

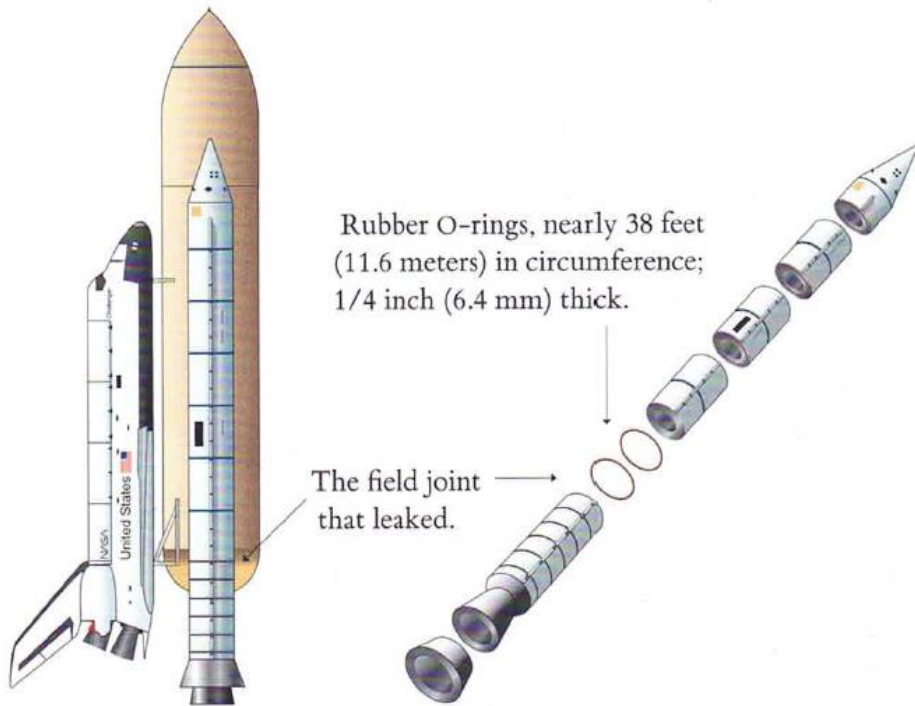
# Mieux faire parler les données

**Eric Mauvière, Icem7**

Toulouse, 1<sup>er</sup> octobre 2021

# Le contexte de la décision de lancer la navette Challenger, qui explose peu après son décollage, le 28 janvier 1986

Un joint d'assemblage d'un des boosters latéraux connaît une rupture, ayant perdu de son élasticité à cause du froid régnant la nuit précédant le lancement.

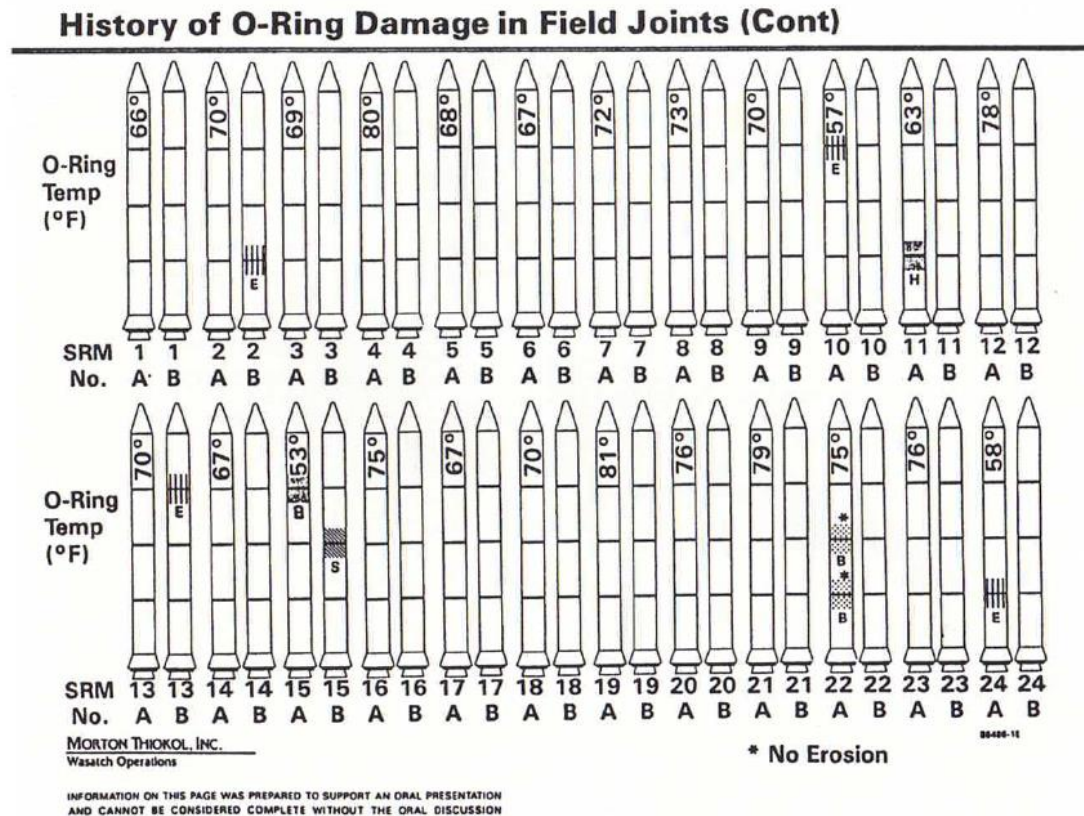


Origine : Edward Tufte – Visual explanations

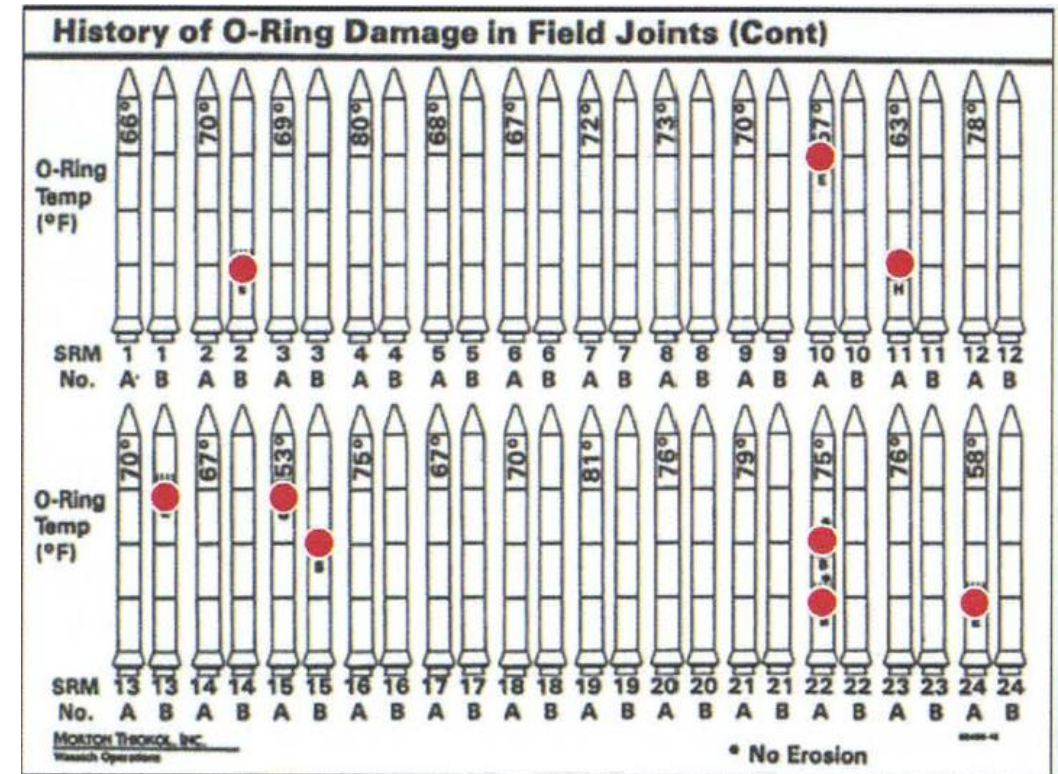
*Une fuite sur le booster droit projette un faisceau de flammes vers le réservoir principal*

# Des ingénieurs la veille au soir redoutaient la catastrophe...

Au fil d'une vingtaine de lancements, et au vu de certains **défauts sur les joints relevés ex-post**, ils avaient le sentiment que **plus il faisait froid, plus il y avait risque de défaillance**.



Graphique établi après la catastrophe par le concepteur des boosters



Même graphique avec ajout de pastilles rouges pour souligner les défauts



# ...mais ils ont malheureusement échoué à convaincre la Nasa

Tufte met en cause la mauvaise qualité des visuels présentés avant et même après la catastrophe : notations incohérentes, données partielles, symboles peu clairs, tri par date de lancement...

