

Centre Émile Durkheim

Atelier Méthodes

Visualisation de données

SOMMAIRE

01. INTRODUCTION

Présentations
Pourquoi la dataviz ?

02. CE QU'ON PEUT FAIRE

La dataviz c'est large!
Les compromis

03. CE QU'ON DOIT FAIRE

Lecteurs
Format final
Objectif

04. CE QU'IL FAUT ÉVITER

Data-ink ratio
Légendes
Échelles
3D
Bonus : dans les médias

05. COMMENT LE FAIRE

Outils
Pros & Cons
Cas d'usage 1
Cas d'usage 2

06. LA SUITE?

01

INTRODUCTION

1 - INTRODUCTION

Joseph Barbier



barbierjoseph.com

github.com/JosephBARBIERDARNAL

Data science

- Chercheurs
- Entreprises
- Service public

Open source

- Dataviz
- Machine Learning
- Stats

Matplotlib Journey



r-graph-gallery.com
python-graph-gallery.com
react-graph-gallery.com
d3-graph-gallery.com
data-to-viz.com
dataviz-inspiration.com

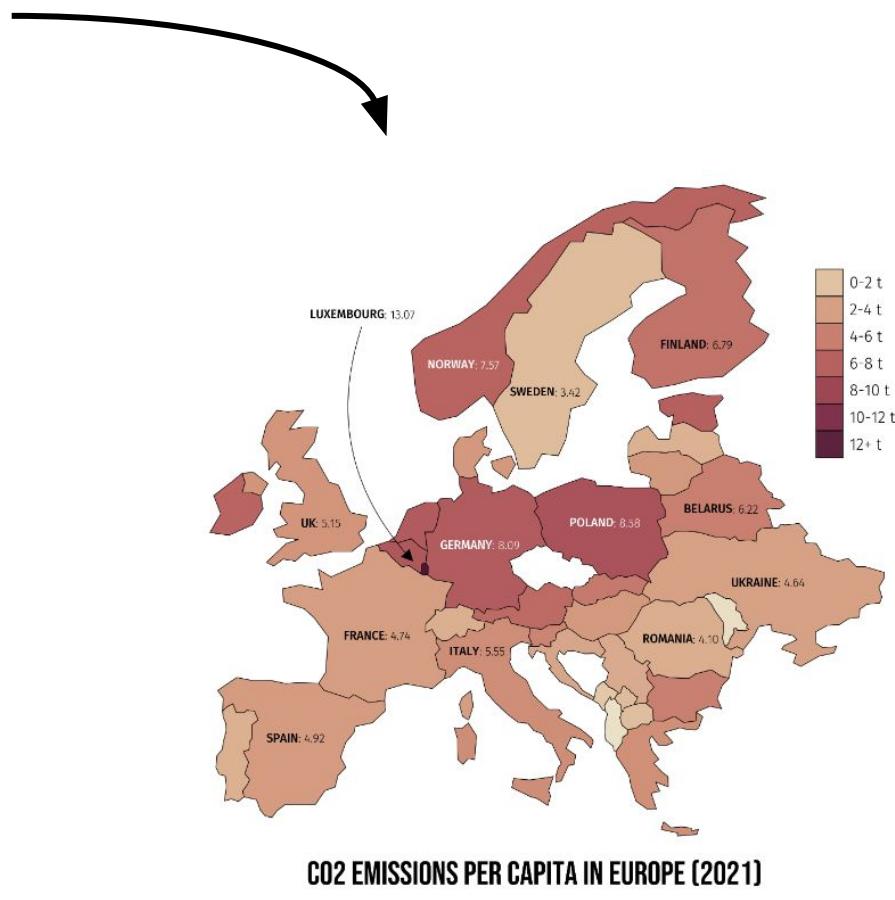
1 - INTRODUCTION

POURQUOI

L'objectif premier de la visualisation de données est de rendre accessible l'information.

Il ne suffit d'avoir l'information pour qu'elle soit facilement consommable.

| Country | Total | Year |
|-----------|-----------|--------|
| Russia | 12.098650 | 2021.0 |
| Norway | 7.573273 | 2021.0 |
| France | 4.741312 | 2021.0 |
| Sweden | 3.424918 | 2021.0 |
| Belarus | 6.222741 | 2021.0 |
| Ukraine | 4.637058 | 2021.0 |
| Poland | 8.577376 | 2021.0 |
| Austria | 7.243330 | 2021.0 |
| Hungary | 4.990287 | 2021.0 |
| Moldova | 1.829658 | 2021.0 |
| Romania | 4.104127 | 2021.0 |
| Lithuania | 4.981189 | 2021.0 |
| Latvia | 3.877550 | 2021.0 |
| Estonia | 7.864087 | 2021.0 |
| Germany | 8.089741 | 2021.0 |

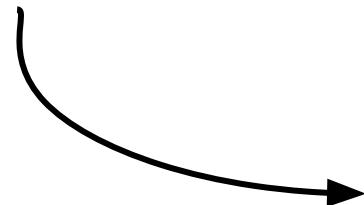


1 - INTRODUCTION

POURQUOI

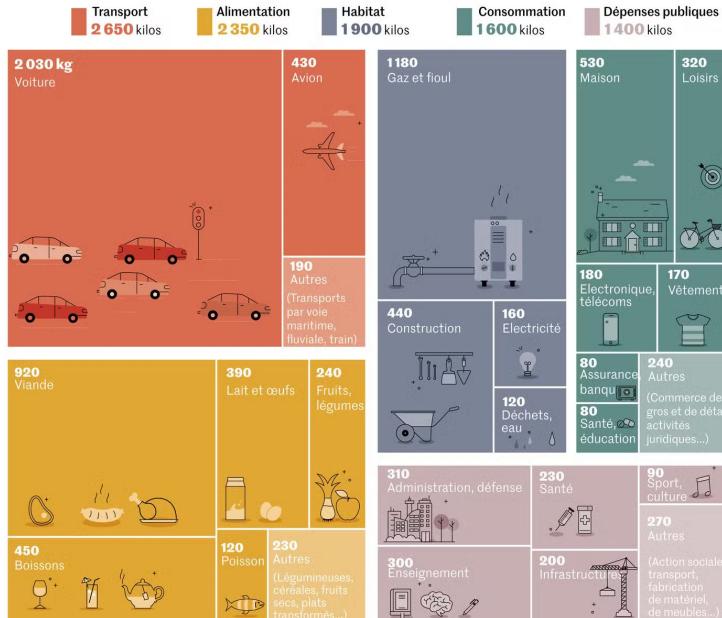
En un coup d'oeil, on voit que les transports et l'alimentation représentent ¼ des parts d'émissions de CO₂.

| catégorie | source | valeur |
|--------------|----------------|--------|
| transport | voiture | 2030 |
| transport | avion | 430 |
| transport | autres | 190 |
| alimentation | viande | 920 |
| alimentation | lait et oeufs | 390 |
| alimentation | fruits/légumes | 240 |
| alimentation | boissons | 450 |
| alimentation | poisson | 120 |
| alimentation | autres | 230 |



Les leviers pour agir sur son empreinte carbone

Répartition des **9,9 tonnes** d'émissions en kilos équivalents CO₂ par habitant en France, en 2019



Sources : Aderme, Rapport des inégalités du monde 2022, MyCO₂ par Carbone 4, d'après le ministère de la transition écologique et le Haut Conseil pour le climat

1 %
des Français les plus riches émet, par individu, **15 fois plus de CO₂ que les 50 % les moins riches**. Les 10 % les plus aisés émettent 1,7 fois plus de gaz à effet de serre que l'ensemble du reste de la population.

Surconsommation

10 millions de tonnes

C'est la quantité de nourriture gaspillée en France chaque année, soit l'équivalent de **150 kilos/hab./an**.

73,9 millions

C'est le nombre d'appareils électroménagers achetés en 2021 en France. Un record.

2 ans

Les Français achètent en moyenne un smartphone tous les deux ans.

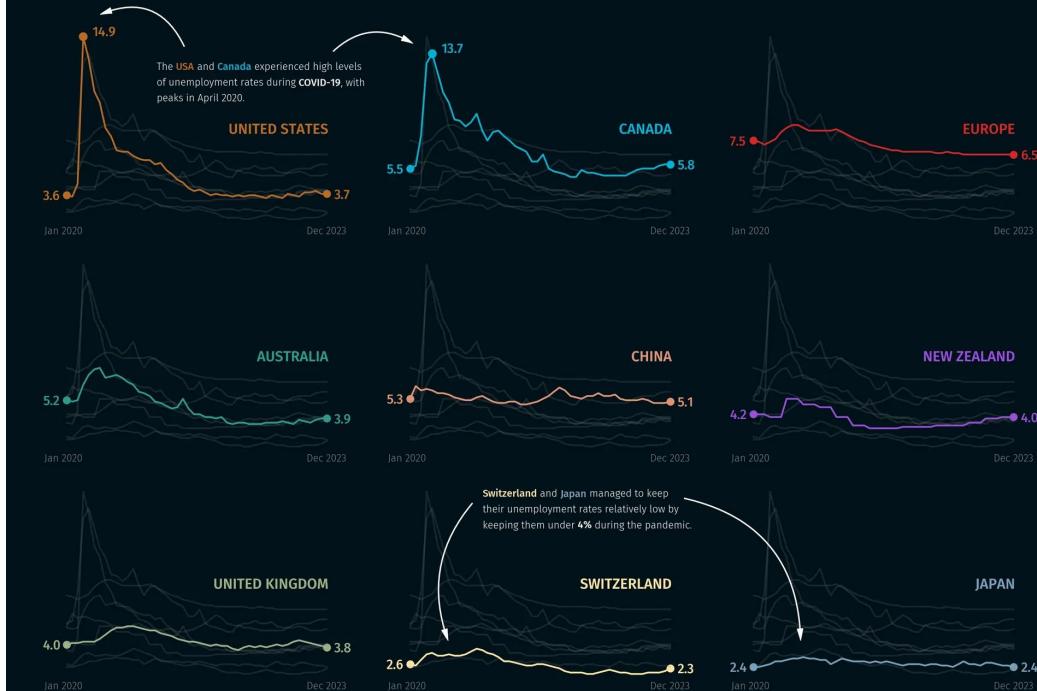
Infographie : Le Monde ; Marianne Pasquier

1 - INTRODUCTION

POURQUOI

La complexité à lire une table montre très rapidement vers des cas où l'on est obligé de trouver un moyen de "représenter" les valeurs.

How countries have been affected by the **COVID-19** pandemic in terms of **unemployment**?
Unemployment rates (in %) in different regions between 2020 and 2023.
Regions are sorted by their maximum unemployment rate during the period.



design: @joseph_barbier
Data: kaggle.com/datasets/keneticenergy/economic-data-life-after-covid

02

CE QU'ON PEUT FAIRE

2 - CE QU'ON PEUT FAIRE DATA-TO-VIZ.COM

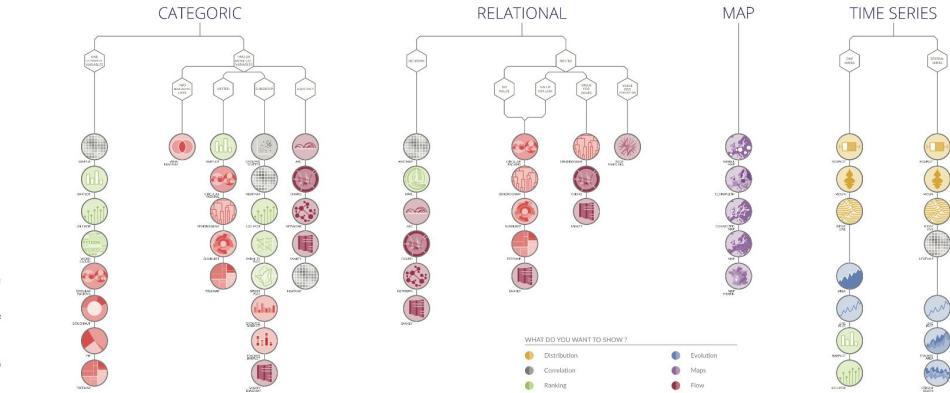


'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps.

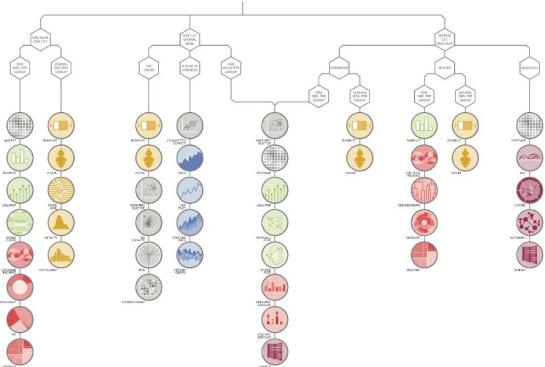
- ① Identify what type of data you have.
- ② Go to the corresponding decision tree and follow it down to a set of possible charts.
- ③ Choose the chart from the set that will suit your data and your needs best.

