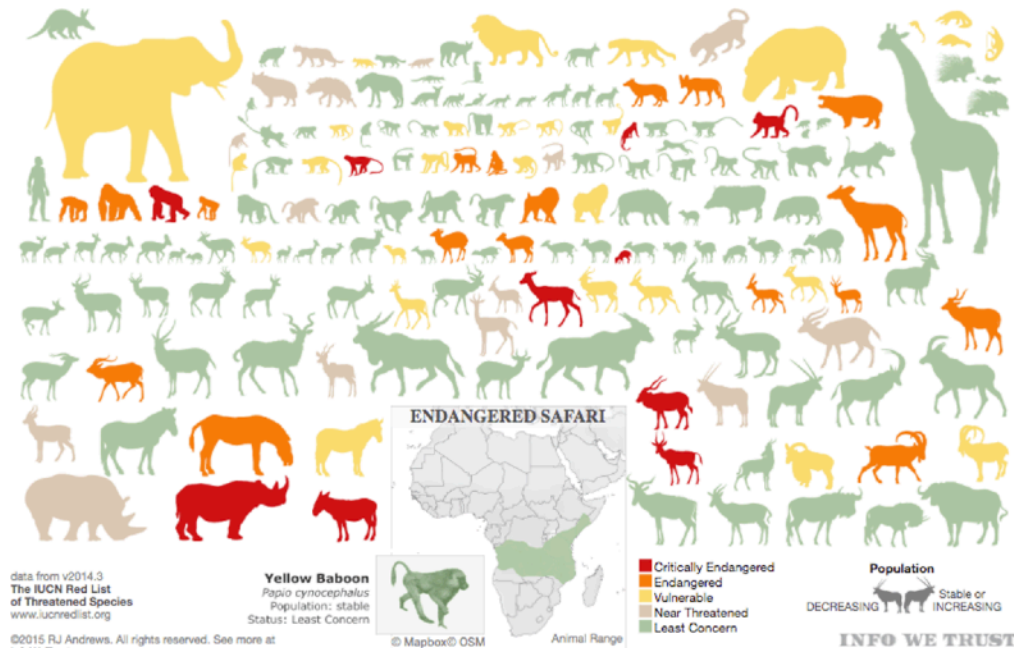
Regiones de producción de los 5 principales estados:



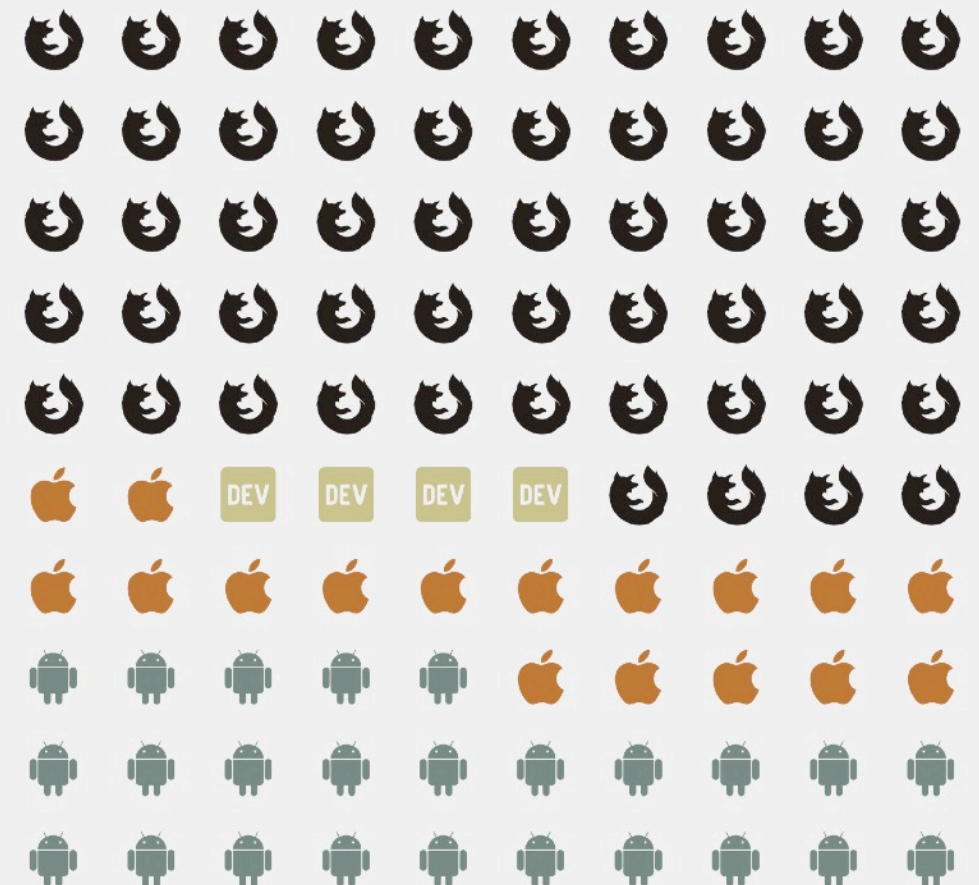
Uso de pistas contextuales



Uso de pistas contextuales



#30DayChartChallenge Twitter clients used by unique users



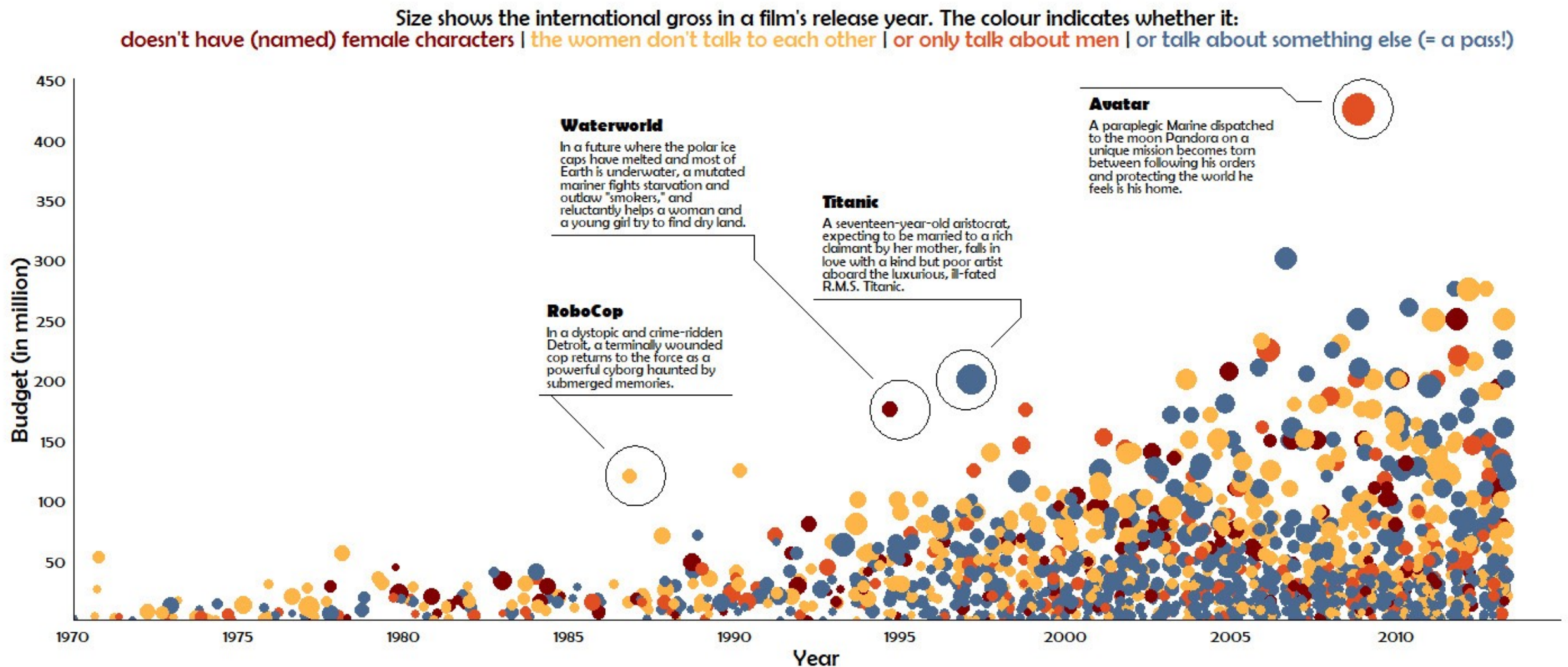
Twitter for Android
iOS/macOS apps
Bots and other
Twitter Web App