Un tri selon la température au sol aurait mieux mis en évidence l'accumulation de problèmes plus il faisait froid.

Mais ce n'était pas encore suffisant, SRM22 à droite apparaissant comme contre-exemple pour la Nasa...

History of O-Ring Damage in Field Joints (Cont)

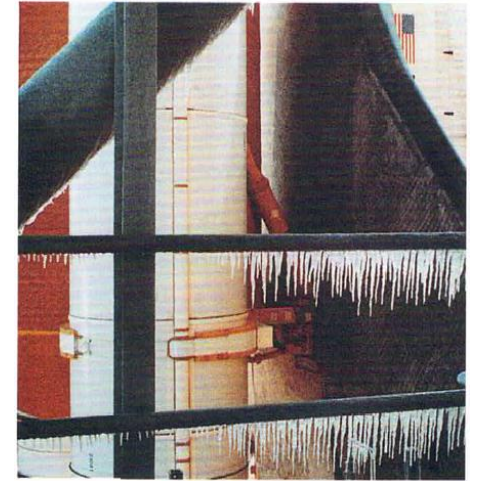
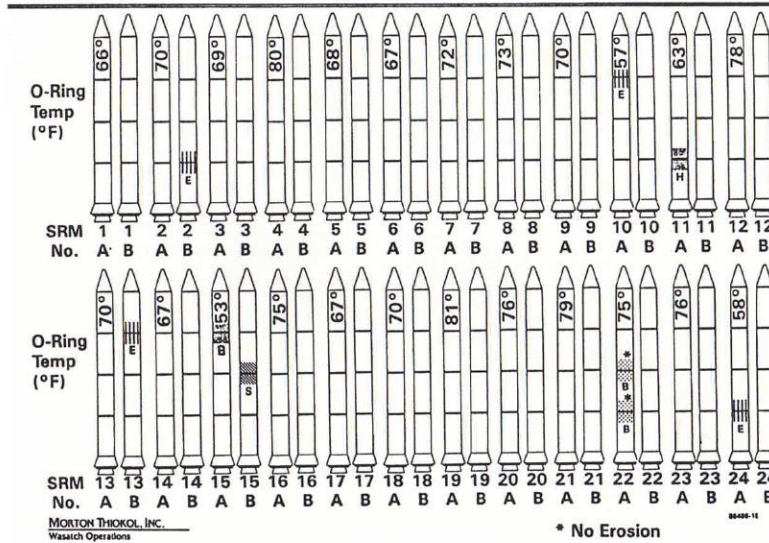
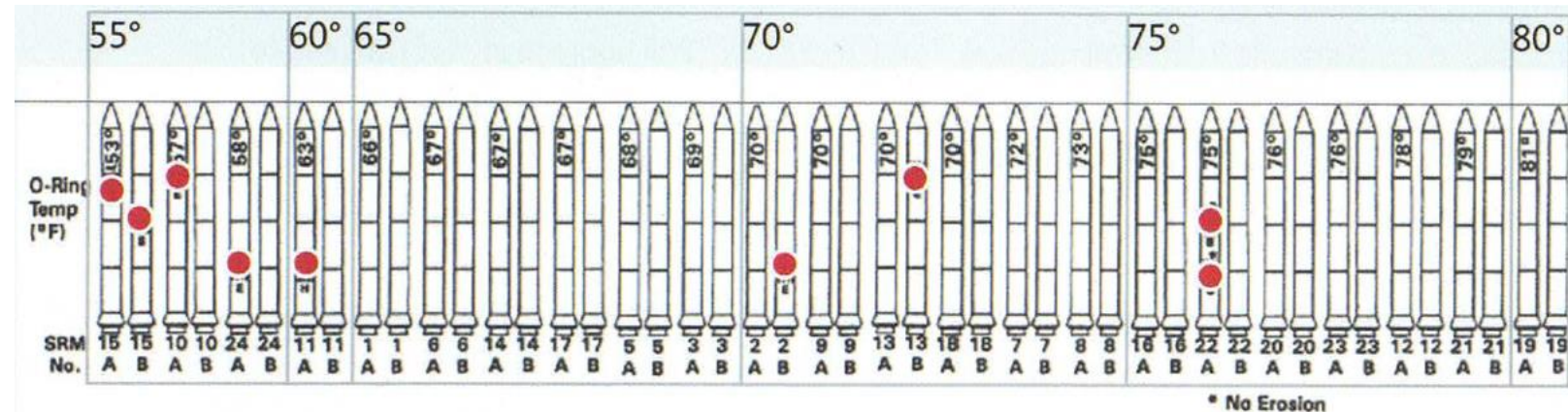


Photo du pas de tir avant le lancement, confirmant les températures négatives du moment



Origine : Edward Tufte - Visual explanations. 32 °F = 0 °C, 75° F = 24 °C

# Deux niveaux de gravité distincts dans les défauts observés

La commission d'enquête a distingué parmi les défauts constatés auparavant les simples « blow-by » (soufflage limité de gaz brûlés) des « erosions » (joints nettement détériorés).

Les « erosions » sont bien plus graves, car irréversibles.

Tufte établit le tableau complet ci-contre et construit un index pondéré de gravité. Le tableau est trié par températures croissantes.

La concentration des dommages en haut de tableau apparaît plus évidente encore.

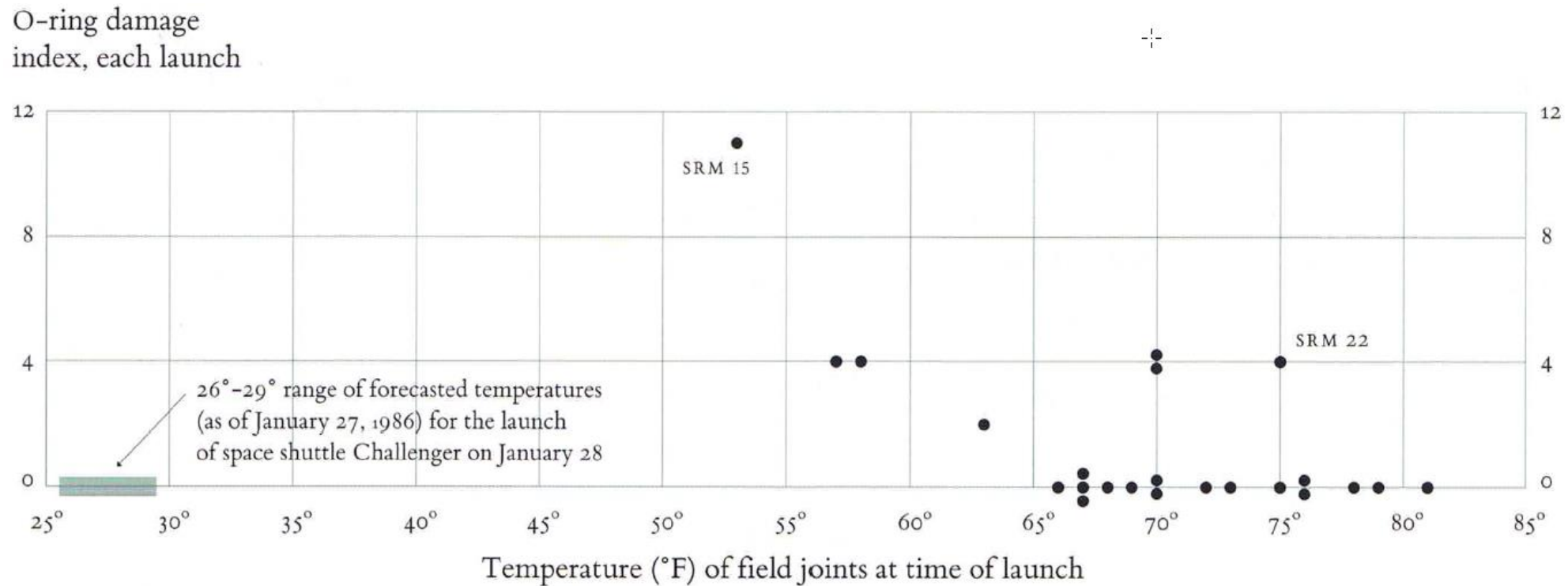
Flight	Date	Temperature °F	Erosion incidents	Blow-by incidents	Damage index	Comments
51-C	01.24.85	53°	3	2	11	Most erosion any flight; blow-by; back-up rings heated.
41-B	02.03.84	57°	1		4	Deep, extensive erosion.
61-C	01.12.86	58°	1		4	O-ring erosion on launch two weeks before Challenger.
41-C	04.06.84	63°	1		2	O-rings showed signs of heating, but no damage.
1	04.12.81	66°			0	Coollest (66°) launch without O-ring problems.
6	04.04.83	67°			0	
51-A	11.08.84	67°			0	
51-D	04.12.85	67°			0	
5	11.11.82	68°			0	
3	03.22.82	69°			0	
2	11.12.81	70°	1		4	Extent of erosion not fully known.
9	11.28.83	70°			0	
41-D	08.30.84	70°	1		4	
51-G	06.17.85	70°			0	
7	06.18.83	72°			0	
8	08.30.83	73°			0	
51-B	04.29.85	75°			0	
61-A	10.30.85	75°		2	4	No erosion. Soot found behind two primary O-rings.
51-I	08.27.85	76°			0	
61-B	11.26.85	76°			0	
41-G	10.05.84	78°			0	
51-J	10.03.85	79°			0	
	06.27.82	80°			?	O-ring condition unknown; rocket casing lost at sea.
51-F	07.29.85	81°			0	

