



Módulo 3 – Análisis Exploratorio y Programación Estadística

Visualización de Datos

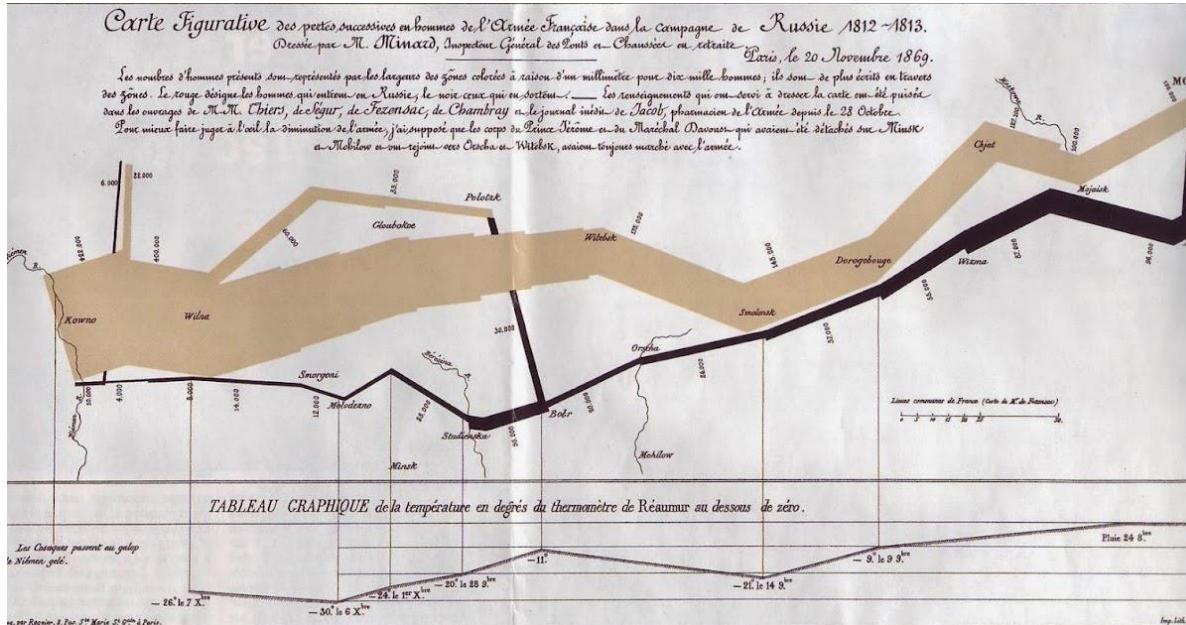
Ciencia de Datos

Visualización de Datos



Visualización de Datos.

Entre los que dibujan datos en periódicos, nadie discute la condición fundadora de la infografía del magnífico trabajo de Charles Minard sobre la campaña de Napoleón a Rusia, entre los años 1812 y 1813. En esta infografía, publicado en un diario de la época, se ilustra cómo 442 mil hombres ingresan a Rusia (color ocre) y sólo 10.000 regresan. Esta fue la peor campaña de Napoleón.



Nótese cómo en esta infografía se mezclan elementos visuales para narrar una historia centrada en base a los datos.

Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

Dessiné par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en lettres dans les zones. Le rouge désigne les hommes qui ont été en Russie, le noir ceux qui en sortent. Les renseignements qui ont servi à dresser la carte ont été pris dans les ouvrages de M. M. Chiers, de Séguir, de Fezensac, de Chambray et le journal médical de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout, qui avaient été détachés sur Minsk et Mogilow et qui rejoignirent Orléans et Witebsk, avaient toujours marché avec l'armée.