Data is a word 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 of this chart, visit data-to-viz.com

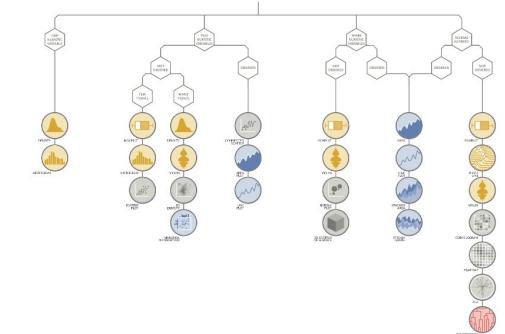
data-to-viz.com



CATEGORIC AND NUMERIC



NUMERIC



Meilleur site pour vous guider dans le choix du bon graphique

2 - CE QU'ON PEUT FAIRE

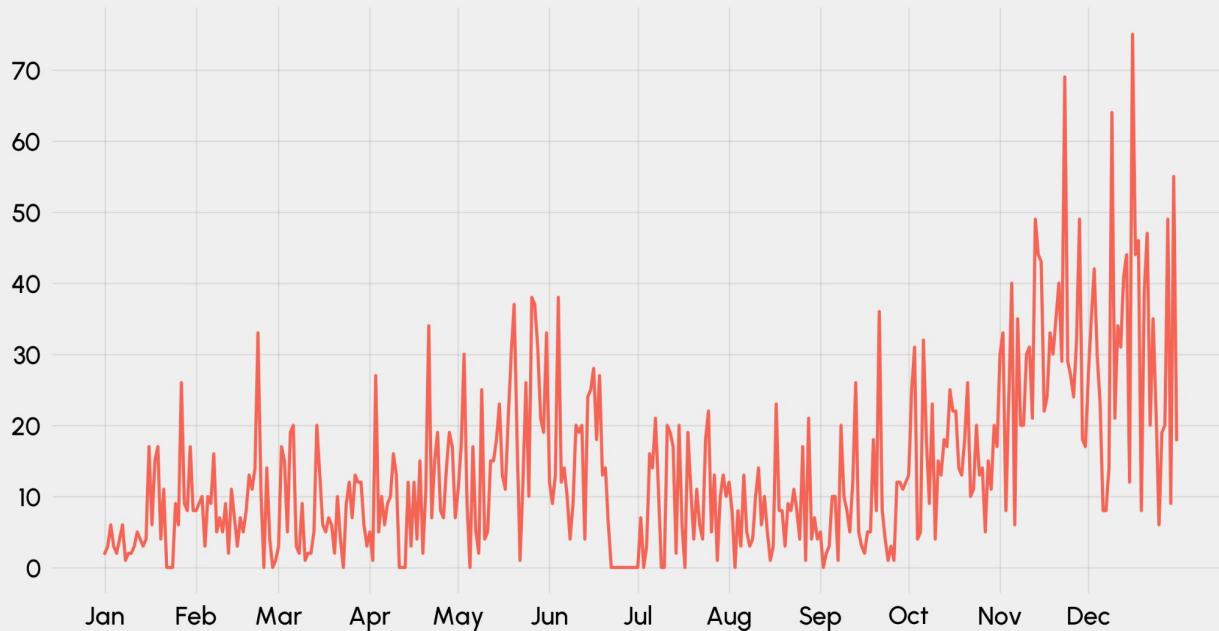
TWEETS MUSK

Imaginons que l'on travaille avec ce jeu de données très simple

| Date | n_tweets |
|------------|----------|
| 2022-01-01 | 2 |
| 2022-01-02 | 3 |
| 2022-01-03 | 6 |
| 2022-01-04 | 3 |
| 2022-01-05 | 2 |
| 2022-01-06 | 4 |
| 2022-01-07 | 6 |
| 2022-01-08 | 1 |
| 2022-01-09 | 2 |
| 2022-01-10 | 2 |

2 - CE QU'ON PEUT FAIRE LINE CHART

Total number of Elon Musk tweets in 2022



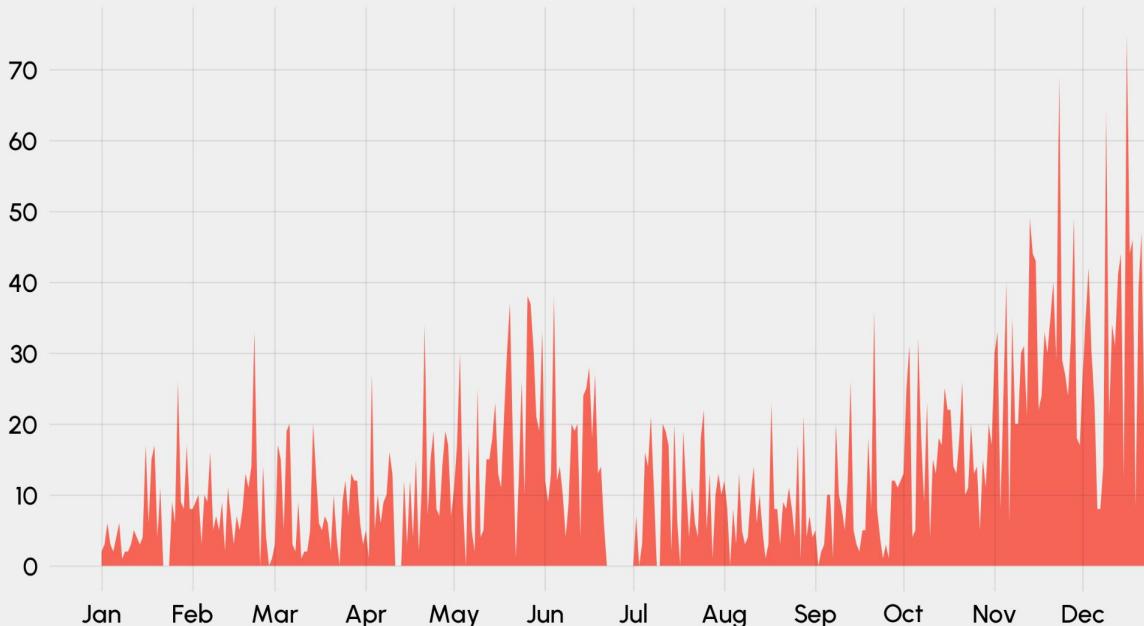
Graphique de série temporelle par défaut

2 - CE QU'ON PEUT FAIRE AREA CHART

Variante du line chart qui donne un effet de "montant" : quelque chose que l'on compte (discret/continu)

- Évolution de la taille : pas très pertinent
- Nombre de jours ensoleillés par semaine à Bordeaux : pourquoi pas

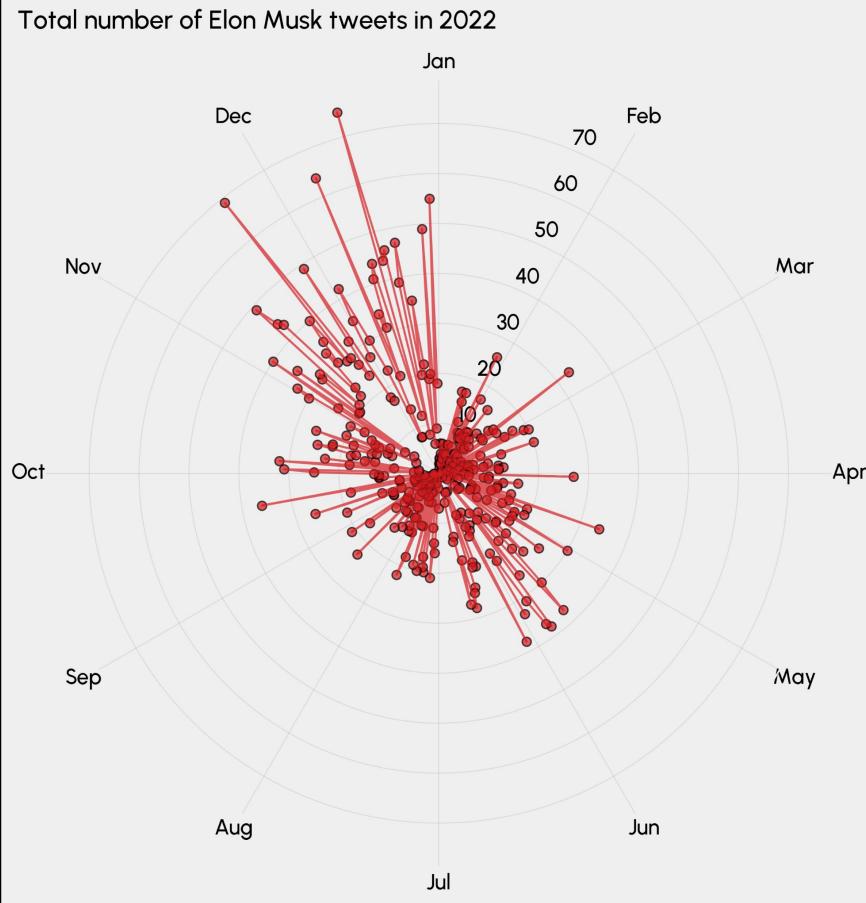
Total number of Elon Musk tweets in 2022



2 - CE QU'ON PEUT FAIRE

CIRCULAR CHART

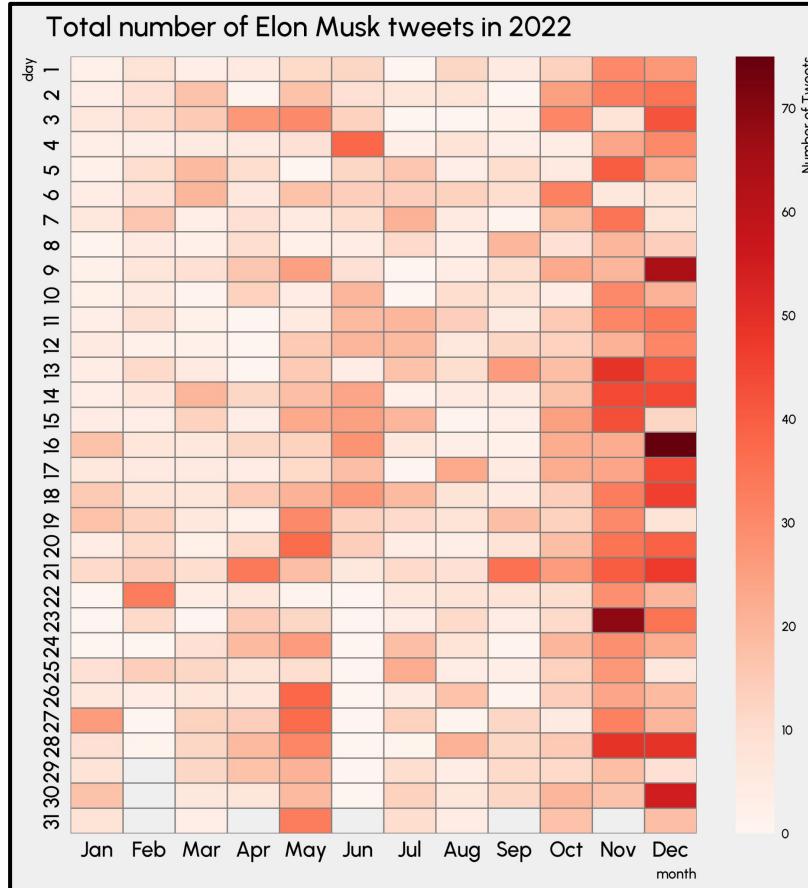
Plus difficile à lire, mais permet de raconter une histoire plus facilement -> annotations