Origine : Edward Tufte – Visual explanations

# Le graphique qui aurait pu convaincre de reporter le lancement

Tufte propose enfin ce « scatterplot » épuré, qui traduit le tableau précédent et y figure **la forte excentricité à gauche de la situation du jour du lancement**, avec une température au sol bien plus critique que celles observées jusqu'alors :

**- 3 °C pour Challenger contre 12 °C**, lors de la situation la + froide connue antérieurement (SRM 15).





# Mais Tufte reconstruit en partie l'histoire et fait preuve d'excès dans ses jugements

## 30 Years After Explosion, Challenger Engineer Still Blames Himself

January 28, 2016 By [Howard Berkes](#)



(Left) Bob Ebeling in his home in Brigham City, Utah. (Right) The Challenger lifts off on Jan. 28, 1986, from a launchpad at Kennedy Space Center, 73 seconds before an explosion killed its crew of seven.

((Left) Howard Berkes/NPR; (Right) Bob Pearson/AFP/Getty Images)

Source : Wbur, 2016

*Ebeling retired soon after Challenger. He suffered deep depression and has never been able to lift the burden of guilt.*

*In 1986, as he watched that haunting image again on a television screen, he said, "I could have done more. I should have done more."*

Tufte a été critiqué par un autre des ingénieurs impliqués ce soir-là (René Boisjoly), considérant qu'il était difficile de faire bien mieux avec les données partielles dont ils disposaient alors.

Ce sont surtout les décideurs finaux (Nasa) qui ont fait valoir d'autres priorités, stratégiques et politiques.

La démonstration de Tufte n'en reste pas moins lumineuse quant aux vertus d'une bonne représentation.

# Jacques Bertin

« Si pour obtenir une réponse  
**correcte et complète à une question donnée,**  
et toutes choses égales,  
une construction requiert  
**un temps d'observation plus court**  
qu'une autre construction,  
on dira **qu'elle est plus efficace** pour cette question ».

*Sémiologie graphique, 1967*

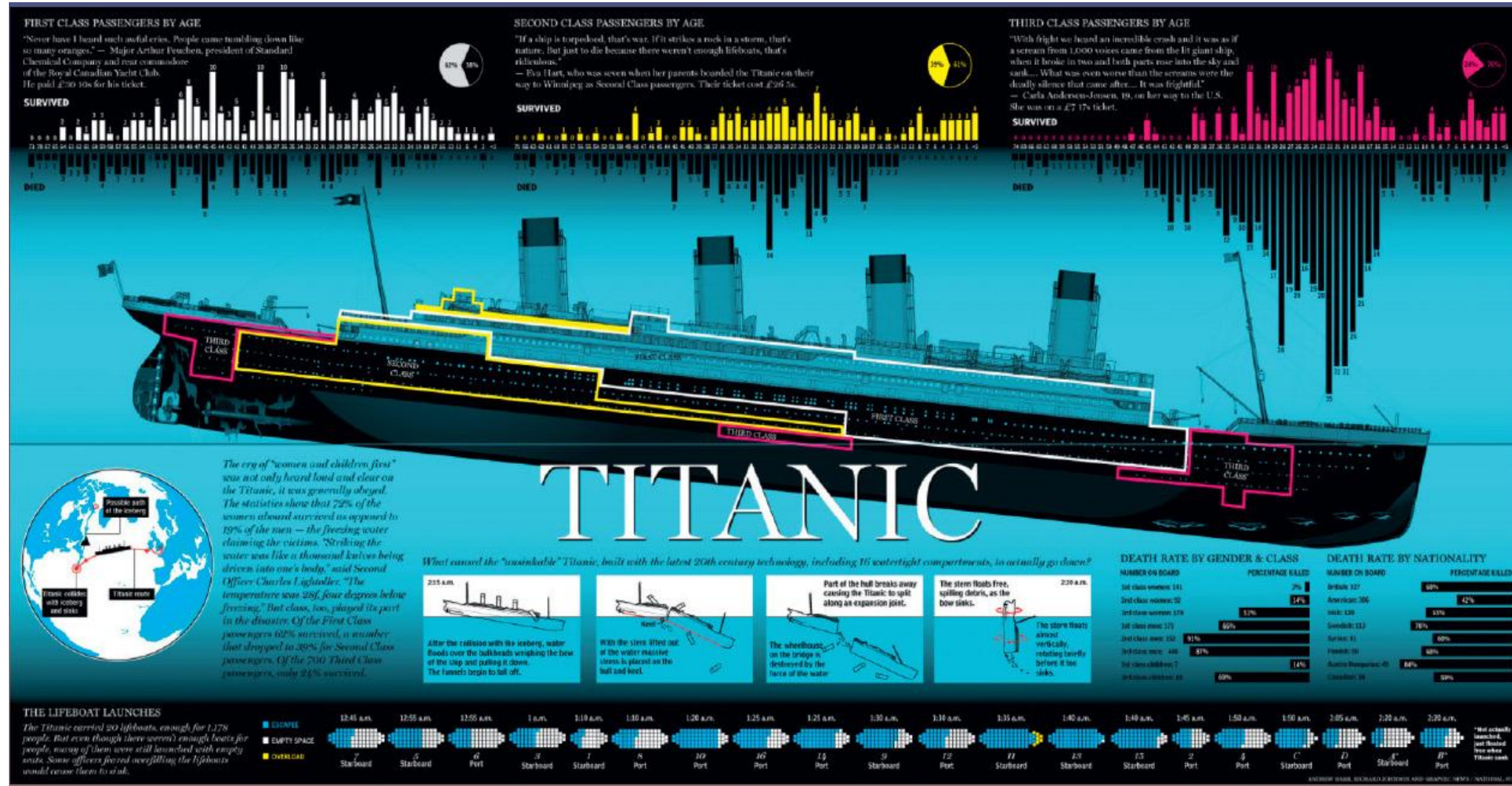


# Edward Tufte

« Un graphique d'excellence  
est celui qui donne au lecteur  
le plus grand nombre d'informations  
dans **le temps le plus court**  
dans **l'espace le plus petit**  
et avec **le moins d'encre** possible ».