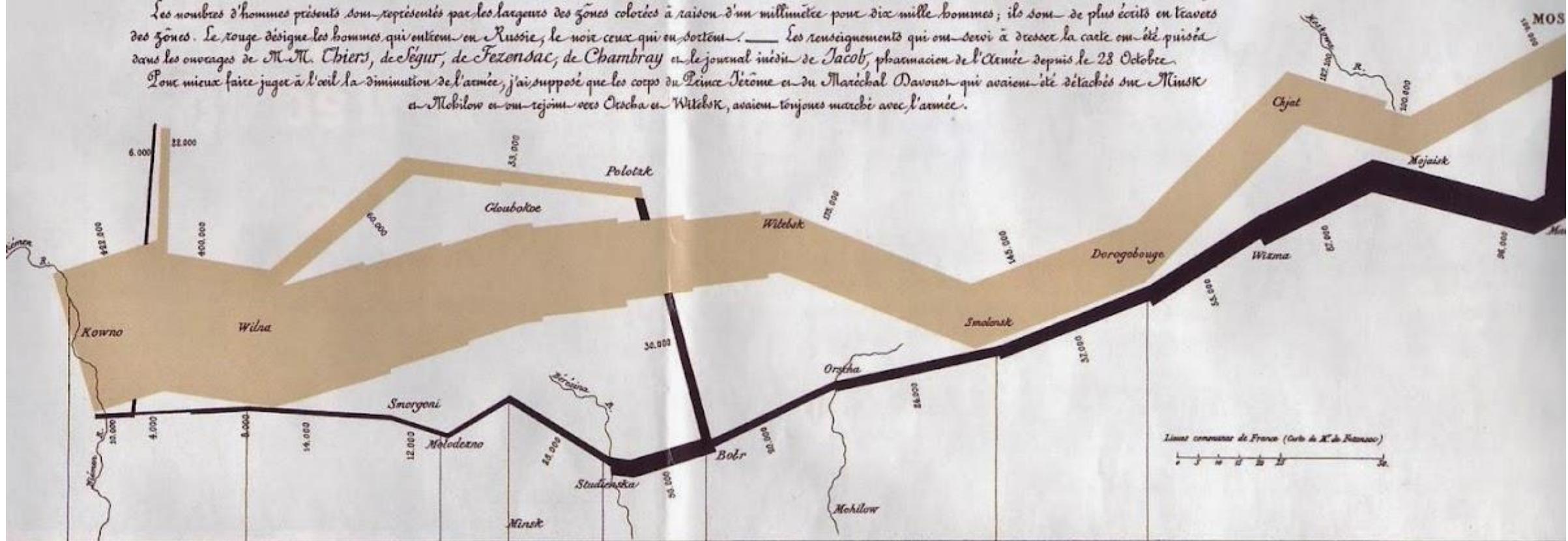
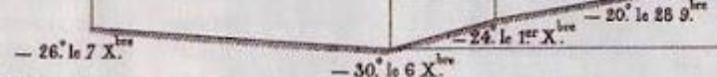


TABLEAU CRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop
le Nilmen gelé.



Visualización de Datos

- La visualización de datos es el proceso de representar datos e información de una manera visual, gráfica o diagramática, con el objetivo de facilitar la comprensión, el análisis y la comunicación de patrones, tendencias, relaciones y otros insights relevantes que puedan extraerse de ellos.
- La visualización de datos puede ser utilizada en una variedad de contextos, desde la investigación científica hasta los negocios, la educación y el periodismo, y puede ser realizada mediante diferentes herramientas y técnicas, como gráficos, mapas, diagramas de flujo, infografías, entre otros.
- La visualización de datos se considera una herramienta poderosa en la toma de decisiones, ya que puede ayudar a identificar patrones y tendencias que de otra manera serían difíciles de detectar, y a comunicar de manera más efectiva información compleja a una audiencia no especializada.

Visualización de Datos



"El mayor valor de una imagen sucede cuando nos fuerza notar algo que nunca esperamos ver"

-- *John Tukey (matemático y estadístico)*

Puntos ciegos de la visualización de datos

La visualización y la cognición del ser humano están relacionadas. Las percepciones visuales inciden sobre los procesos cognitivos, y, por ende, son fundamentales en la comunicación. Sobre todo, cuando se busca comunicar datos.

Sin embargo, tal como muchas veces se favorece el entendimiento, hay oportunidades en la cual sucede el fenómeno opuesto. Es decir, se transmite una idea que no se condice con los datos presentados induciendo a errores en los receptores.

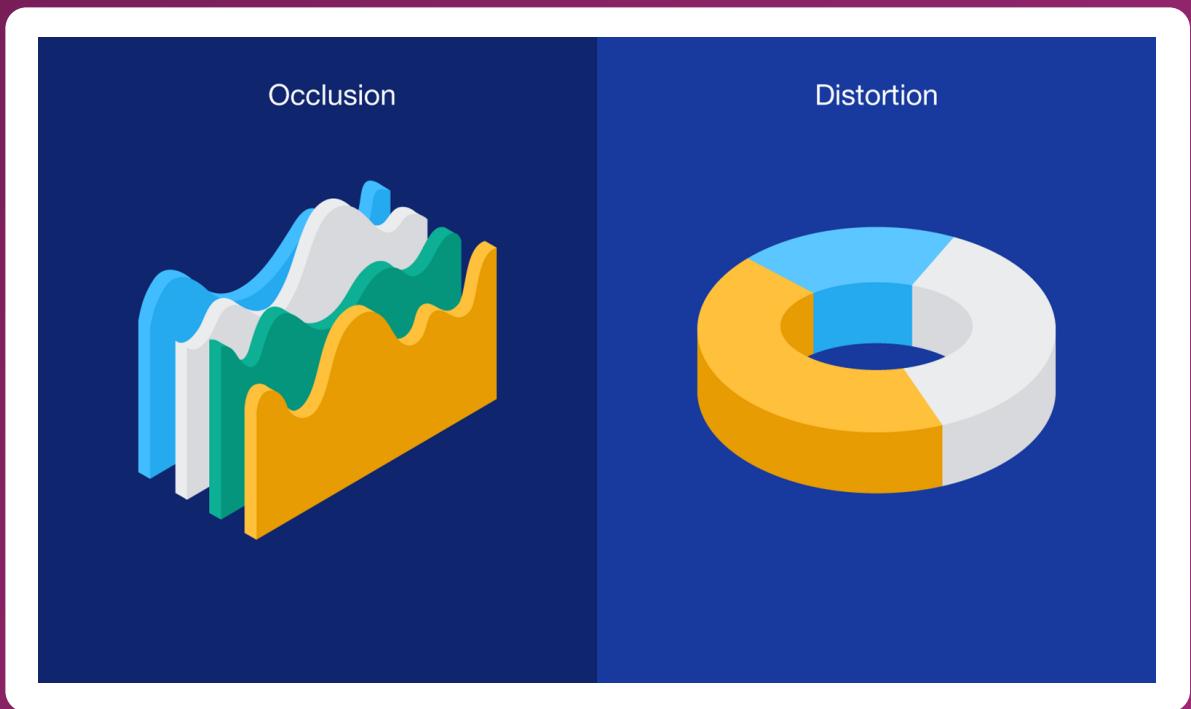
Para ser justo, hay que decir que muchas veces esto sucede de forma involuntaria, pero también habrá oportunidades en que efectivamente se hizo una manipulación adrede.