2 - CE QU'ON PEUT FAIRE

CALENDAR HEATMAP

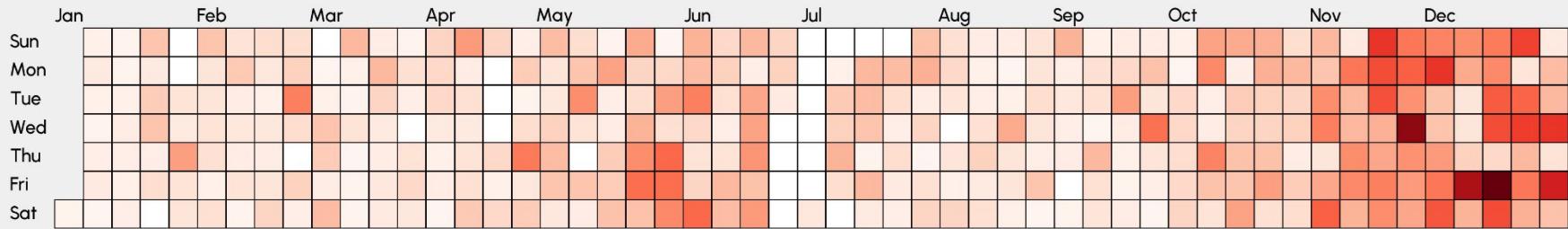
Mettre en avant les patterns mensuels



2 - CE QU'ON PEUT FAIRE

CALENDAR HEATMAP

Total number of Elon Musk tweets in 2022

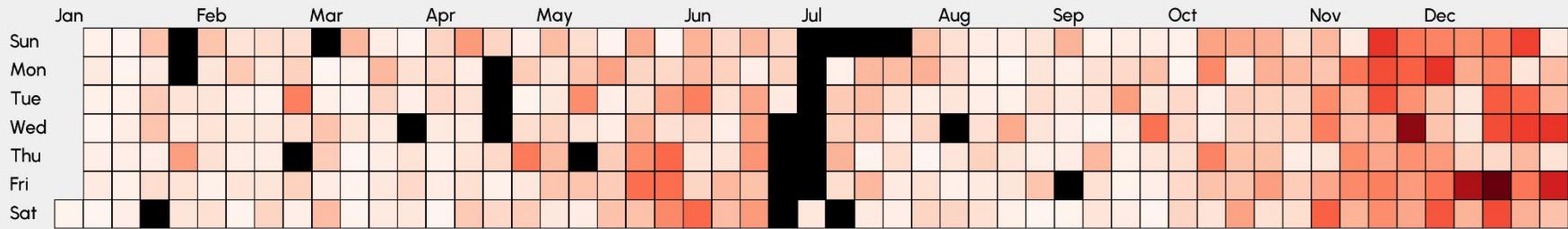


Mettre en avant les patterns hebdomadiers

2 - CE QU'ON PEUT FAIRE

CALENDAR HEATMAP

Total number of Elon Musk tweets in 2022

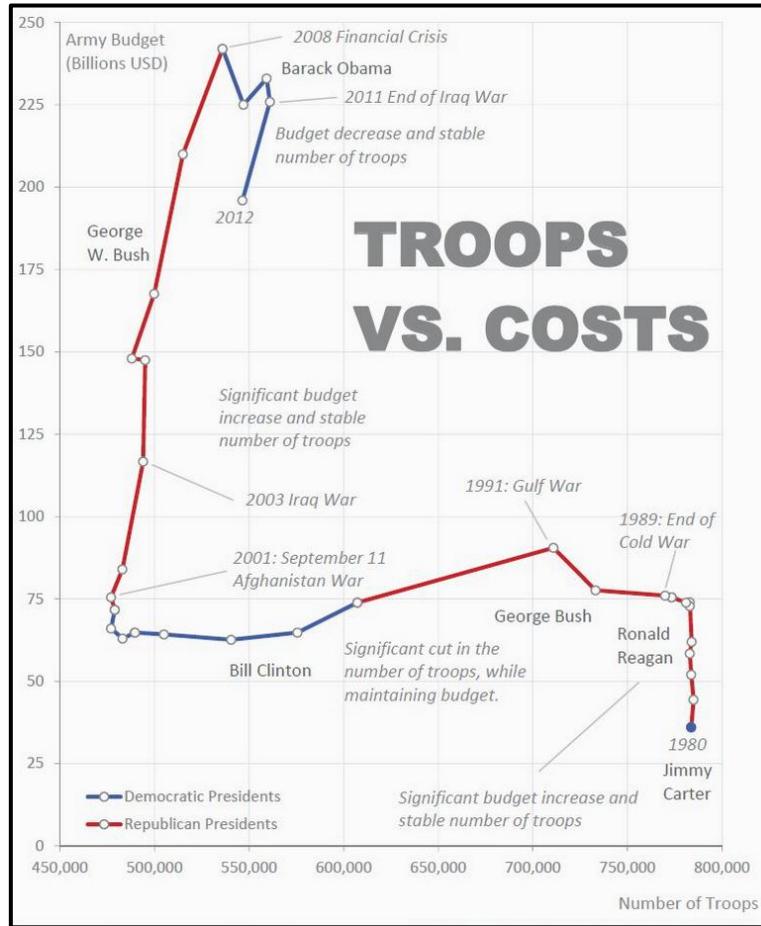


Mettre en avant des jours en particuliers

2 - CE QU'ON PEUT FAIRE AUTRES

Connected scatter plot

Storytelling ++
Précision +++
Lecture -

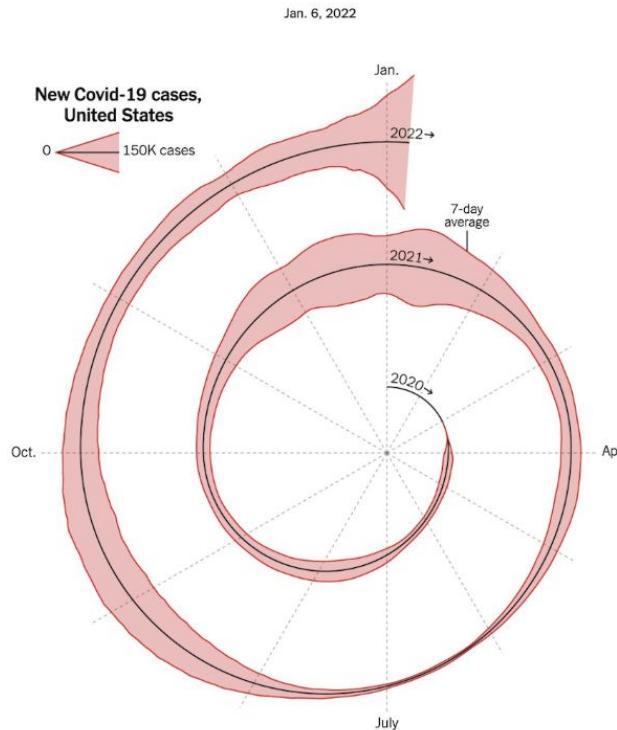


2 - CE QU'ON PEUT FAIRE **AUTRES**

Area circular chart

Storytelling +++
Tendance ++
Précision --
Lecture --

Here's When We Expect Omicron to Peak



2 - CE QU'ON PEUT FAIRE AUTRES

Calendar heatmap

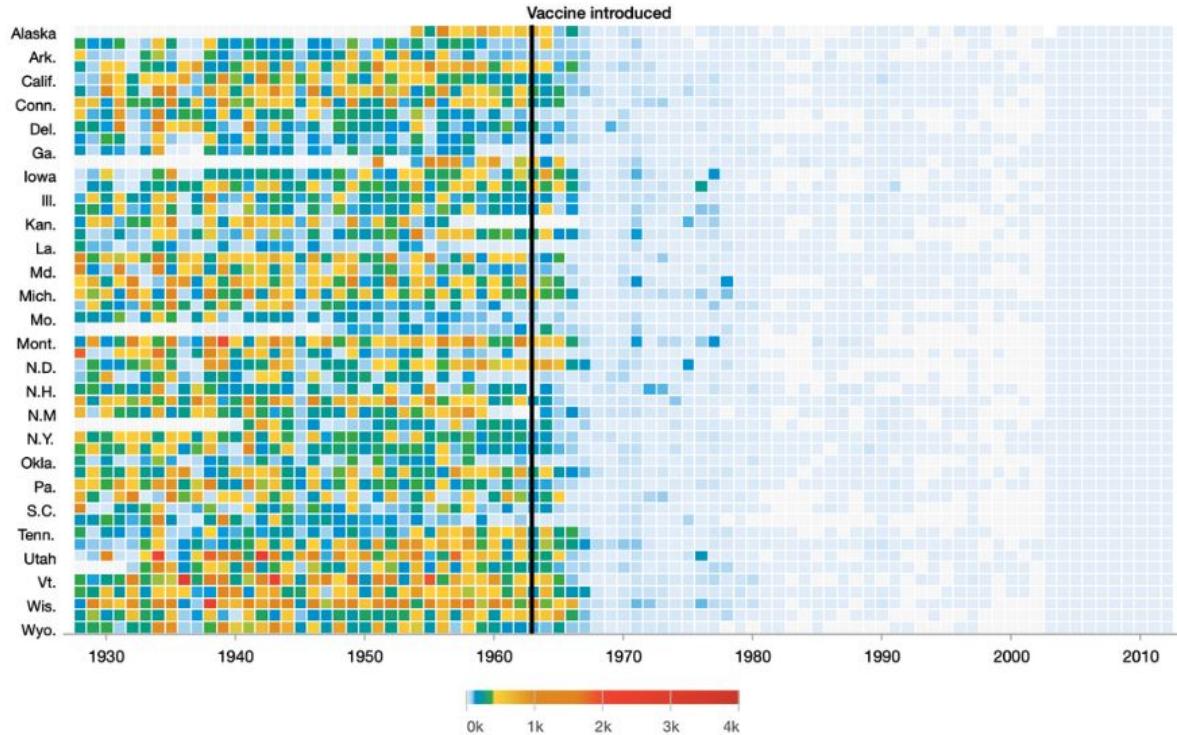
Storytelling +++

Tendance +++

Précision ---

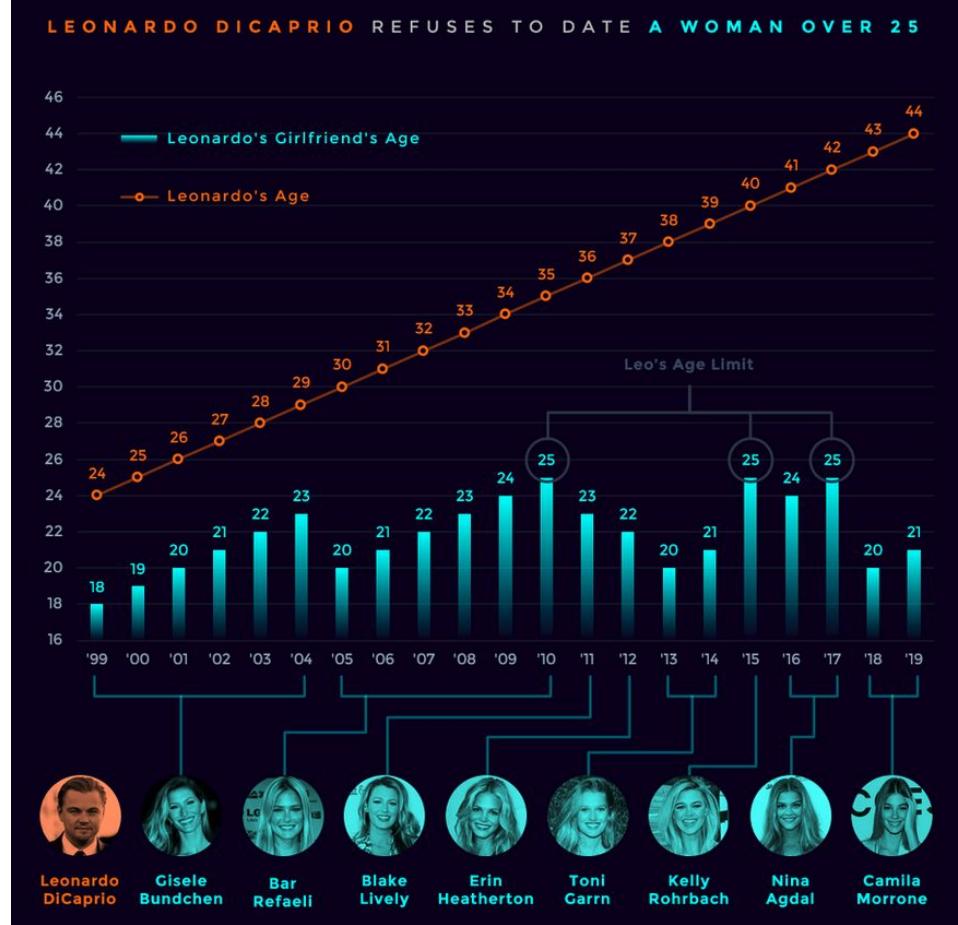
Lecture ++

Measles



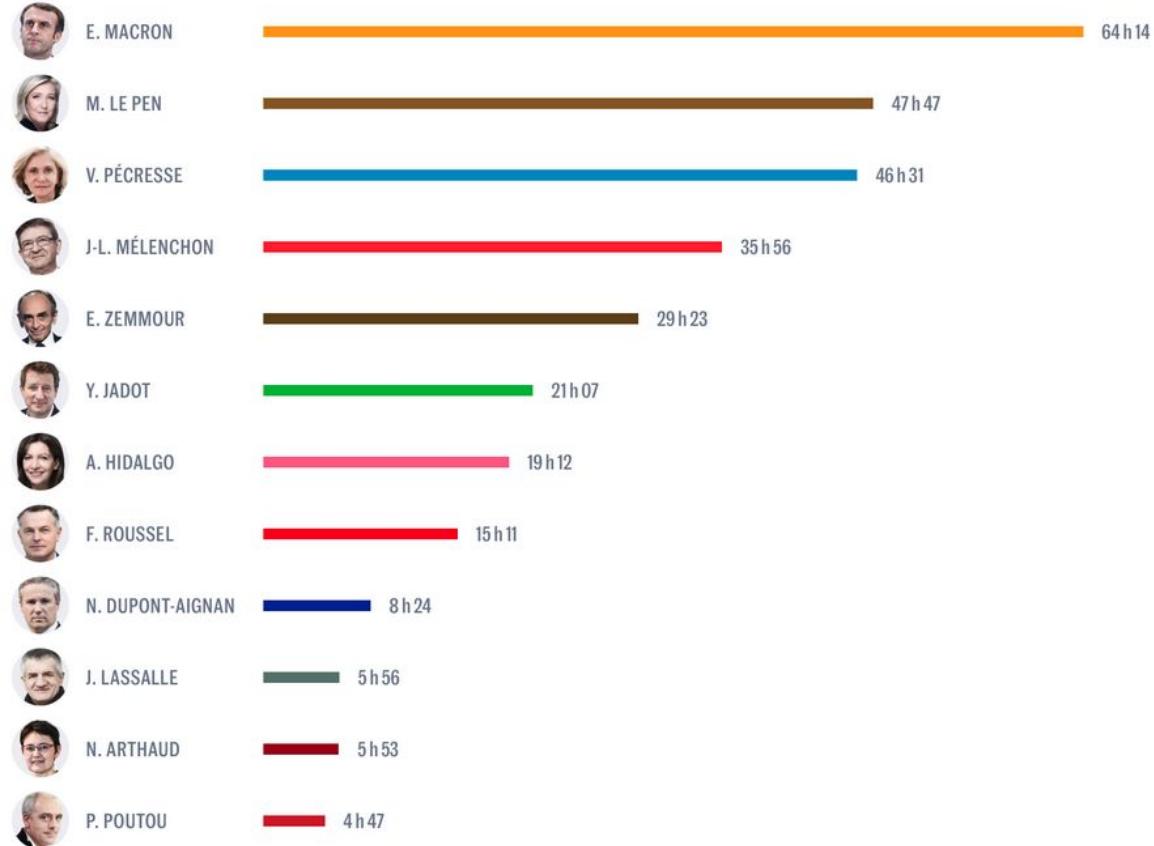
2 - CE QU'ON PEUT FAIRE AUTRES

Données non temporelles ?
Eh bien si !