*The Visual Display of Quantitative Information, 1983*

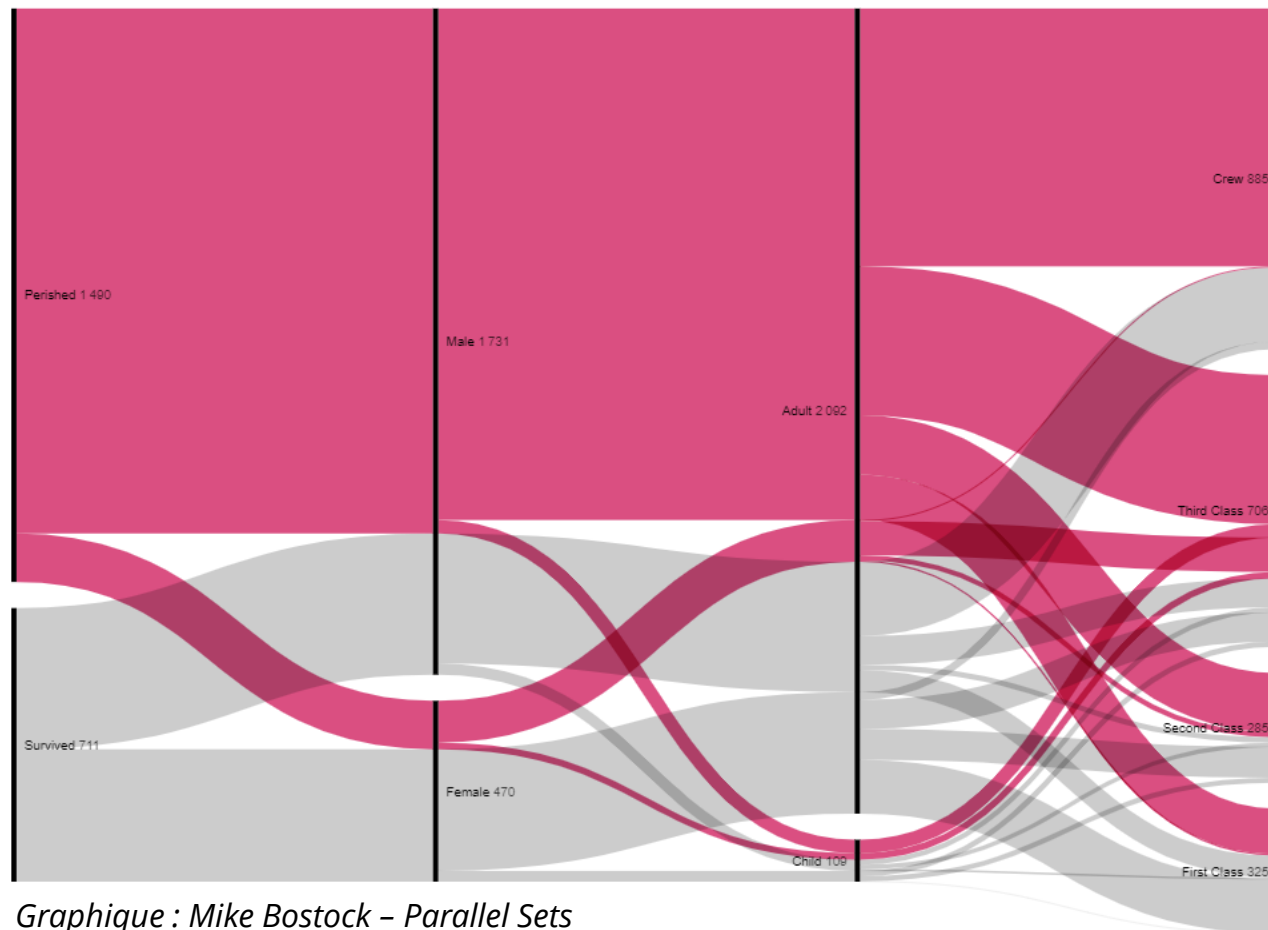
# Comment bien synthétiser les données du naufrage du Titanic, le 14 avril 1912, avec 2 224 personnes à bord



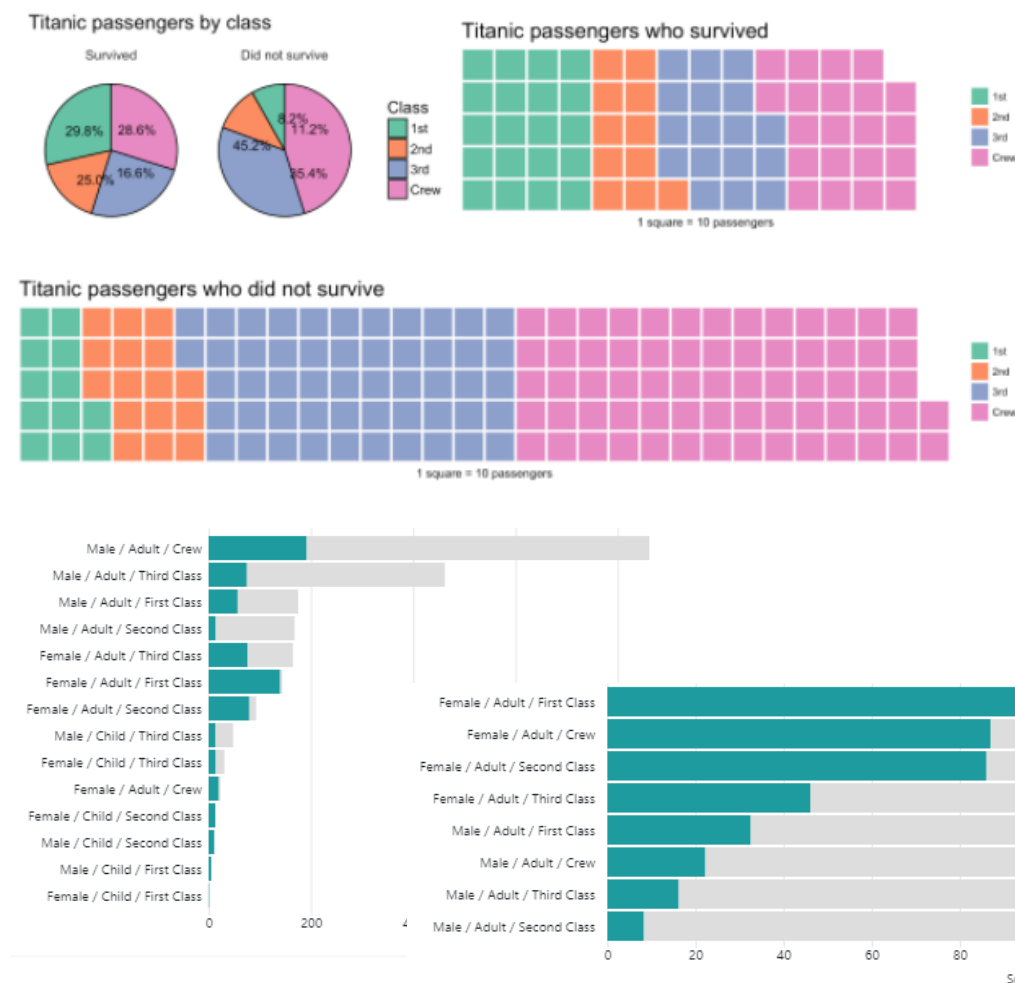
Infographie : Andrew Barr and Richard Johnson, [National Post](#)

# Un jeu de données détaillées très visualisé depuis 1995...

Mis en circulation publique par Robert G Dawson, support de nombreux cours et exercices, Il donne lieu à des graphiques plus ou moins faciles à lire...