A continuación, se presenta algunos aciertos y desaciertos del análisis visual.

Puntos ciegos de la visualización de datos

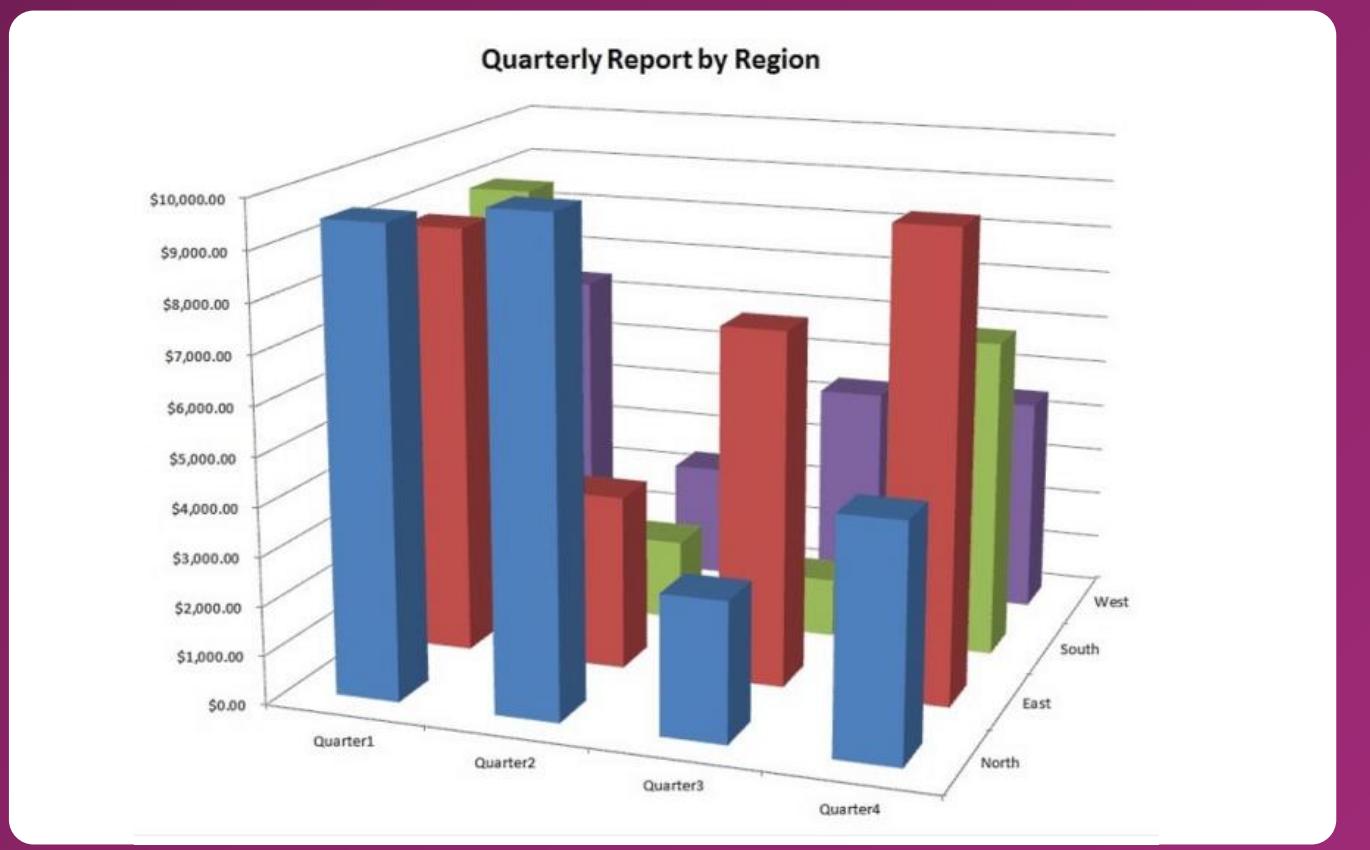
Los gráficos con perspectiva 3D muchas veces pueden producir un efecto distorsionador de las proporciones o bien puede ocultar información relevante.