2 - CE QU'ON PEUT FAIRE AUTRES

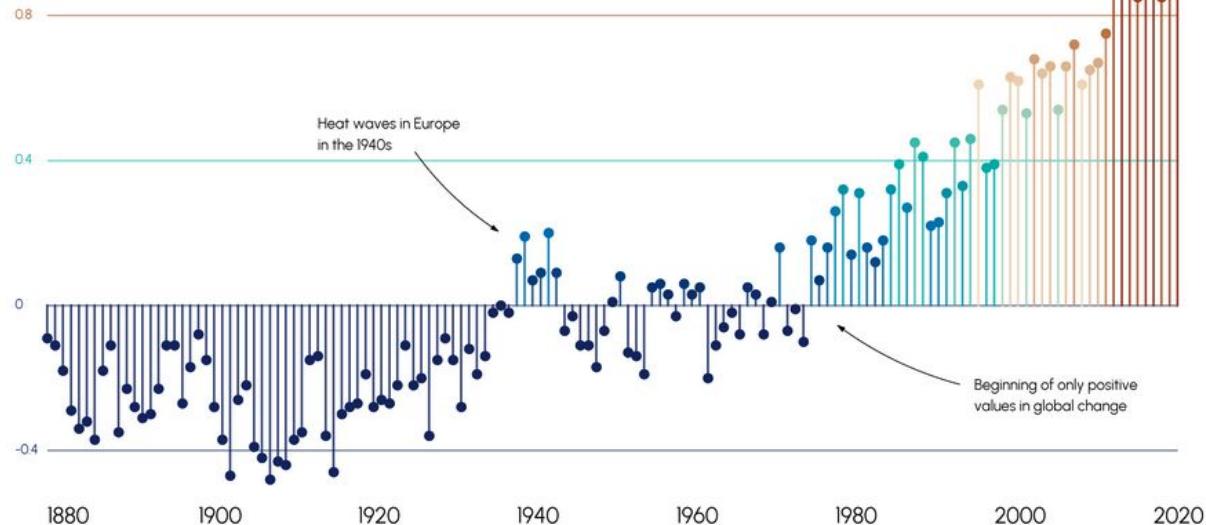
Données temporelles ?
Eh bien non !



2 - CE QU'ON PEUT FAIRE AUTRES

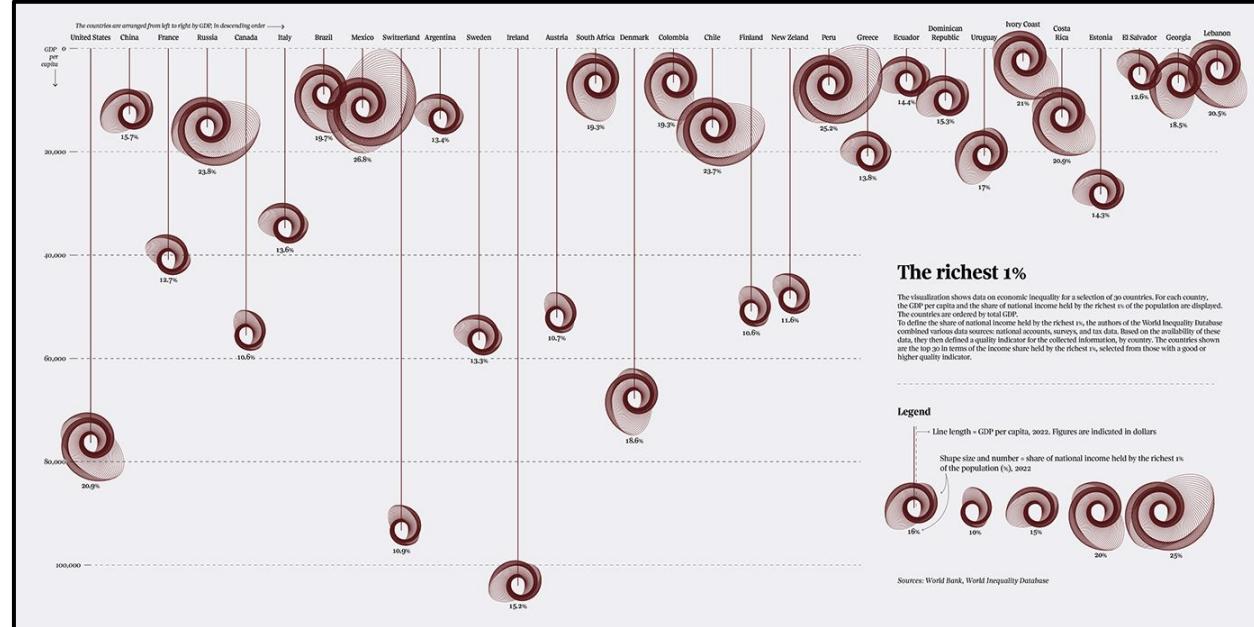
Global Land-Ocean Temperature Index

Change in global surface temperature compared to the long-term average from 1951 to 1980



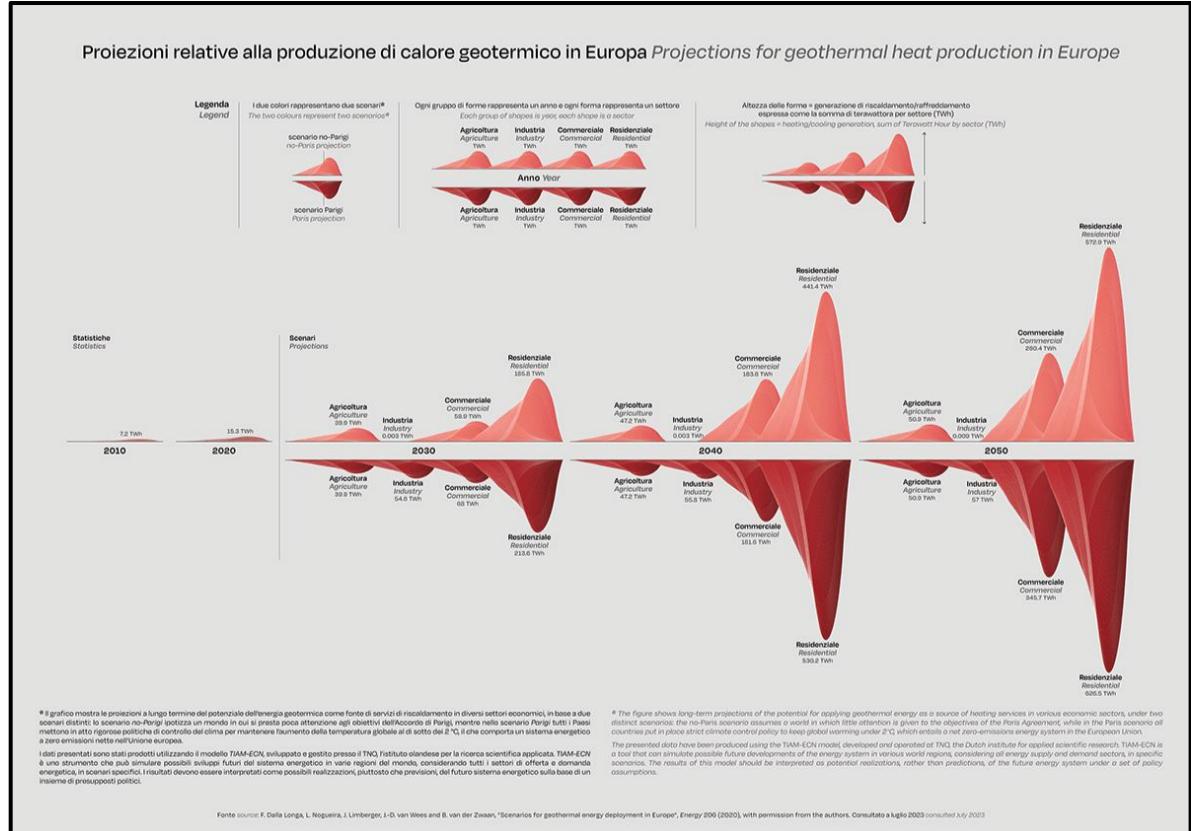
Graph barbierjoseph.com
Data Source: NASA

2 - CE QU'ON PEUT FAIRE COMPROMIS



Federica
Fragapane

2 - CE QU'ON PEUT FAIRE COMPROMIS



Federica
Fragapani

2 - CE QU'ON PEUT FAIRE **COMPROMIS**

| | | | | | |
|--|--|---|--|------|--------|
| | | ChatGPT OpenAI | | 22M | \$62M |
| | | ChatGPT OpenAI | | 40M | \$12M |
| | | Claude by Anthropic Anthropic PBC | | 400K | \$900K |
| | | Claude by Anthropic Anthropic PBC | | 600K | \$400K |

03

CE QU'ON DOIT FAIRE

3 - CE QU'ON DOIT FAIRE **LECTEURS**



CHERCHEURS



JOURNALISTES



GRAND PUBLIQUE



VOTRE GRAND-MÈRE

3 - CE QU'ON DOIT FAIRE **FORMAT**

Web

- Beaucoup plus complexe techniquement
- Storytelling +++

<https://colasdroin.github.io/snsf-dataviz/>

<https://holtzy.github.io/Mortality/>

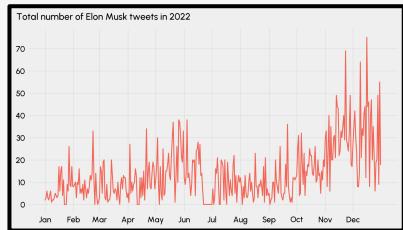
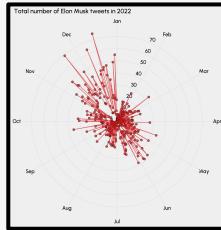
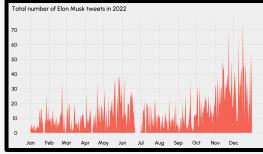
Static

- Design plus complexe
- Plus simple techniquement

Interactif

- R: ggiraph, plotly
- Python: plotly, bokeh, altair
- Bon compromis complexité / fonctionnalité

3 - CE QU'ON DOIT FAIRE OBJECTIF

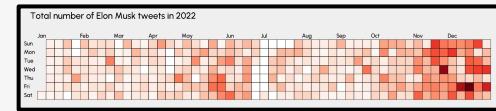
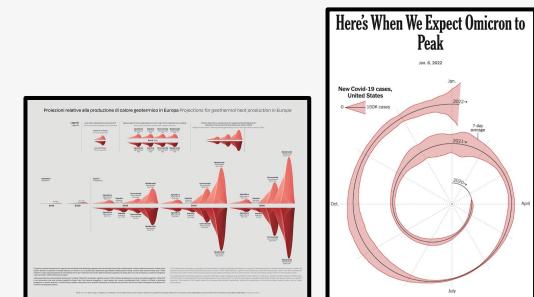


EXPLORATOIRE

Vous êtes en train de vous “balader” dans votre jeu de données et essayez de trouver ce que les données vous raconte.

DESCRIPTIF ⚠

Vous décrivez de manière assez formelle un évènement : **“Evolution du nombre de tweets de Elon Musk”**



STORYTELLING ✓

Présentation & Insight sur un phénomène : **“Depuis Octobre, Elon Musk tweet 30 fois par jours en moyenne”**

04

ÇE QU'IL FAUT ÉVITER

4 - CE QU'IL FAUT ÉVITER

DATA-INK RATIO

$$\text{Ratio} = \frac{\text{Eléments informatifs}}{\text{Tous les éléments du graphique}}$$

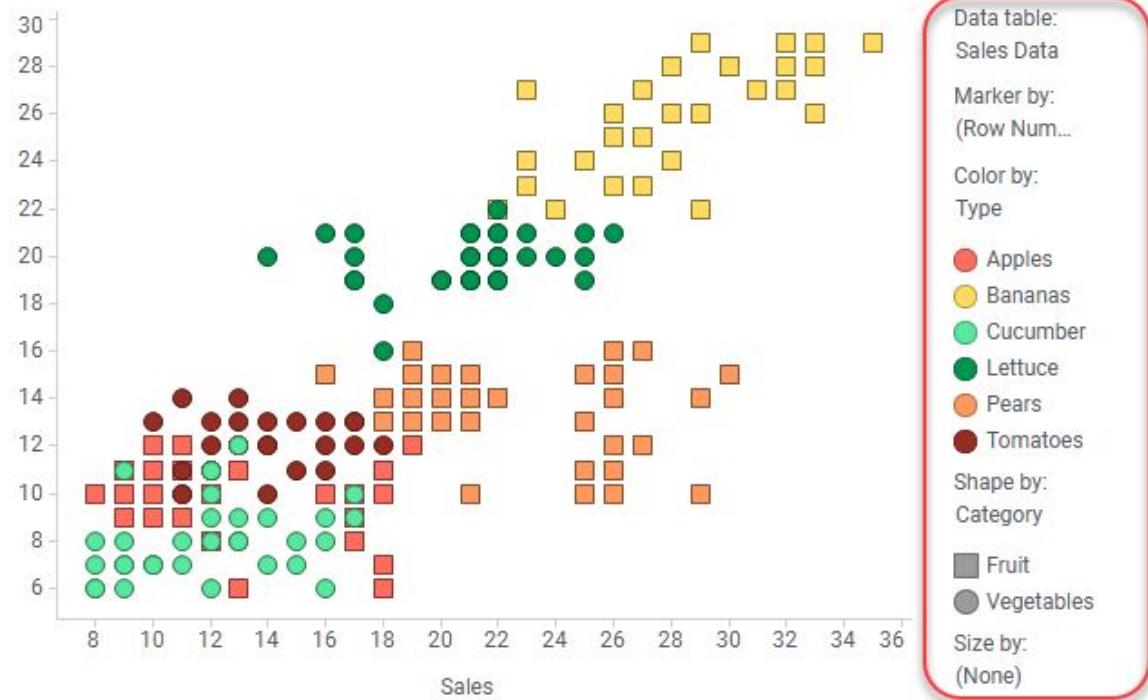
"Above all else show the data." Edward Tufte

4 - CE QU'IL FAUT ÉVITER

LEGENDES

Une légende sert à expliquer les couleurs, symboles ou formes utilisés dans un graphique.