Source: Twitter · Graphic: Georgios Karamanis

Uso de texto de manera estratégica

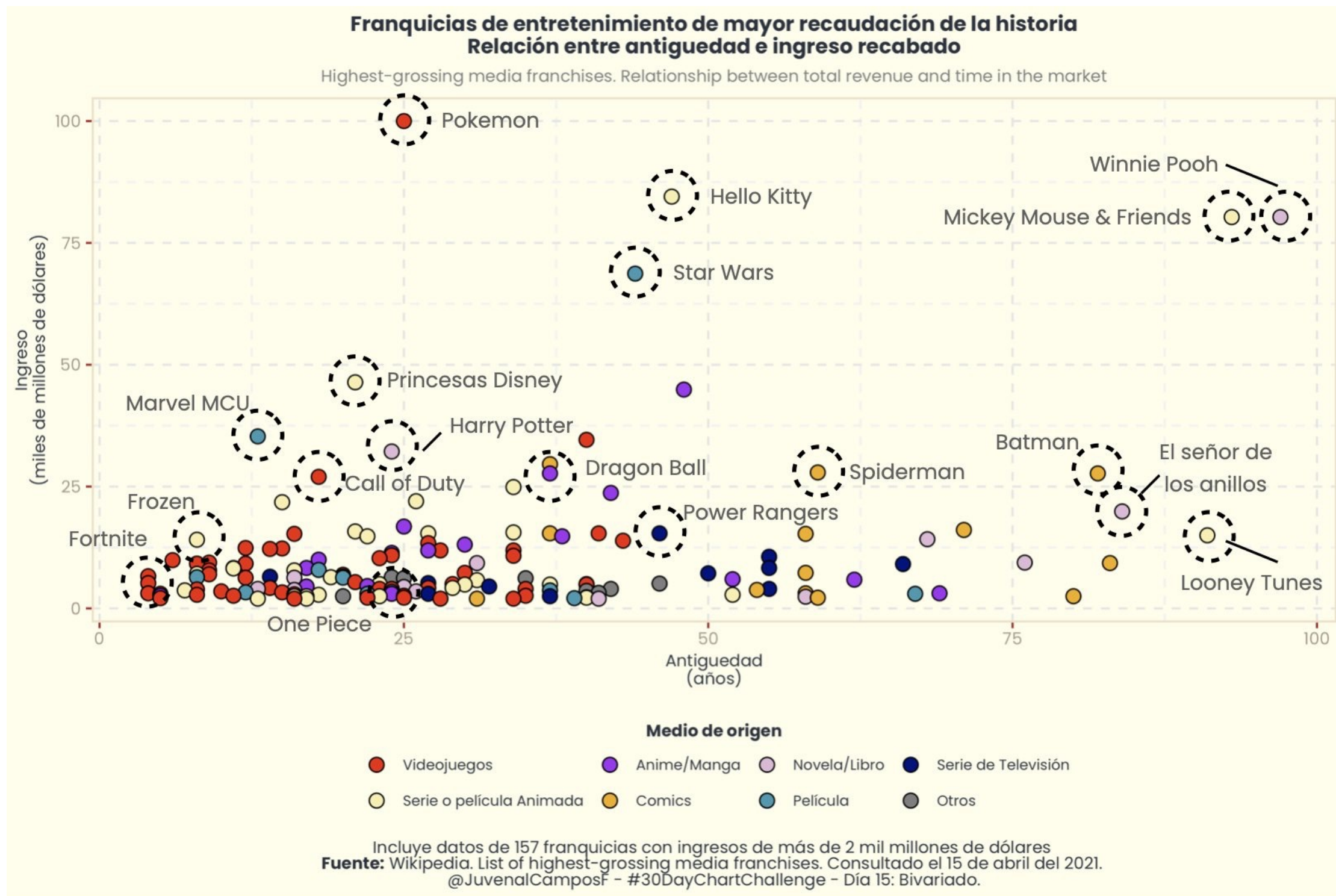
Usamos texto para indicar detalles importantes, sin **distraer**, sin poner **demasiado** y con el **contraste** adecuado.