Nótese cómo en este caso, en el diagrama de dona el segmento de color azul se ve disminuido en comparación al amarillo y blanco.



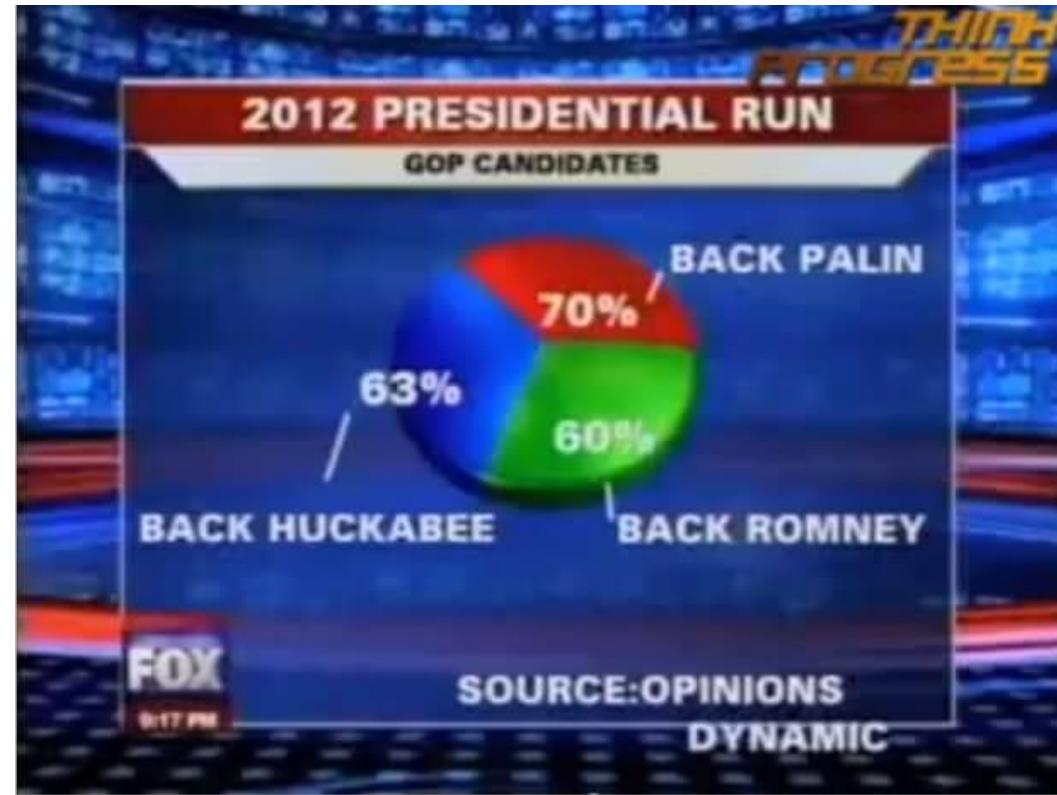
Este es un claro ejemplo de cómo elegir una visualización poco apropiada para una buena comunicación con la audiencia.

En este caso, las perspectivas 3D distorsionan la realidad y hacen más difícil la interpretación.



Puntos ciegos de la visualización de datos

Este es un caso que no necesita mayores comentarios.

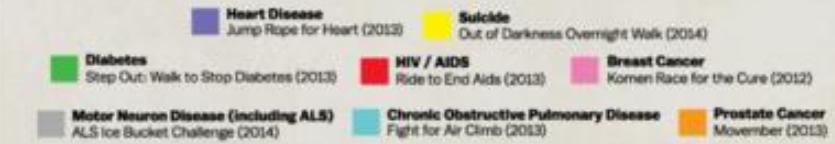


Puntos ciegos de la visualización de datos

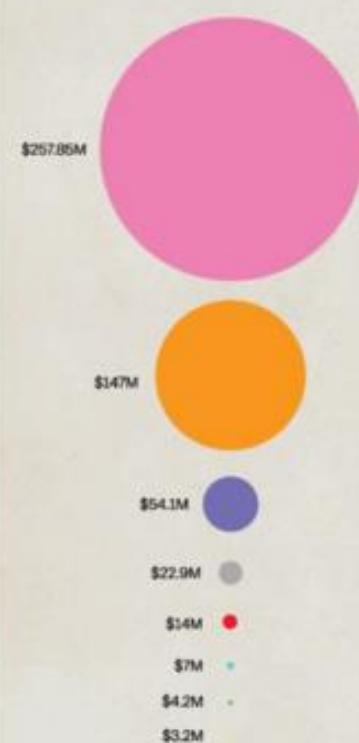
En el siguiente ejemplo, se quiso mostrar cómo ha sido la evolución de la cantidad de personas muertas a causa de una enfermedad mortal, y por otro lado, cómo en el tiempo ha ido aumentando el levantamiento de fondos para combatir dicha enfermedad.