Graphique : Mike Bostock – Parallel Sets



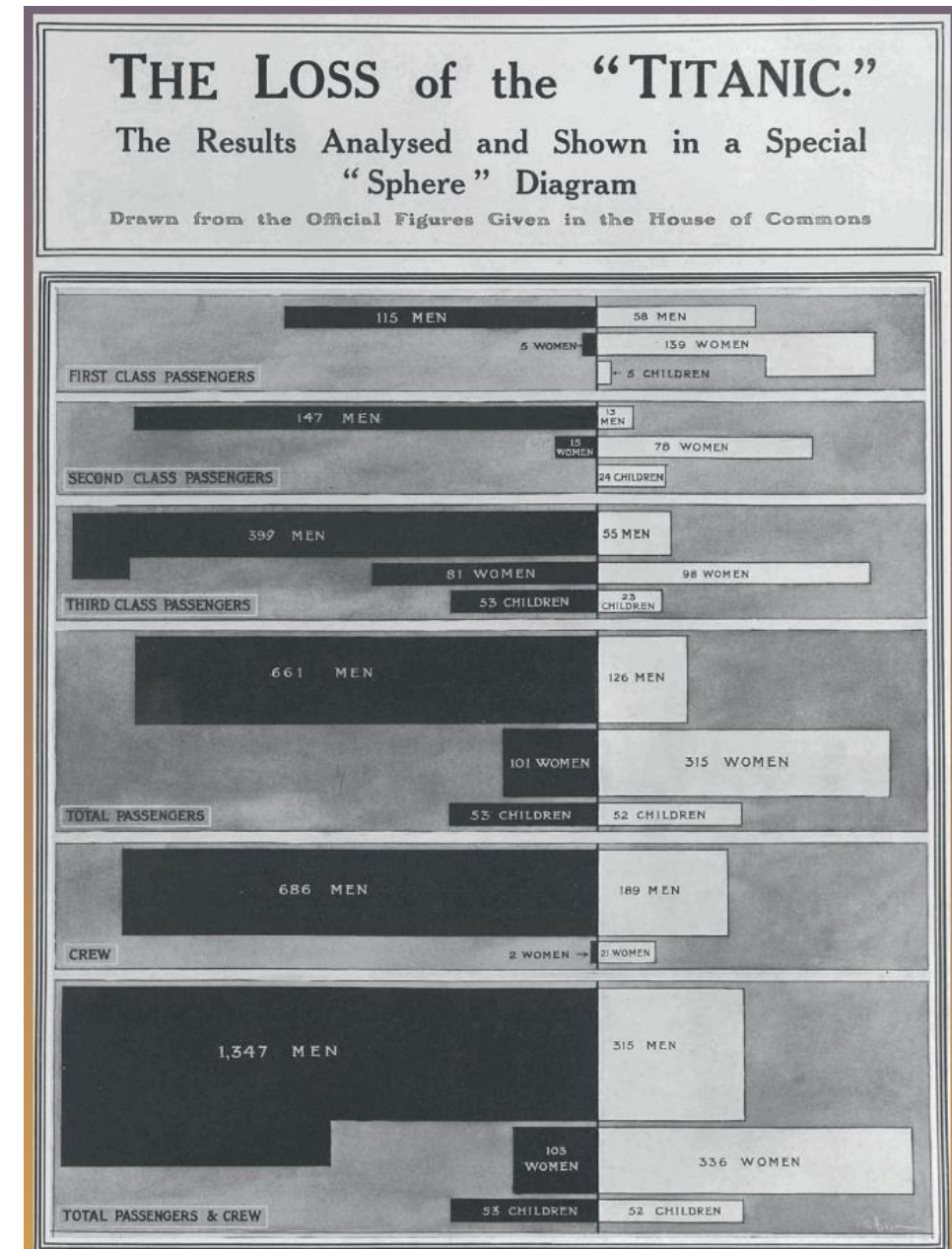


# ...mais 3 semaines après le naufrage, les données étaient déjà mises en images

Cette composition journalistique, exhumée en 2019 par Michael Friendly, utilise un langage étonnamment moderne :

- Collection de « petits graphiques »,
- Mosaïque de rectangles de surface proportionnelle, échelle cohérente dans toute la composition,
- Typographie nette, renseignant les barres.

Et l'on constate que l'usage de la couleur n'est ici nullement nécessaire, deux niveaux de gris suffisent.

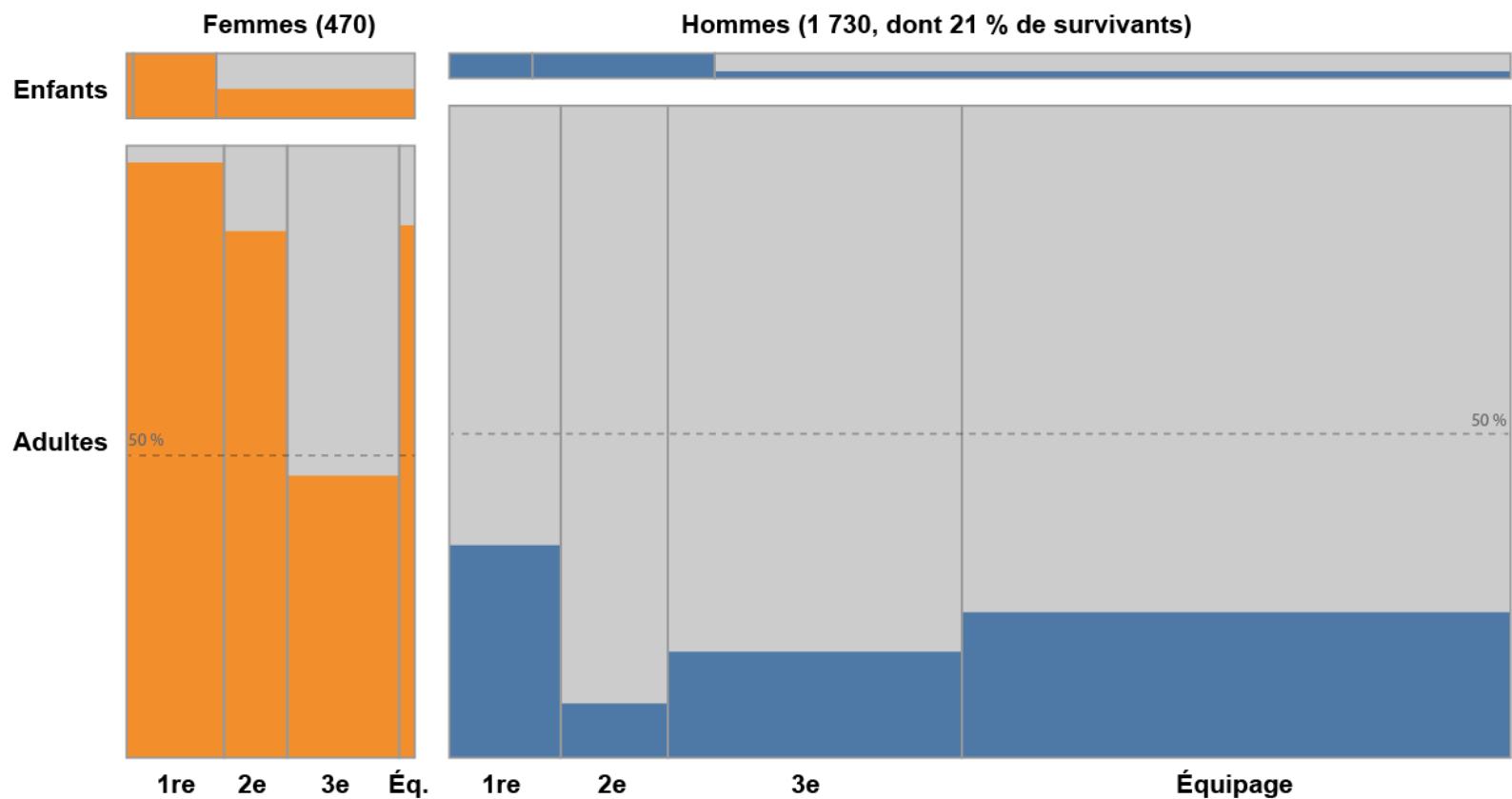


Source : G. Bron - The Sphere - 4 mai 1912

# Une représentation plus moderne et parlante encore

**Titanic : les trois quarts des femmes furent sauvées**

Très peu d'hommes passagers en 2e classe ont survécu



Source : Robert Dawson

**6 enseignements** à imprimer sans (trop) d'effort, avec ce graphique :

- Peu d'enfants étaient à bord,
- Une grande majorité de femmes furent sauvées,
- La plupart des hommes n'ont pas survécu,
- Il y avait trois fois plus d'hommes que de femmes,
- L'équipage représentait plus d'1/3 des personnes à bord,
- Les hommes en 2<sup>e</sup> classe avaient le moins de chance de s'en sortir.

Inspiration du graphique : Bruce Gabrielle – Storytelling with graphs

# Jacques Bertin

« Un graphique  
ne doit pas seulement montrer les feuilles de l'arbre.  
Il doit aussi **montrer les branches**  
**et l'arbre tout entier.**  
L'œil peut alors **aller du détail à l'ensemble**  
et découvrir à la fois  
**la structure générale et ses exceptions ».**

*Sémiologie graphique, 1967*



# Quelques principes généraux de construction d'une dataviz expressive

Rassembler patiemment et ardemment toutes les données pertinentes

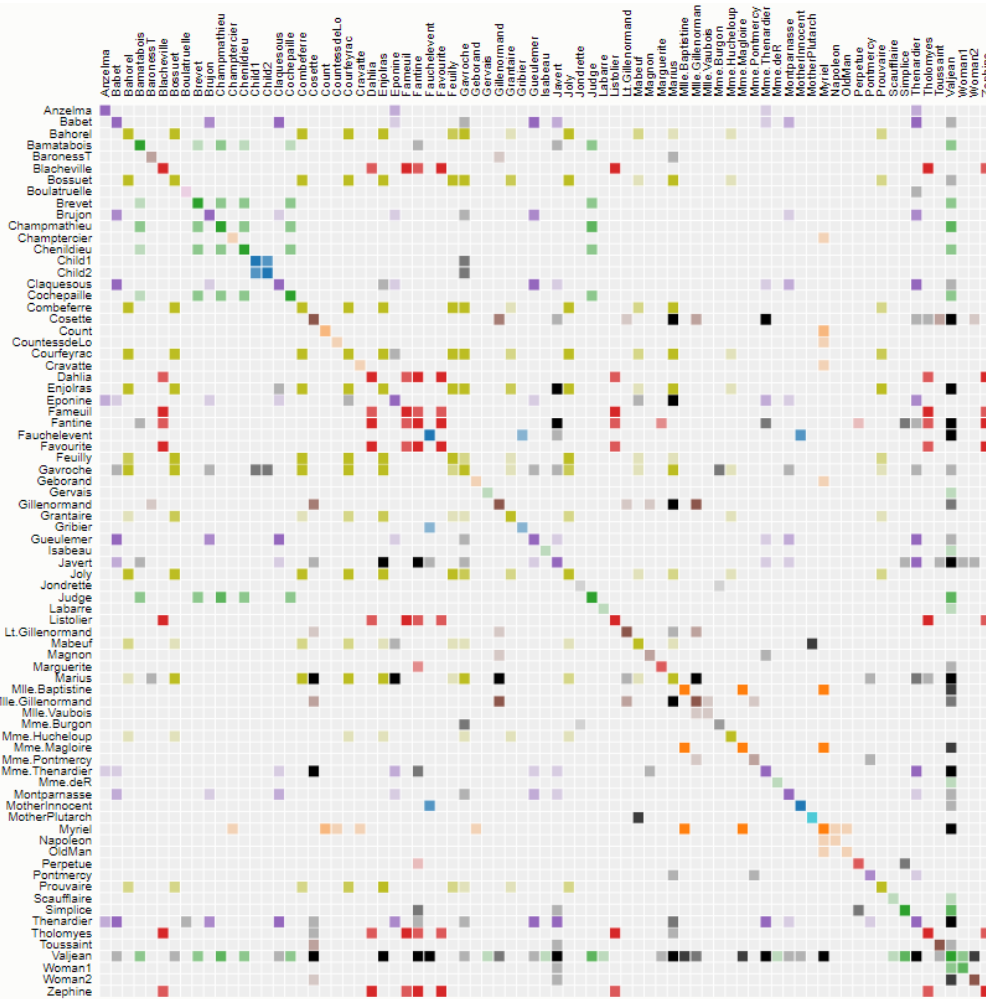
Identifier le ou les tris les plus significants