Elle permet d'identifier les catégories, de comprendre les données représentées, et de faciliter la lecture du visuel.



4 - CE QU'IL FAUT ÉVITER

LEGENDES

Très souvent, on en a **pas trop besoin**.

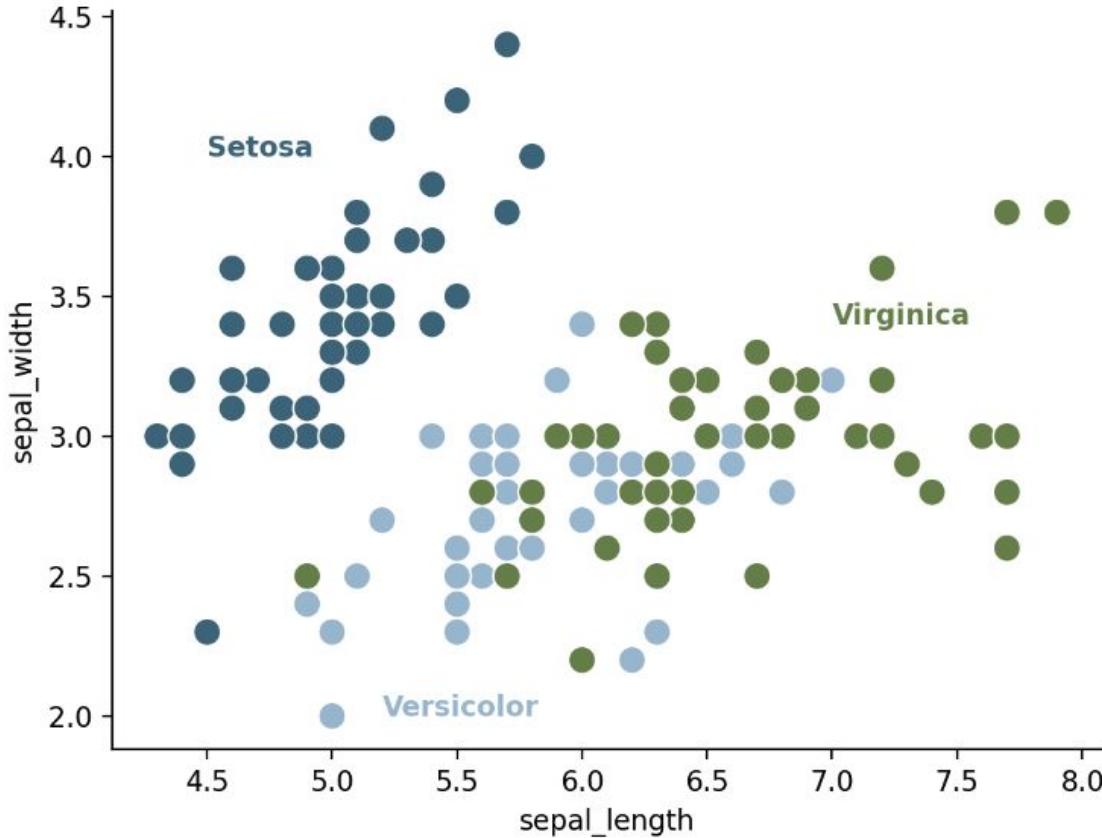
Quelques annotations facilitent énormément la lecture.



4 - CE QU'IL FAUT ÉVITER LEGENDES

Très souvent, on en a pas du tout besoin.

Quelques annotations facilitent énormément la lecture.



LEGENDES



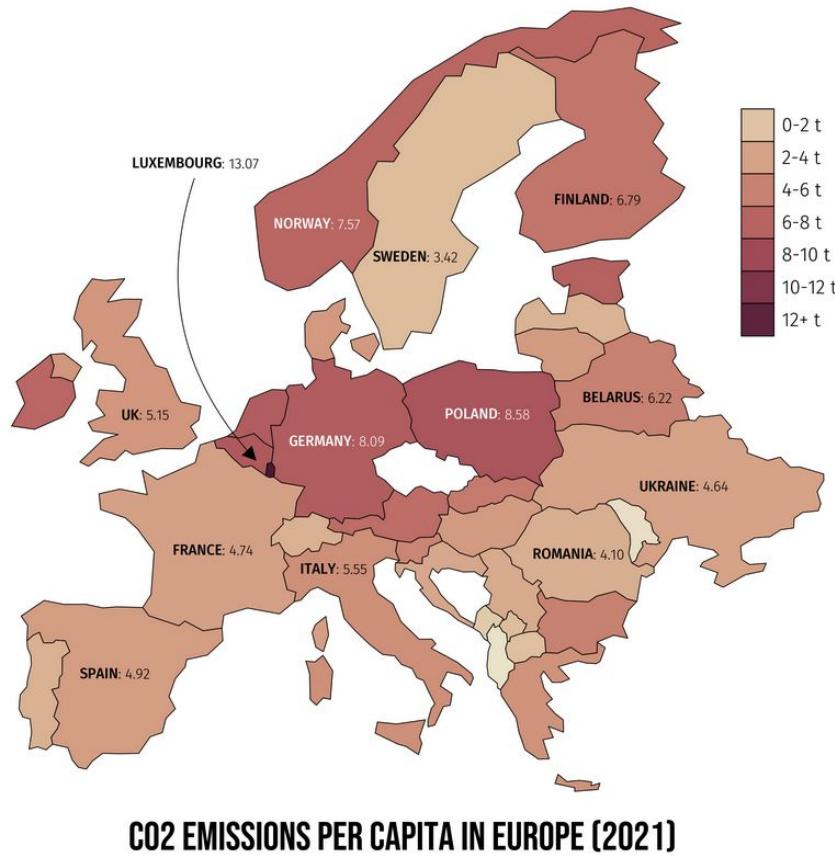
Ne laissez pas votre outil déterminer vos choix de design!!

Que ça soit R, Python, Excel, ou n'importe quoi d'autres, le style par défaut est rarement le bon.

4 - CE QU'IL FAUT ÉVITER

LEGENDES

Quand c'est nécessaire

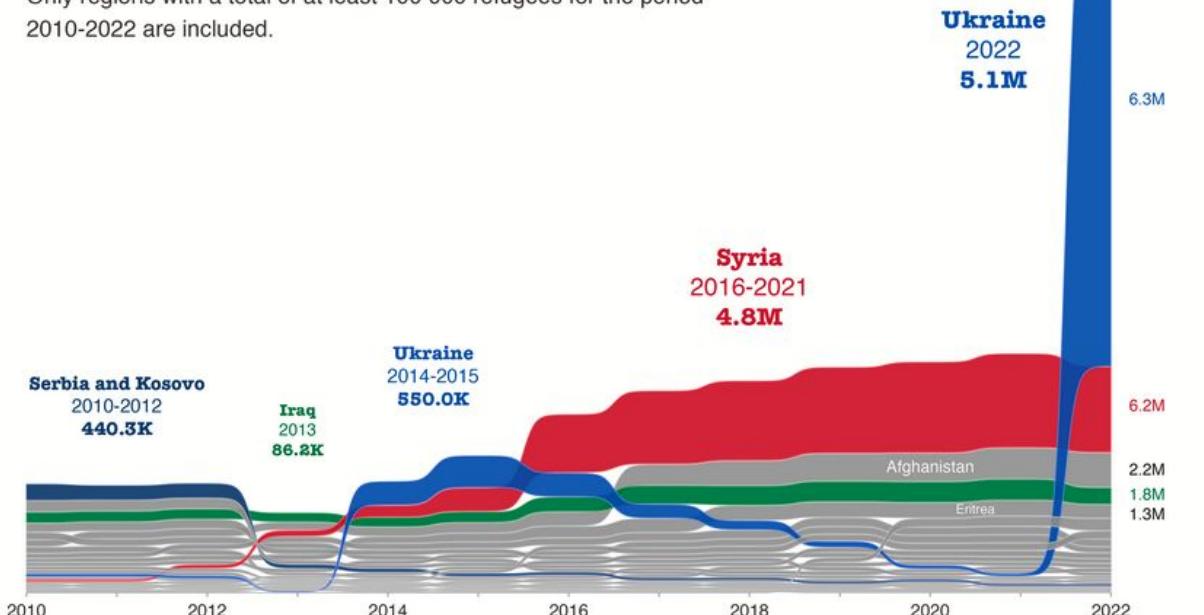


4 - CE QU'IL FAUT ÉVITER LEGENDES

Quand ce n'est **pas nécessaire**

Where do the 22 million refugees in Europe come from?

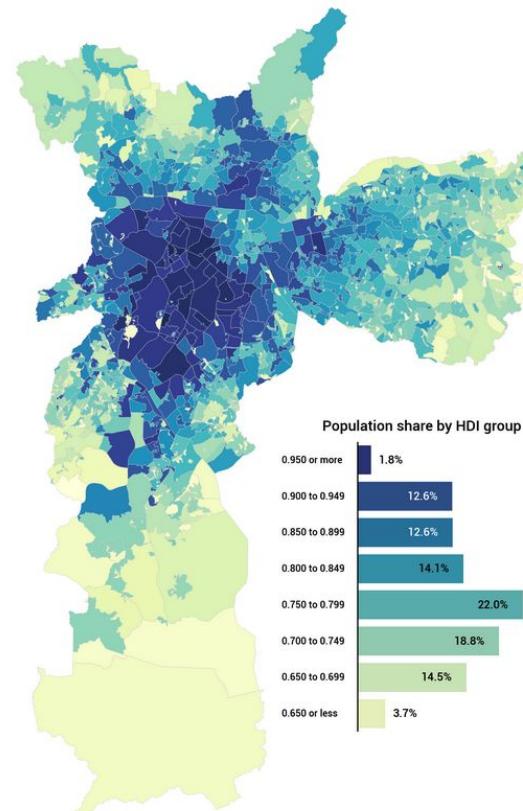
The chart displays rankings and the number of refugees from each region from 2010 to 2022. Top-ranked regions during different periods are highlighted.
Only regions with a total of at least 100 000 refugees for the period 2010-2022 are included.