Tal vez un gráfico de barras en el tiempo o un diagrama de líneas hubiera sido más conveniente puesto que, en este caso, se transmite la idea que mientras más fondos se levantan, más muertes se producen. Cosa que posiblemente no era lo que se quería transmitir.

WHERE WE DONATE VS. DISEASES THAT KILL US



MONEY RAISED



DEATHS (US)



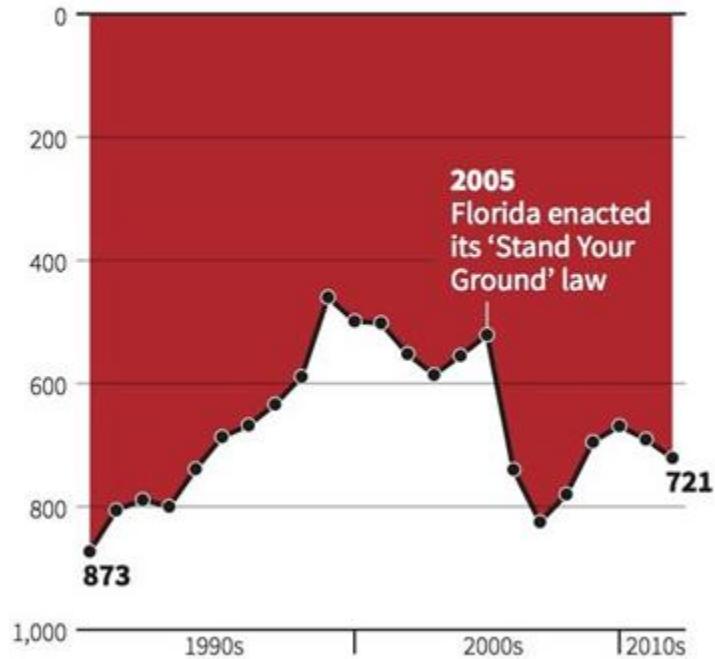
Source: CDC (2011)



Puntos ciegos de la visualización de datos

Gun deaths in Florida

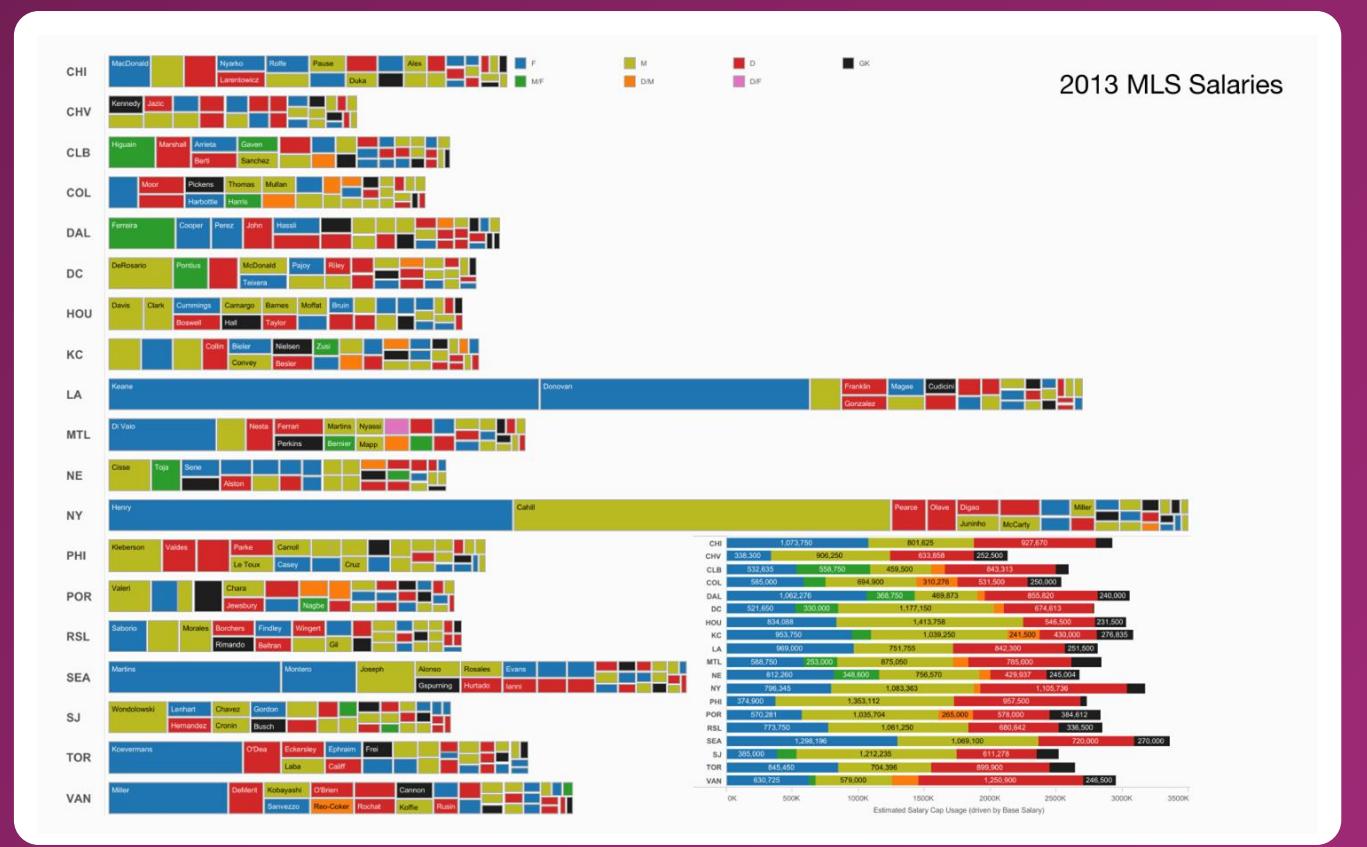
Number of murders committed using firearms