Dégager différents niveaux de lecture et les hiérarchiser

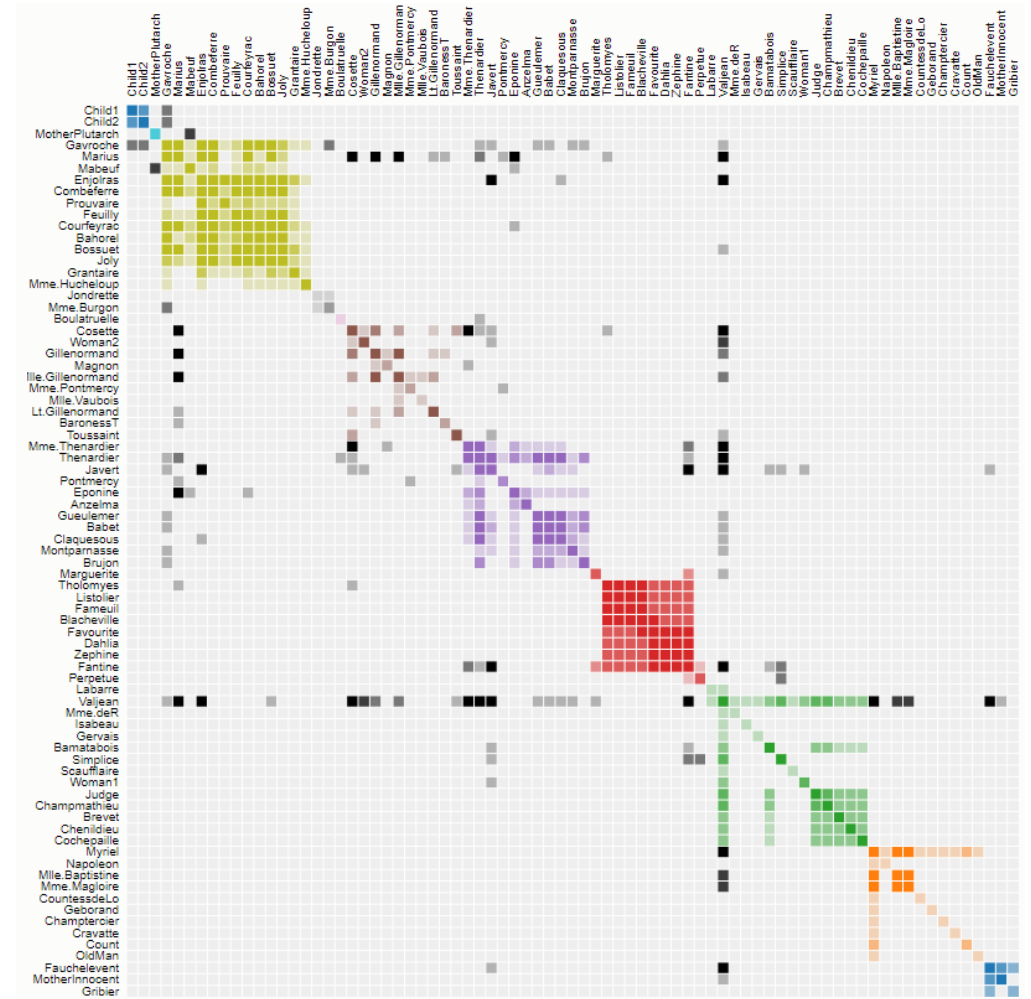
Préférer de petits graphiques bien mis en grille (et triés)  
à un graphique complexe

Utiliser la couleur de façon modérée et maîtrisée

# Les personnages des Misérables et leurs rencontres dans un même chapitre de l'ouvrage



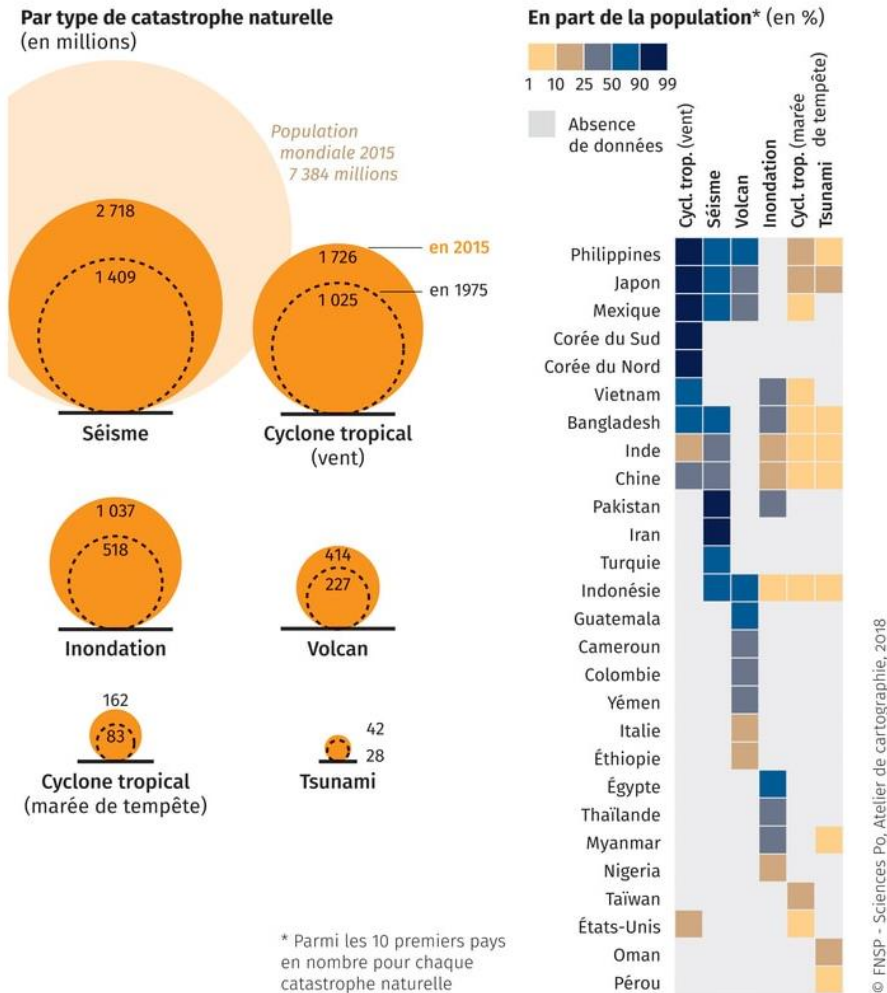
De l'ordre  
platement  
alphabétique  
à un tri  
intelligent



Graphique et animation : Mike Bostock

# De l'importance du tri et de la hiérarchisation de l'information

Population potentiellement exposée aux catastrophes naturelles, 2015



L'Atelier de cartographie de Sciences Po est un des modèles de mise en œuvre élégante et subtile des bonnes pratiques sémiologiques.