Back to the Bechdel test!



Includes 1642 films from 1970 - 2013.
Data source: *Bechdeltest.com* API via Tidy Tuesday

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. Preocupate 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



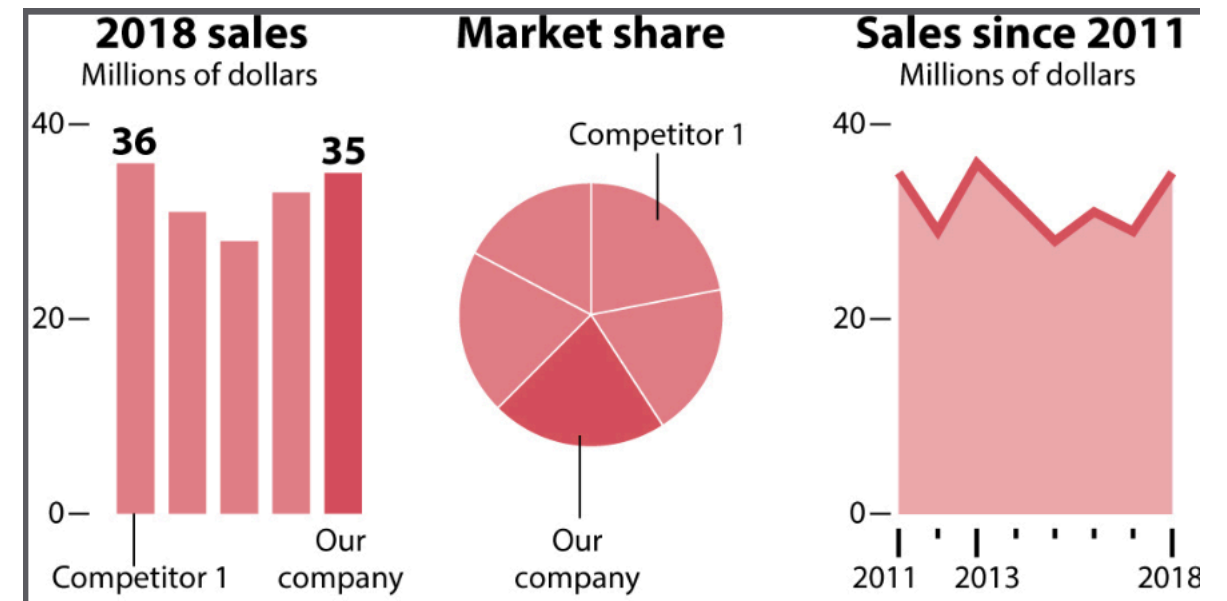
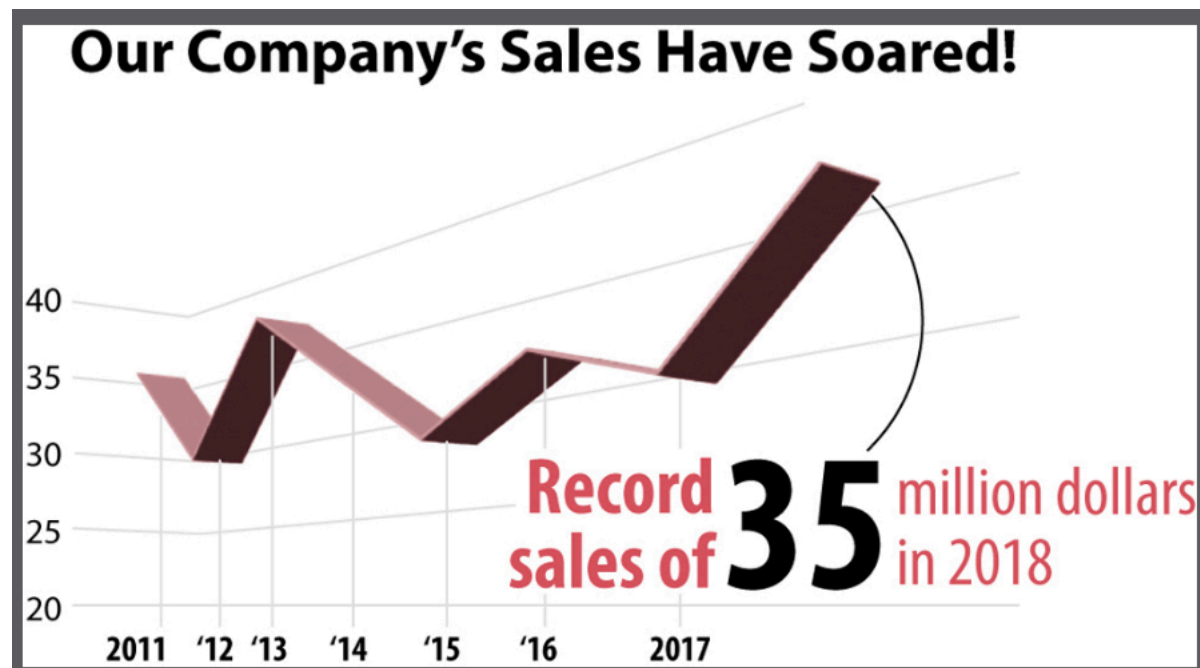
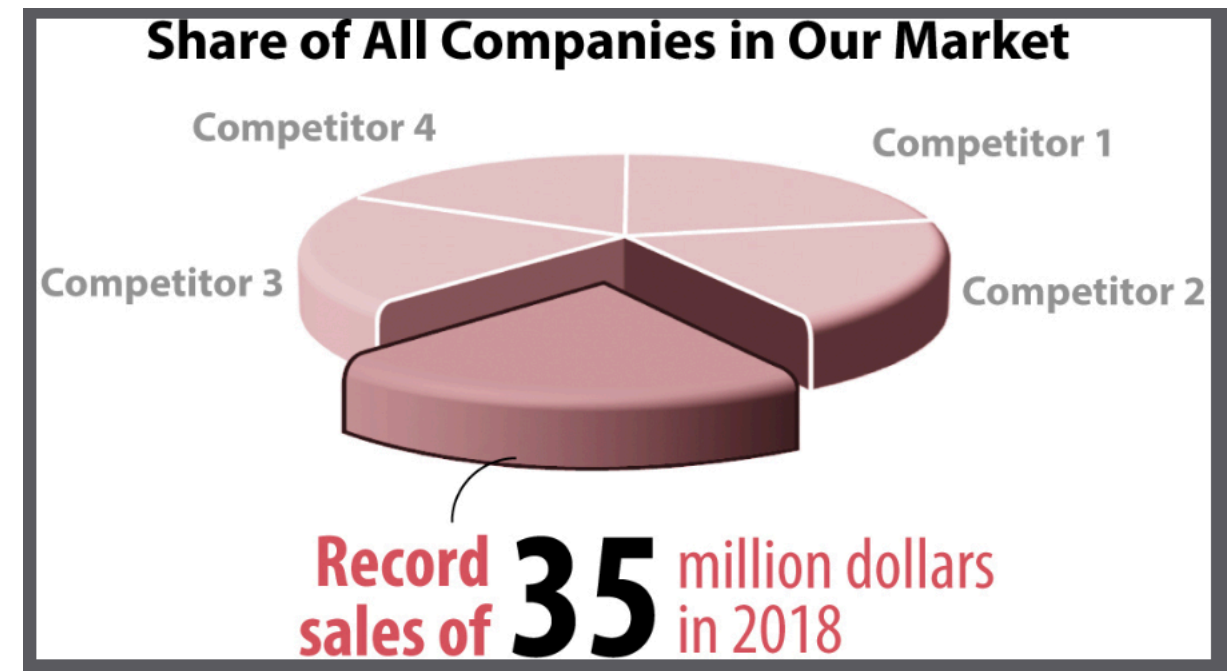
Mind your legend