Aunque no lo parezca, la elección de colores también es importante en una visualización gráfica.

Puntos ciegos de la visualización de datos

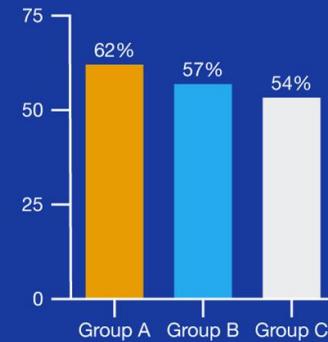
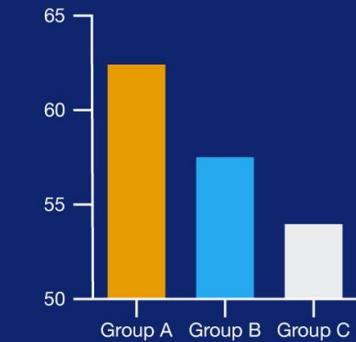
Cuidado con el exceso de datos en las visualizaciones. Muchas veces un exceso de información de visualización hace que se dificulte la transmisión de las ideas principales.



Puntos ciegos de la visualización de datos

El manejo de la escala es algo que también se debe determinar con cuidado pues produce una distorsión en la percepción.

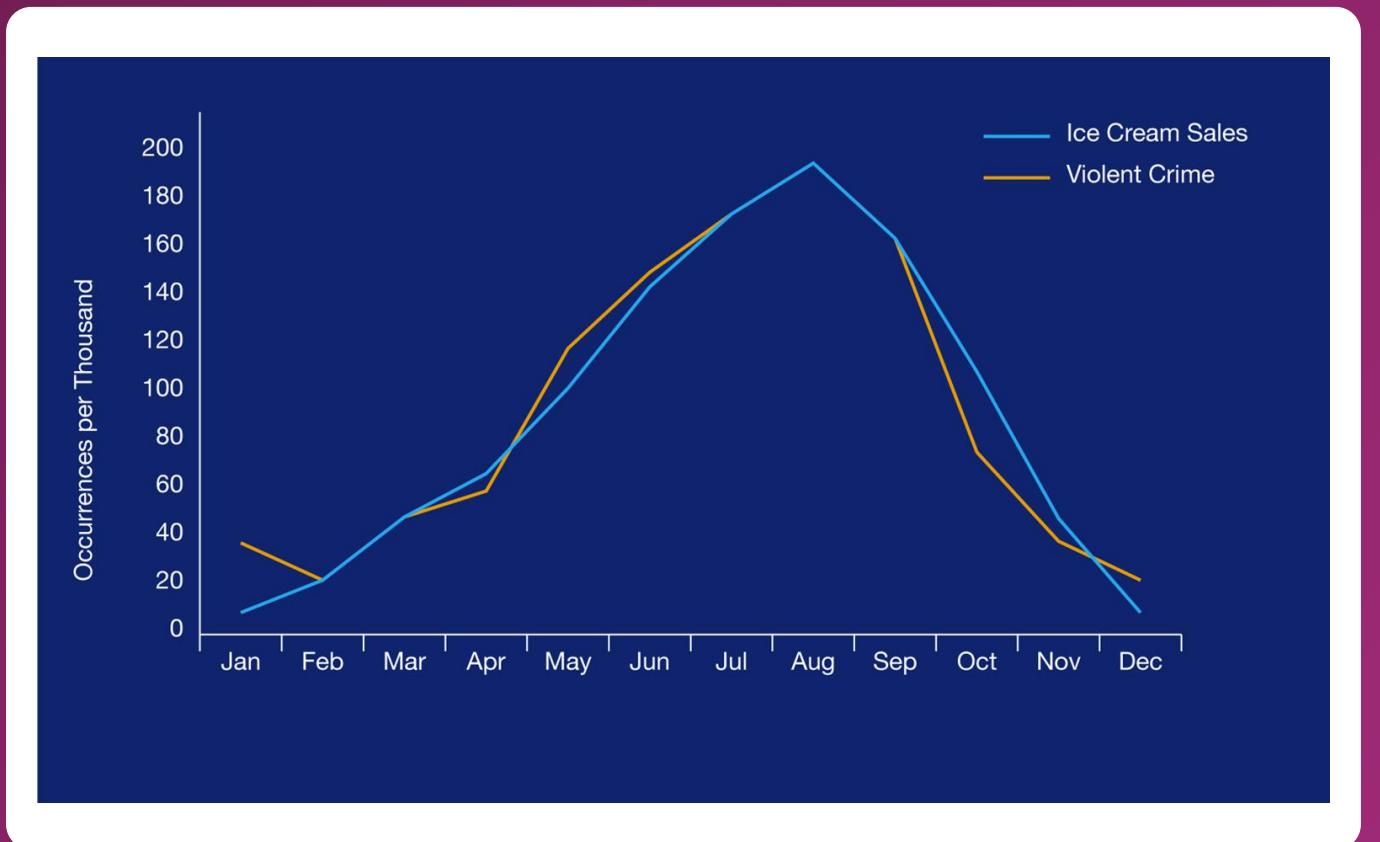