4 - CE QU'IL FAUT ÉVITER

LEGENDES

HDI in Sao Paulo, BR (2010)



Source: Atlas Brasil

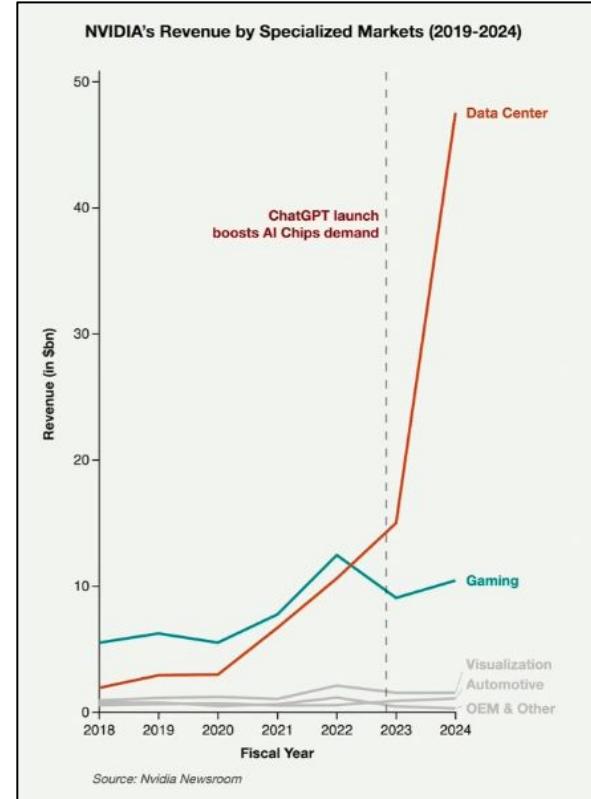
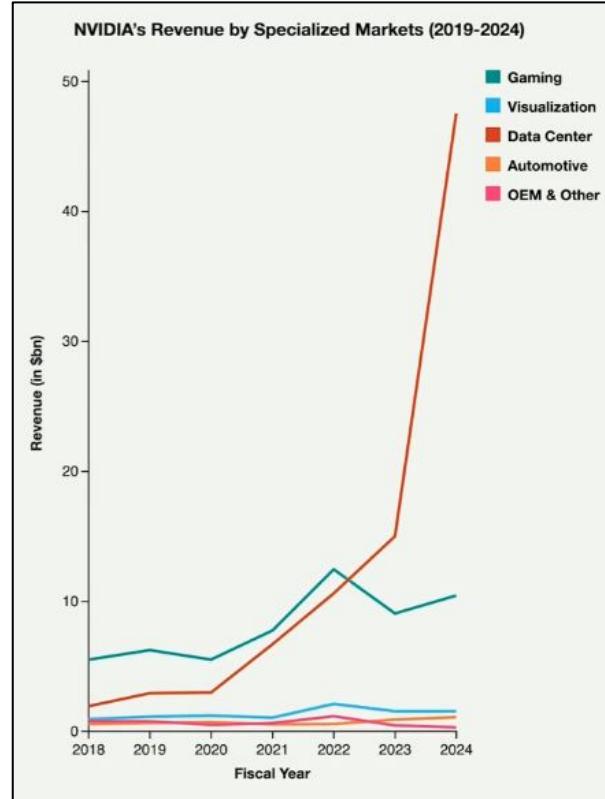
Exemple d'une légende informative +++

4 - CE QU'IL FAUT ÉVITER LEGENDES

Source (et recommandation)

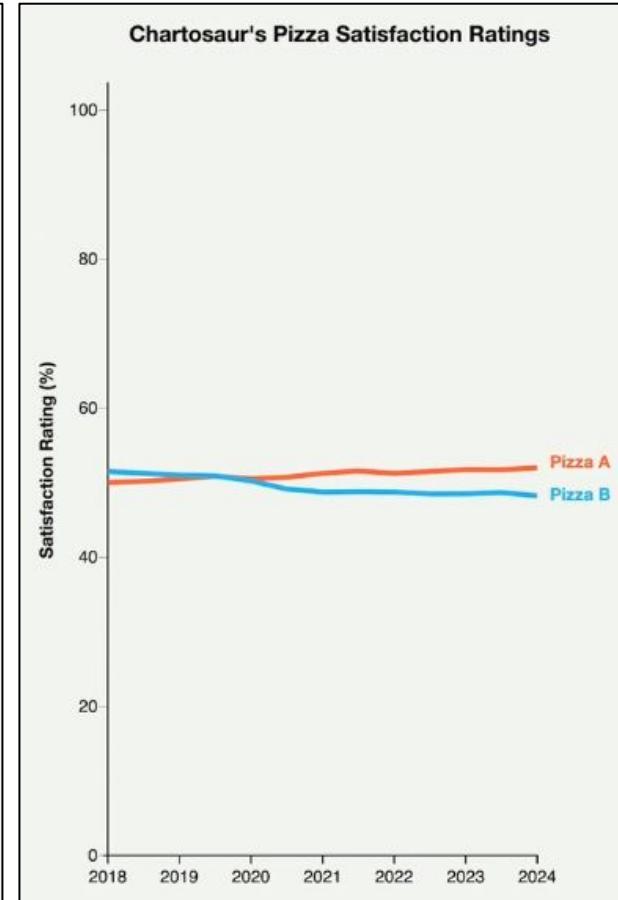
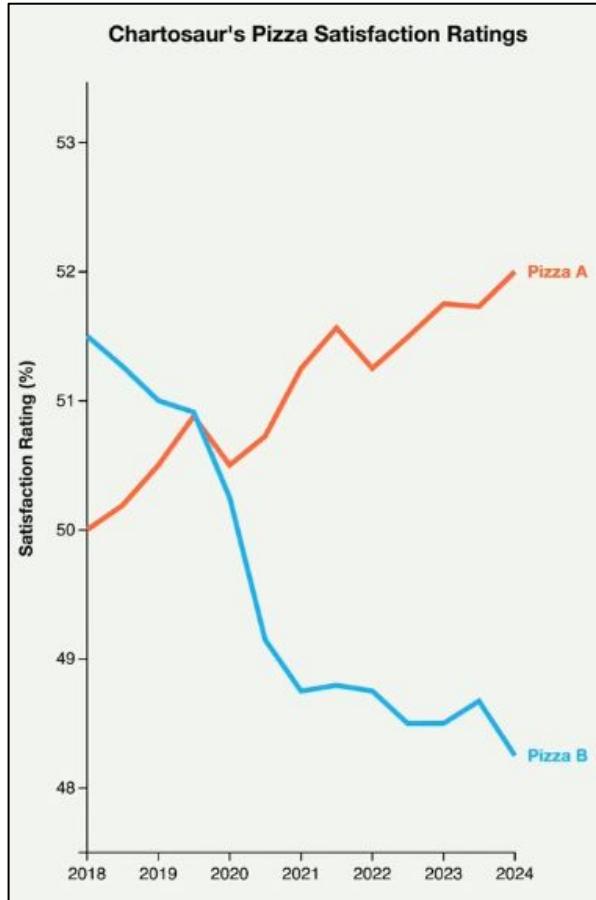


Ne pas utiliser de légende est aussi un bon moyen de raconter une histoire (storytelling) et éviter d'être uniquement dans le descriptif.



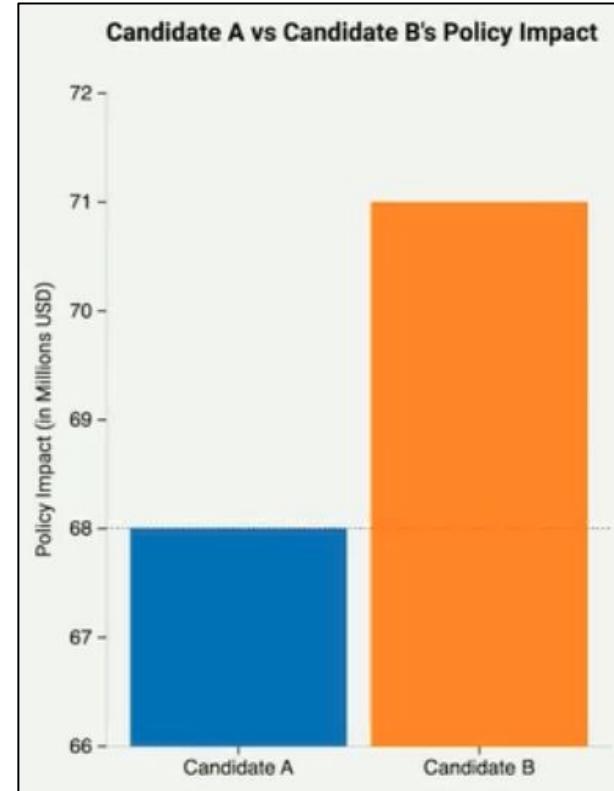
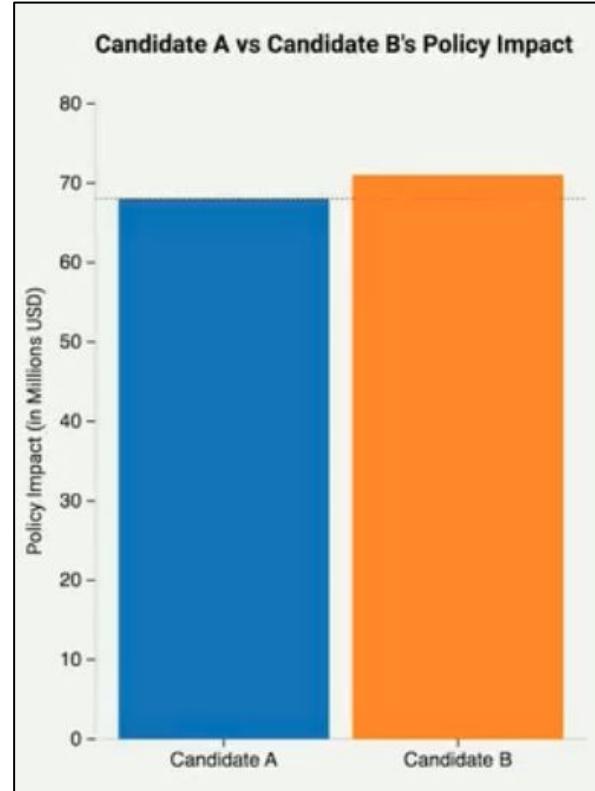
4 - CE QU'IL FAUT ÉVITER ECHELLES

Source (et recommandation)



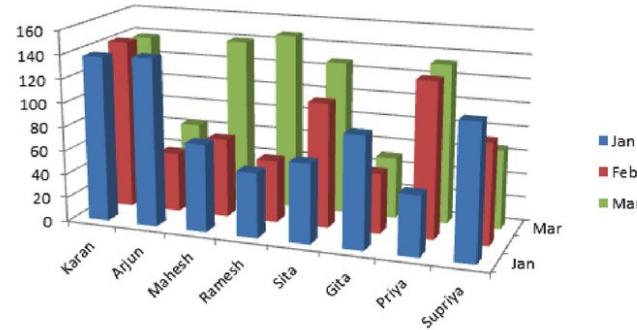
4 - CE QU'IL FAUT ÉVITER ECHELLES

Source (et recommandation)



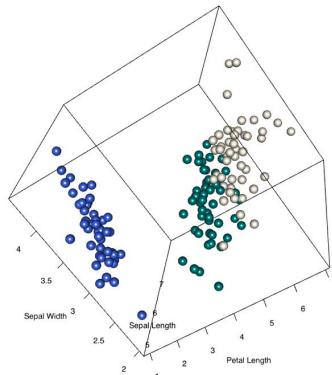
4 - CE QU'IL FAUT ÉVITER

3D

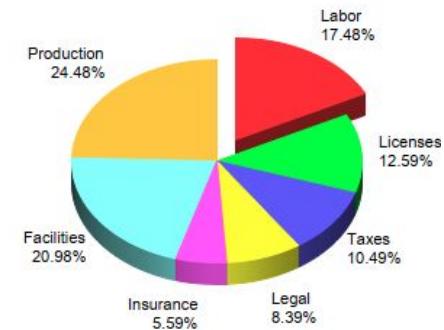


La 3D dans 99% du temps :

- Ça n'apporte rien
- Ça ajoute de la complexité
- C'est difficile à lire
- C'est moche