Mind the aspect ratio

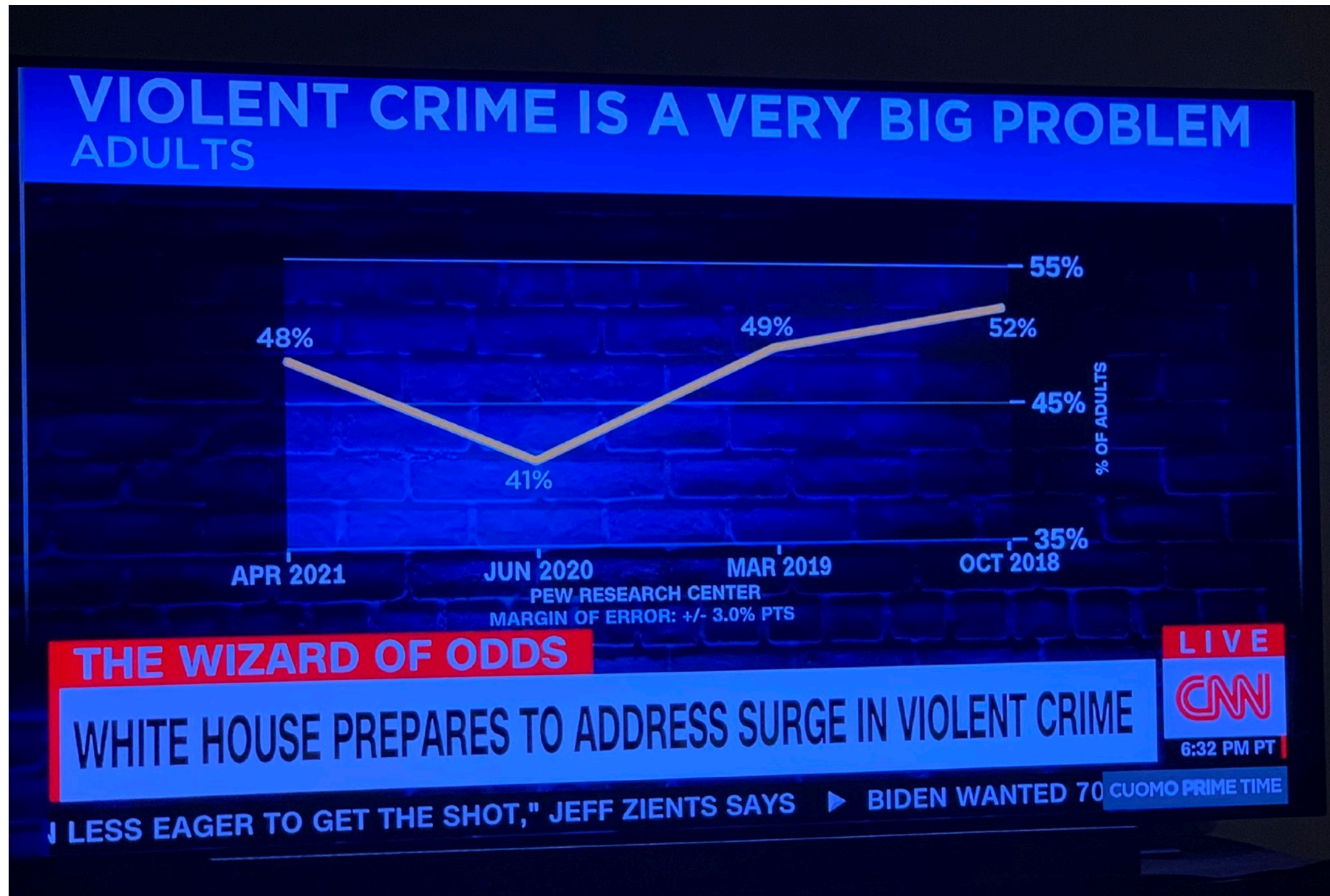
<https://www.data-to-viz.com/caveats.html>

El problema del 3D



Alberto Cairo. How Charts Lie. 2019.

Respetar 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