Nótese cómo en el primer gráfico se da a entender que el grupo A es muy superior a los demás, cosa que no se observa en el gráfico de la derecha al ampliar la escala.



Puntos ciegos de la visualización de datos

A veces es posible que se dé a entender en una audiencia que existe causalidad entre dos variables por el mero hecho que los gráficos muestran una correlación.

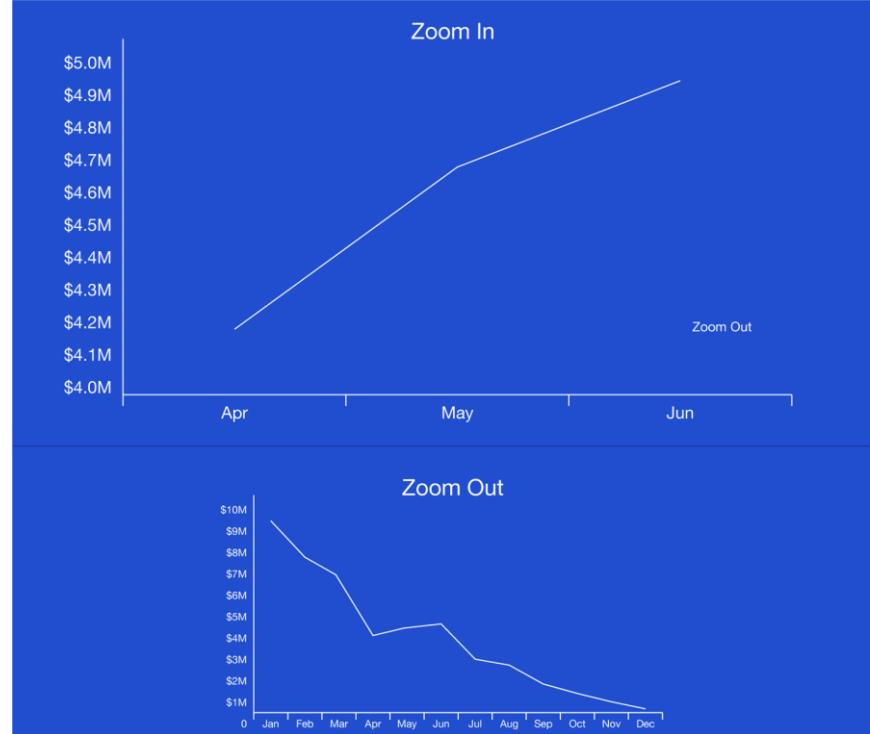
Sin embargo, ya sabemos que correlación no implica causalidad y ya sabemos acerca de las correlaciones espurias.



Puntos ciegos de la visualización de datos

Utilizar el zoom en un gráfico de forma inadecuada puede llevar a mostrar un panorama totalmente distinto a la realidad.