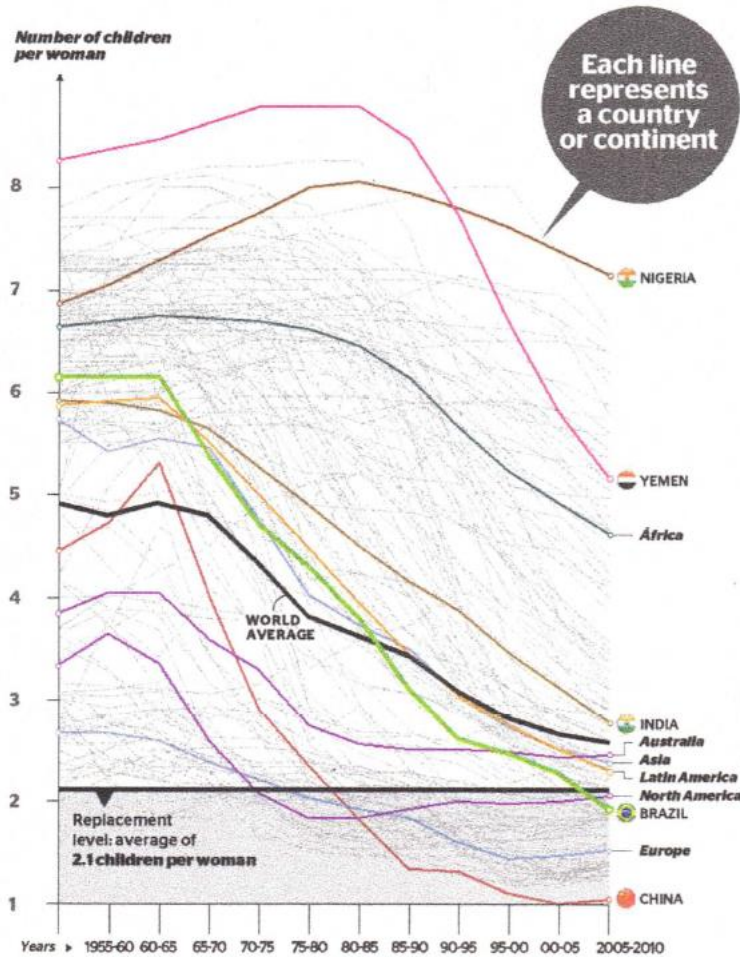
Ici un assemblage de graphiques élémentaires, dans une grille équilibrée.

Le graphique « heatmap » à droite trie les pays par intensité d'exposition décroissante.

L'auteur a résisté à la tentation de colorier différemment les ronds à gauche (par exemple par type de catastrophe).



# Une collection triée de petits graphiques vaut parfois mieux qu'un amas de spaghettis



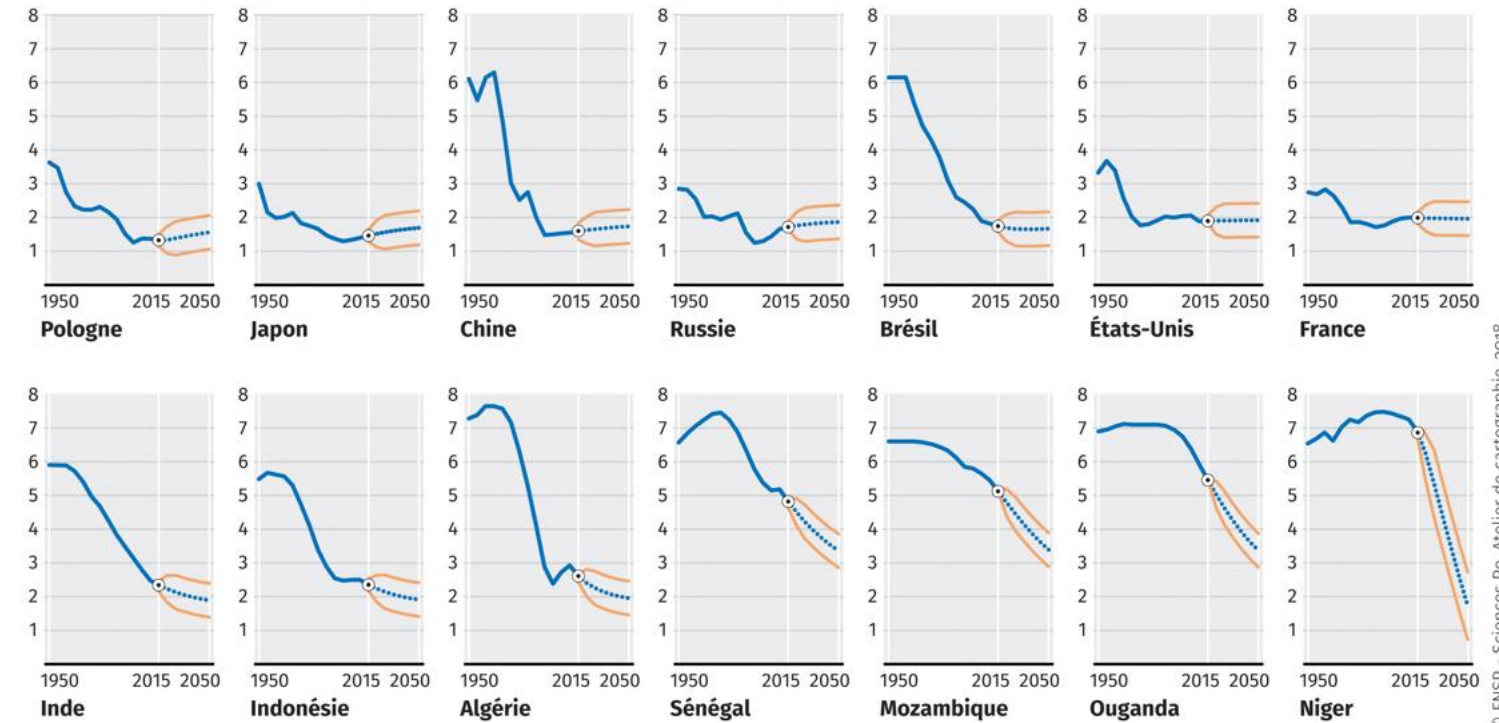
Graphique : Epoca – Alberto Cairo

## Taux de fécondité, 1950-2050

### Nombre d'enfants par femme

Les courbes sont triées par ordre croissant selon la valeur de 2015.

Les pointillés montrent les prévisions moyennes des Nations unies, l'orange représente les prévisions hautes et basses.



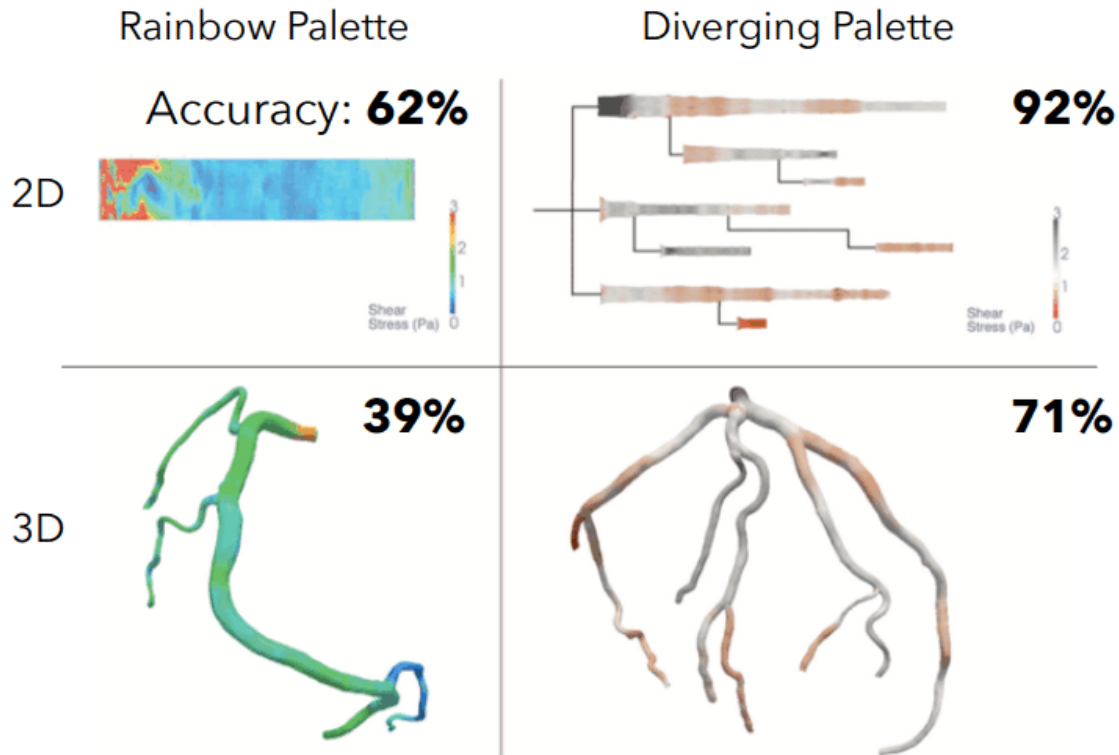
Source : Nations unies, division Population, *World Population Prospects : The 2015 Revision*, [www.un.org](http://www.un.org)

Graphique : Atelier de cartographie – Sciences Po – Espace mondial, l'Atlas

© FNSP - Sciences Po, Atelier de cartographie, 2018

# Du bon usage de la couleur pour une décision rapide et sûre

## Artery Visualization [Borkin et al 11]



Avec l'imagerie médicale, la datavisualisation est une affaire parfois de vie ou de mort, quand il s'agit pour des médecins d'être précis et sûrs dans leurs diagnostics.