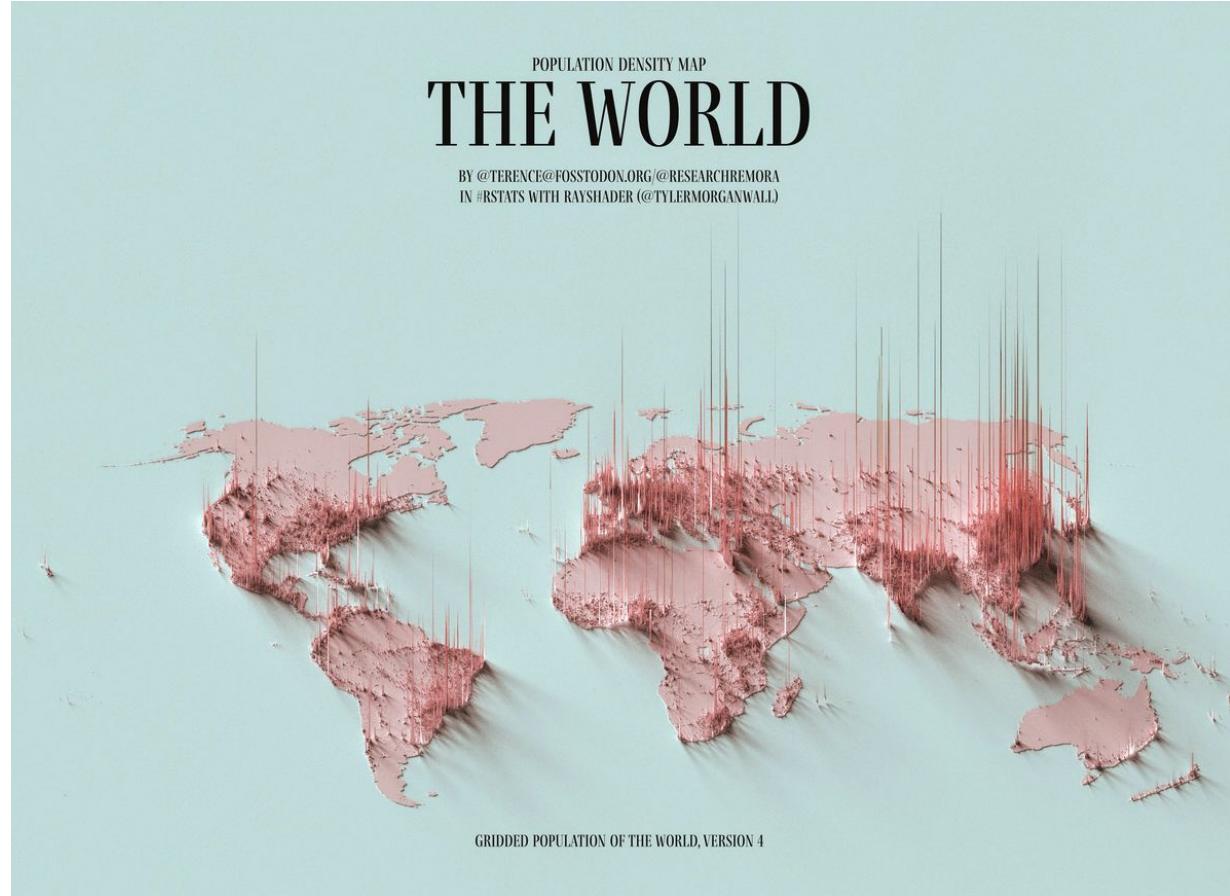


Project Cost Breakdown



4 - CE QU'IL FAUT ÉVITER 3D

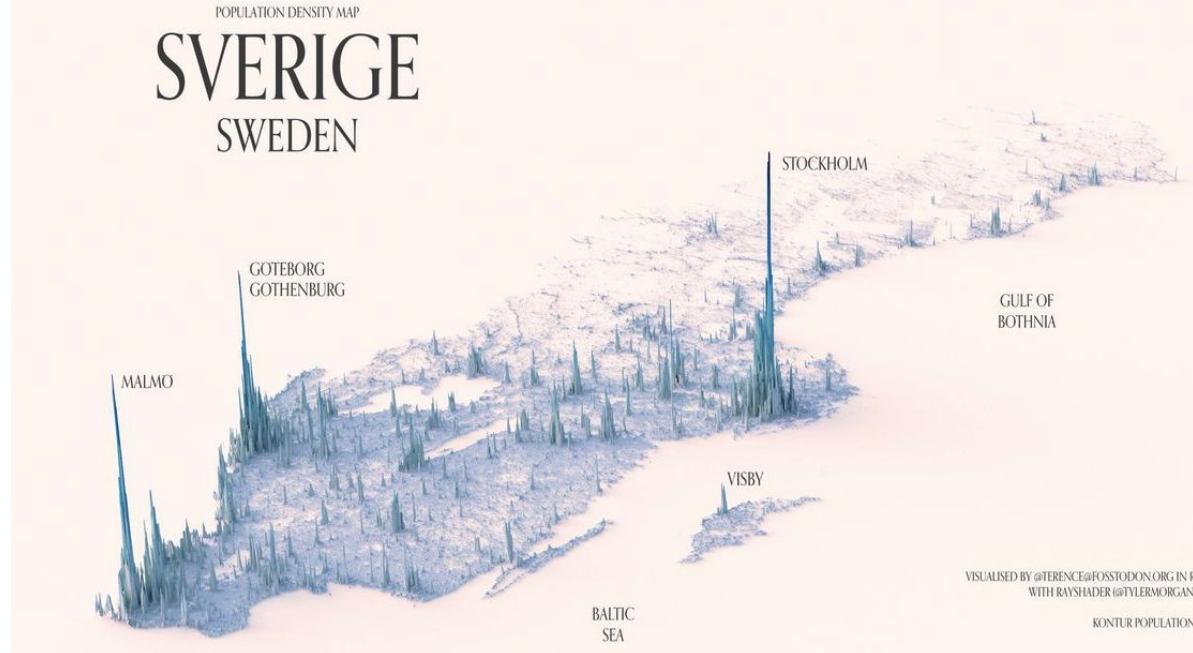
Mais ça peut être très jolie!



4 - CE QU'IL FAUT ÉVITER

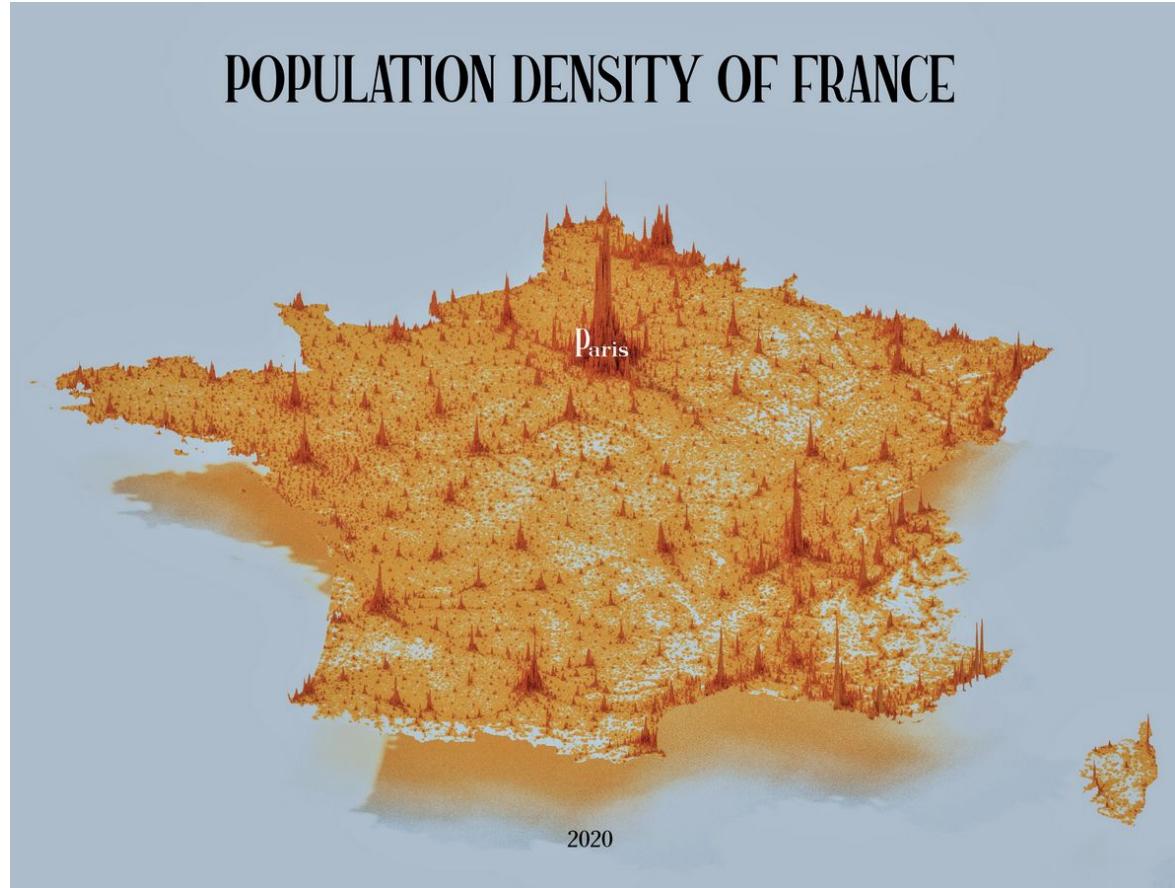
3D

Mais ça peut être très jolie!



4 - CE QU'IL FAUT ÉVITER 3D

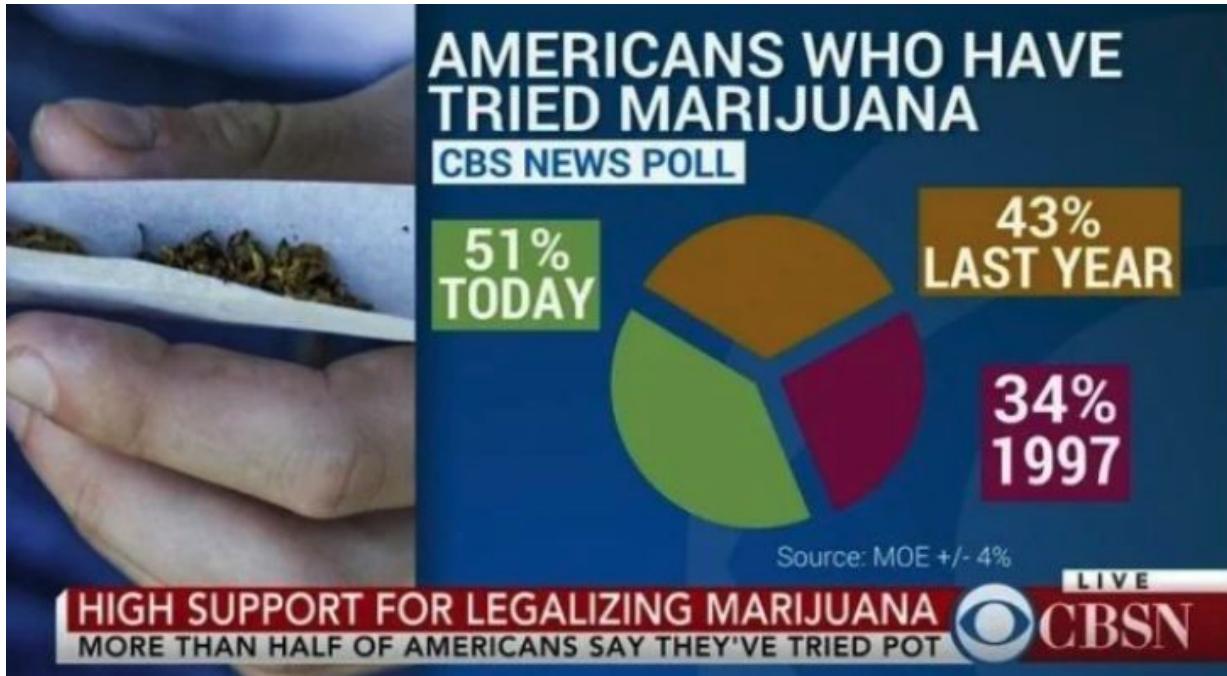
Mais ça peut être très jolie!



4 - CE QU'IL FAUT ÉVITER

BONUS: DANS LES MÉDIAS

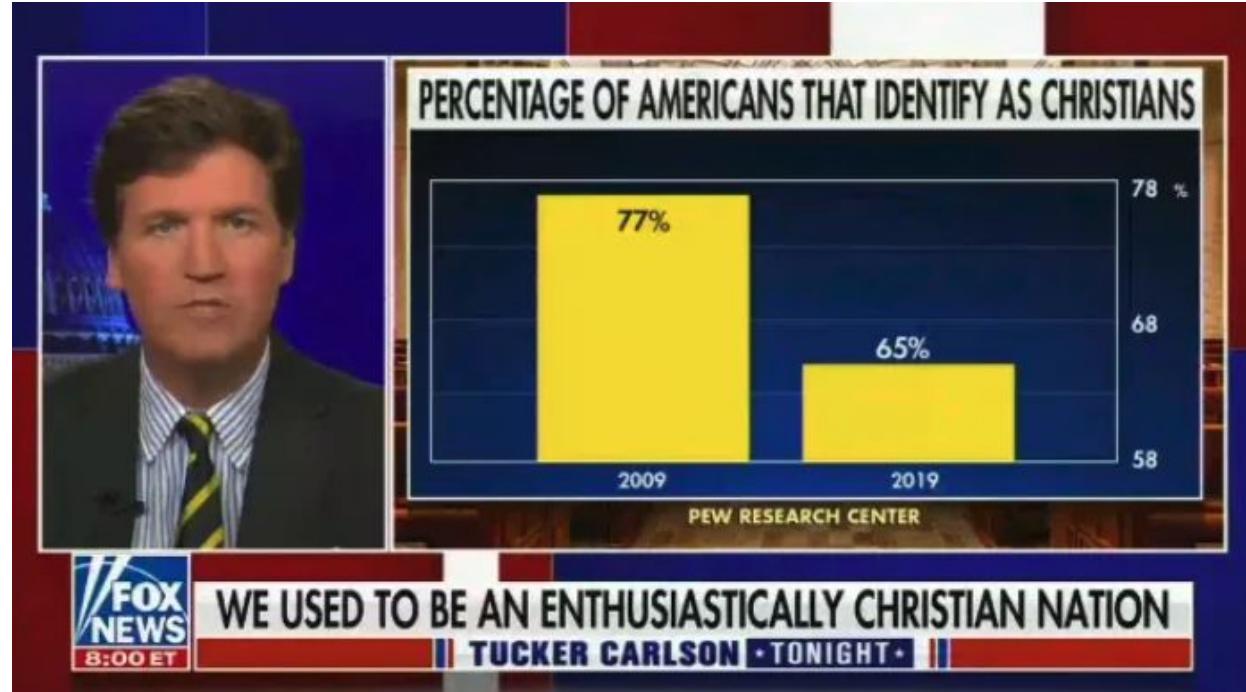
Attention aux conflits d'intérêts !