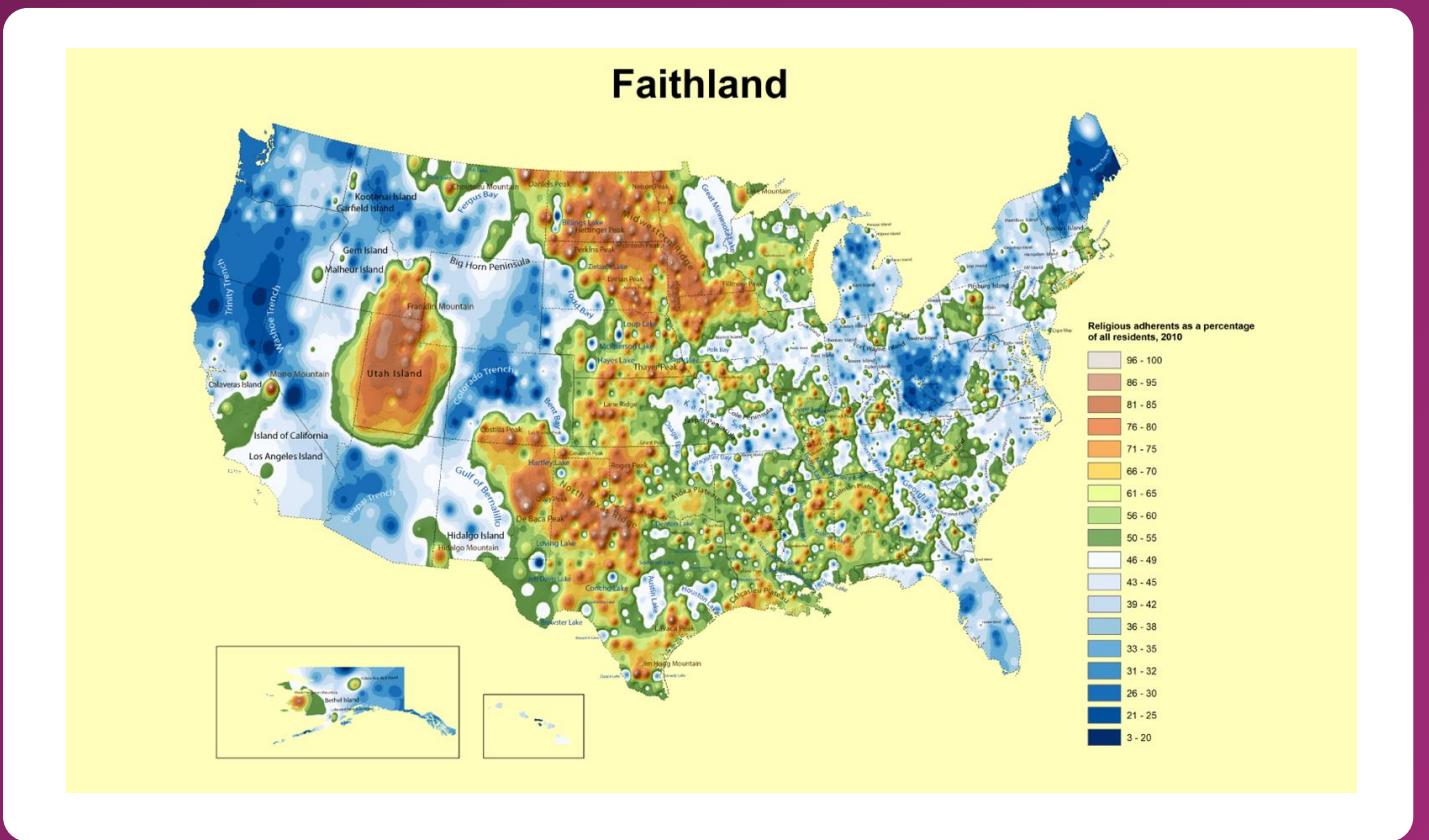
Aunque altamente cuestionado, este es un recurso frecuente.



Puntos ciegos de la visualización de datos

Por último, la elección correcta de colores también es importante al momento de elaborar una visualización.

Nótese cómo los colores azules y verdes del mapa de calor distorsionan la perspectiva geográfica del mapa confundiéndose con océano y tierra.



Librerías de Visualización

Python cuenta con varias librerías para visualización. Las más utilizadas son las siguientes:

- » **Matplotlib**: para gráficas sencillas con bars, pies, lines, scatter plots, etc.
- » **Seaborn**: para visualización estadística, para crear mapas de calor o de alguna manera resumiendo los datos y aún desea mostrar la distribución de los datos.
- » **Plotly y Bokeh**: para visualización interactiva. Si los datos son tan complejos (o no puede ver la información de sus datos), se recomienda utilizar Plotly y Bokeh para crear visualizaciones interactivas que permitan a los usuarios explorar los datos mismos.

Librería Matplotlib



Dudas y consultas



KIBERNUM



Fin presentación



KIBERNUM