4 - CE QU'IL FAUT ÉVITER

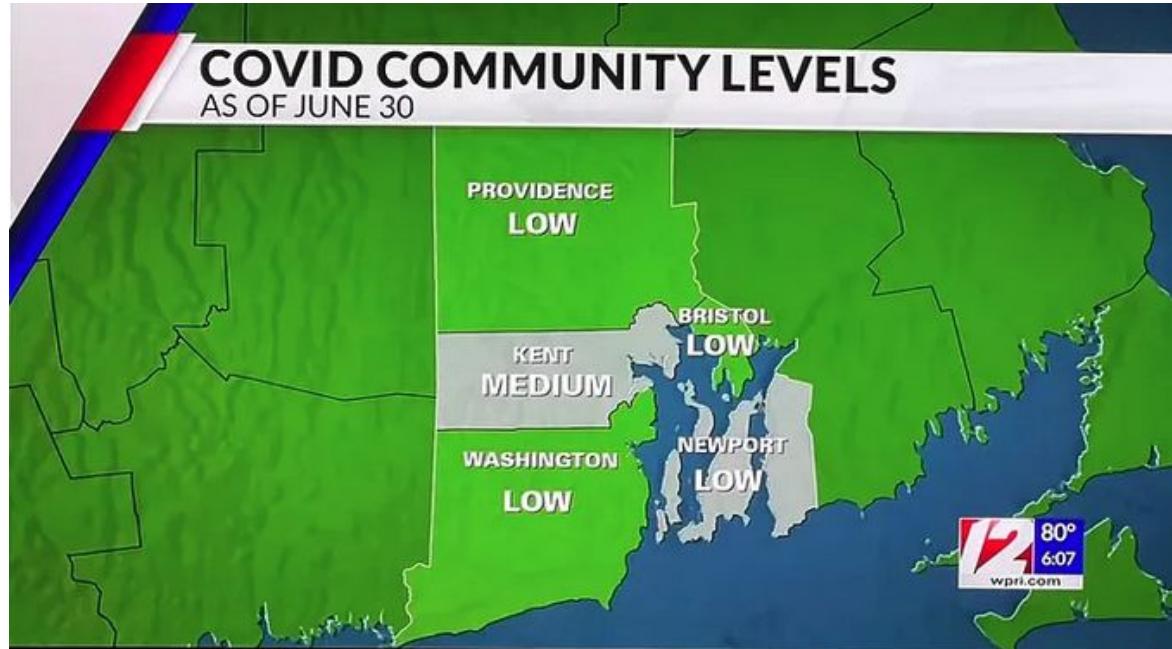
BONUS: DANS LES MÉDIAS

Attention aux conflits d'intérêts !



4 - CE QU'IL FAUT ÉVITER

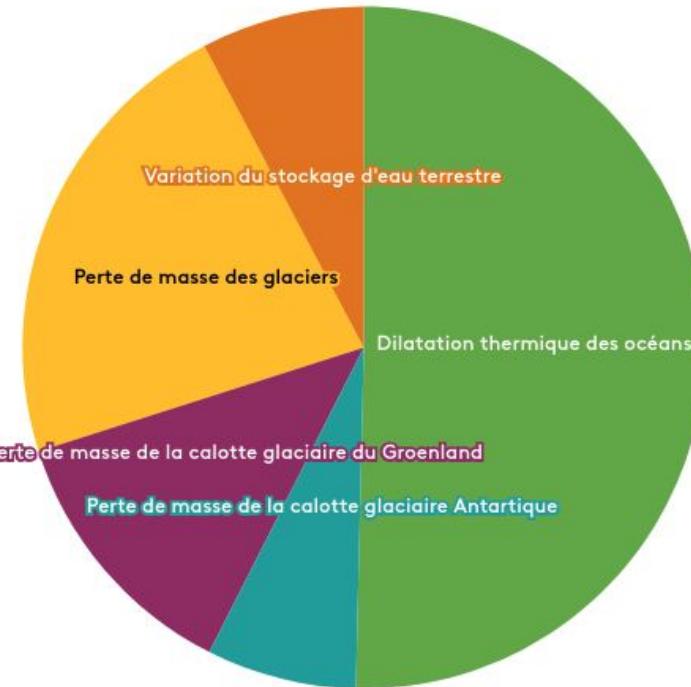
BONUS: DANS LES MÉDIAS



4 - CE QU'IL FAUT ÉVITER

BONUS: DANS LES MÉDIAS

Contributions à l'élévation du niveau global de la mer entre 1971 et 2018

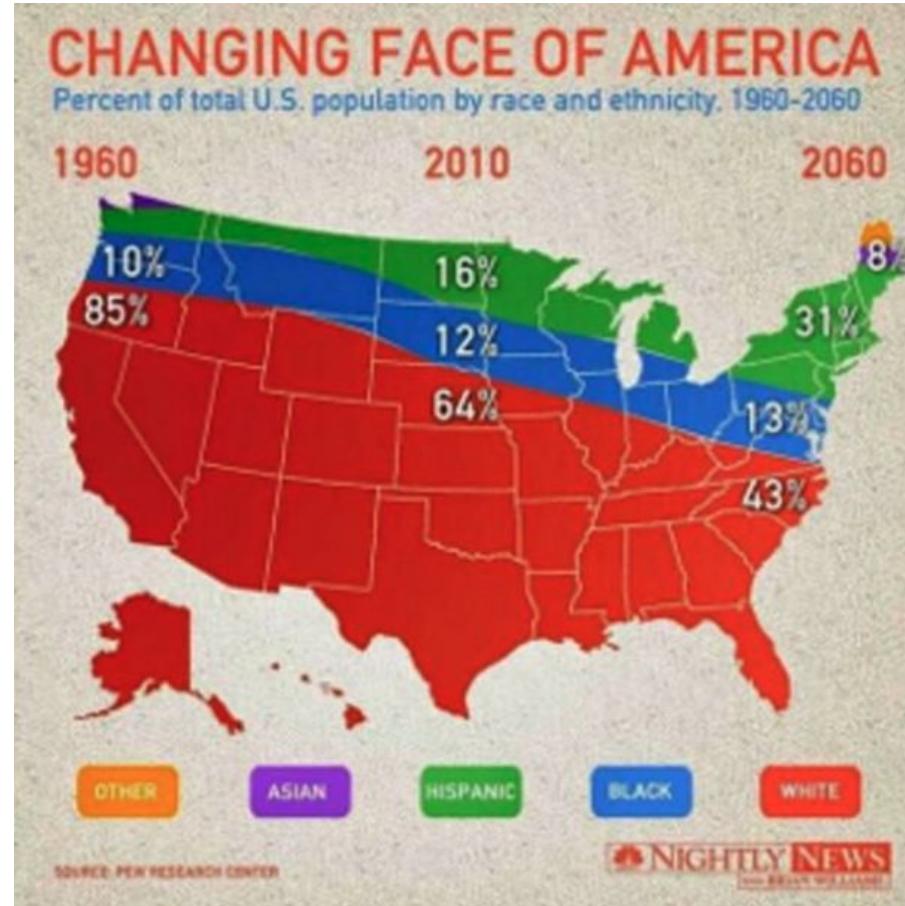


Source : [sixième rapport du Giec](#)

franceinfo:

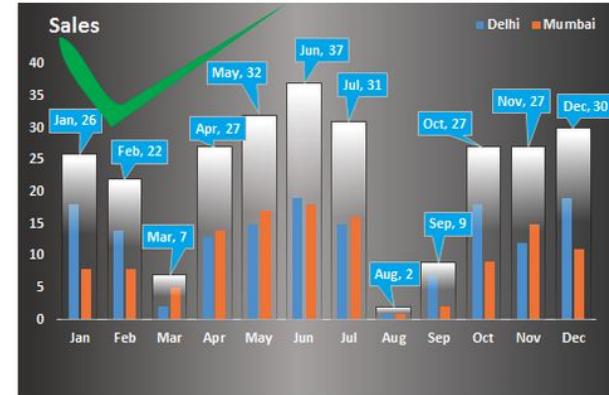
4 - CE QU'IL FAUT ÉVITER

BONUS: DANS LES MÉDIAS



4 - CE QU'IL FAUT ÉVITER

BONUS



Internet regorge de conseils douteux

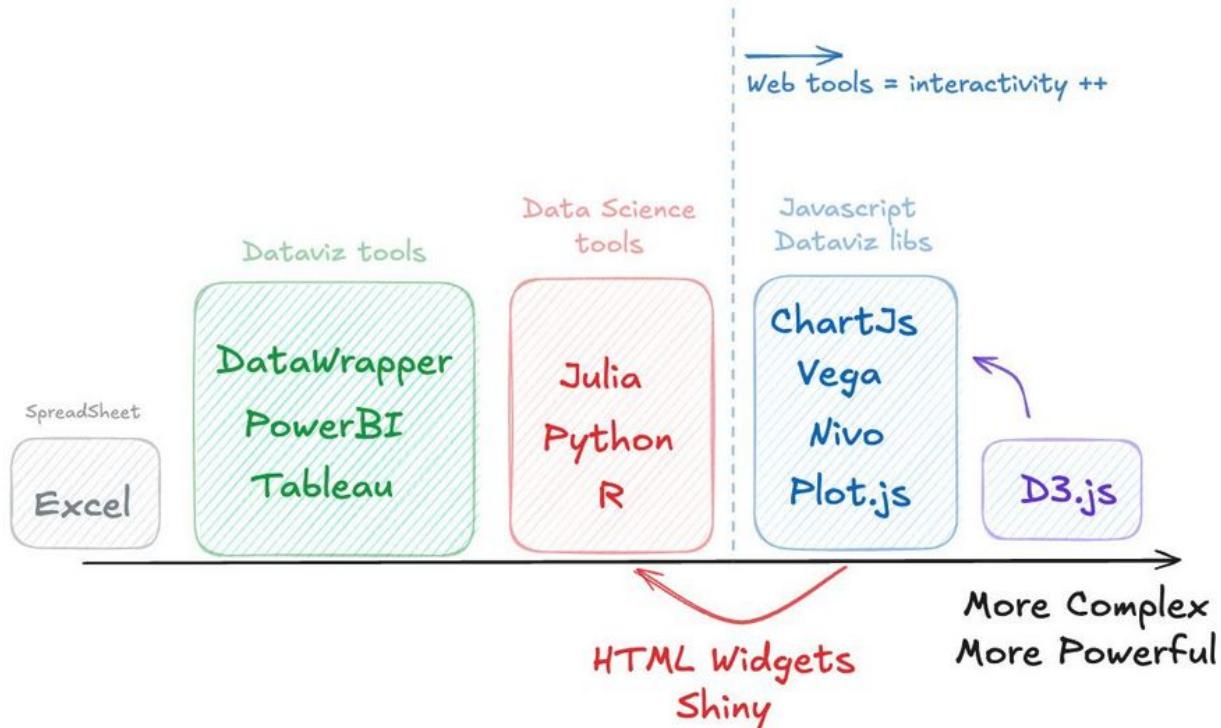
05

COMMENT FAIRE

5 - COMMENT FAIRE OUTILS

Éléments à prendre en compte :

- Facilité d'usage
- Intégration avec d'autres outils (connexion base de données, etc)
- Reproductibilité → va falloir coder 😞
- Prix et Licenses
- Interactivité ou non ?



Source : Yan Holtz

5 - COMMENT FAIRE

PROS & CONS

No-code

- Simplicité d'usage
- Prototypage rapide
- Intégration avec d'autres outils
- Payant
- Dépendant d'une entité externe (sécurité, confidentialité, etc)
- Reproductibilité

Python ou R

- Gratuit (et pour toujours!)
- Reproductible
- Communauté (aide, extensions, etc)
- Large éventail de possibilités
- Intégration avec d'autres outils
- Peut devenir complexe dans certains cas
- Possibilités larges mais pas infinies

JavaScript

- Tout ce qui est dans Python ou R +
- Tout (oui, tout) est possible
- Contrôle sur tout
- Très complexe → requiert des connaissances importantes du web et développement logiciel en général → développeur + data analyst + chercheur/journaliste

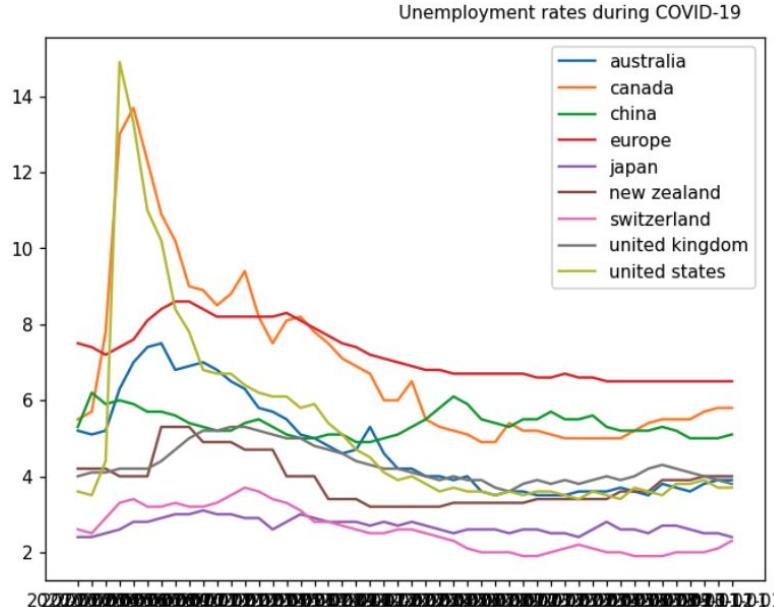
CAS D'USAGE

Quarto + R + Github = ❤

On va créer un site avec des graphiques interactifs et customisés,
et le mettre en ligne gratuitement !

5 - COMMENT FAIRE
**CAS
D'USAGE**

On prend un graphique et
on le rend cool 😎



06

LA SUITE ?

LA SUITE ?

Pratiquer

- Prendre le temps
- Lire des articles
- Partager et récupérer du feedback !



Ressources

- R/Python Graph Gallery
- Data To Viz

Personnes à suivre

- Yan Holtz
- Nicola Rennie
- Ansgar Wolsing
- Cara Thompson
- Georgios Karamanis
- Morgan Depenbusch

Chat

- Parler avec des gens du milieu !

Merci !

Joseph Barbier

barbierjoseph.com

github.com/JosephBARBIERDARNAL

Slides (pdf et pptx) : github.com/JosephBARBIERDARNAL/vizprep