Cette célèbre étude porte sur les probabilités de se tromper ou pas, dans un temps de test limité, dans une analyse d'imagerie cardiaque. Les chiffres traduisent la part des diagnostics corrects.

Source : Michelle Borkin et alii  
*Evaluation of Artery Visualizations for Heart Disease Diagnosis*

# Du bon usage de titres et annotations éclairants

Summary data from respondent countries indicate that median government expenditure on sanitation and drinking water is one-third of that for health and one-sixth of that for education (Table 3.4).

TABLE 3.4 Government expenditure on health, education and WASH

Country	Government expenditure on health (% of GDP)	Government expenditure on education (% of GDP)	Government expenditure on WASH (% of GDP)
Algeria	1.2	3.8	0.4
Angola	4.4	4.7	0.5
Benin	1.9	3.8	0.8
Burkina Faso	4.8	—	1.2
Burundi	1.3	—	0.7
Cameroon	2.3	4.7	0.7
Chad	—	—	—
Cote d'Ivoire	2.1	6.9	1.1
Egypt	8.4	—	2.6
Guinea	3.7	3.2	0.4
Kenya	1.7	4.7	0.8
Madagascar	6.1	3.9	0.8
Mali	3.7	4.1	0.8
Morocco	1.4	—	0.7
Mozambique	1.1	3.2	0.3
Niger	6.4	16.9	3.5
Nigeria	2.2	5.5	0.7

2010 data  
2010 data  
Not all countries reported contributions from regional and local governments  
Source: 2011 GLAAS country survey; UNICEF 2010; WFP 2010; World Bank 2010

One-third of countries indicate that revenues cover less than 80% of the cost of providing essential services (Figure 3.11).



Government expenditure (from taxes and transfers) on sanitation and drinking-water ranged from 0.37% to 3.5% of GDP (Figure 3.2).

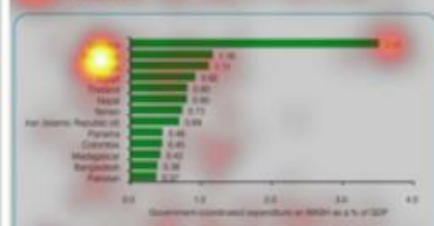


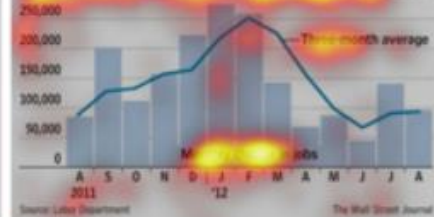
FIGURE 3.2 Public spending from funds obtained through domestic taxes and external transfers on sanitation and drinking-water as a percentage of GDP (2010 data)

Note: Not all countries reported contributions from regional and local governments (e.g., Egypt, Kenya and Jordan)

Source: 2011 GLAAS country survey; World Bank 2010

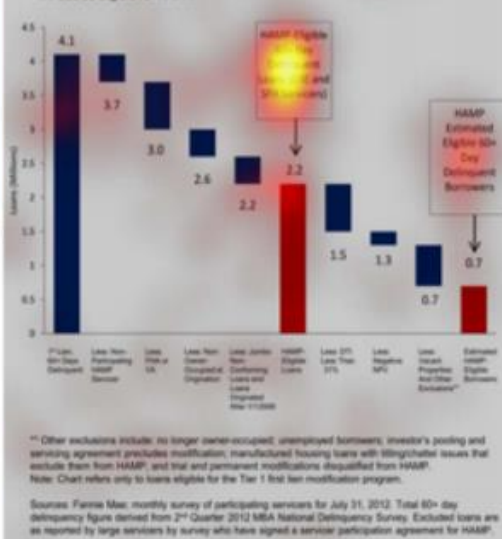
Hobbled Hiring

World Bank's new hiring program (COP) is designed to help hiring job growth.



WASH and Sanitation for Homeowners

Not all 60+ day delinquent loans are eligible for HAMP. Other characteristics may preclude homeowner eligibility. Based on the estimates, of the 4.1 million homeowners who are currently 60+ days delinquent, about 700,000 homeowners are eligible for HAMP. As this represents a point-in-time snapshot of the delinquency population and estimated HAMP eligibility, we expect that as homeowners become seriously delinquent between now and the end of 2013, some of those homeowners will also be eligible for HAMP.



Qu'est-ce qui rend une dataviz « mémorable » ?

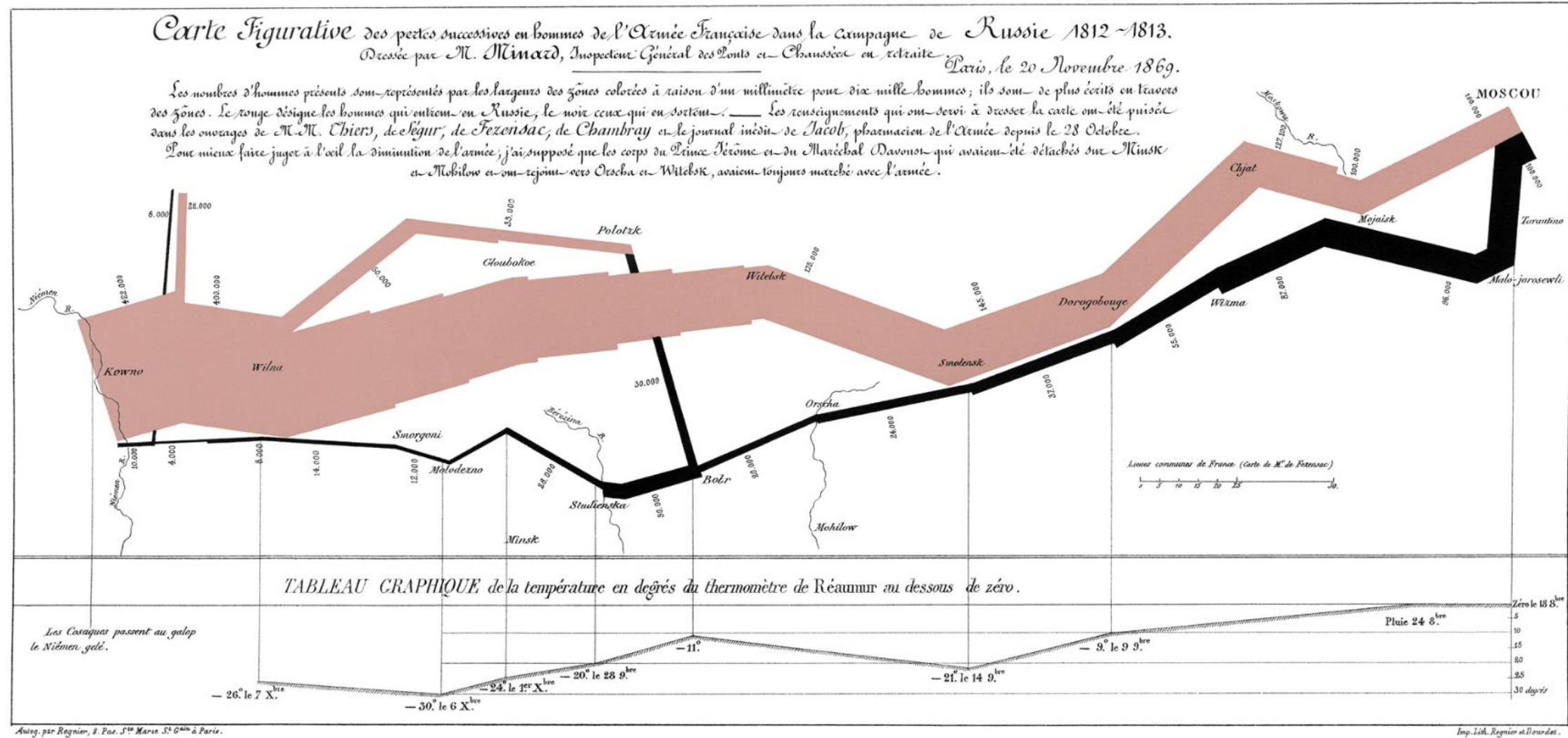
Cette expérience, nourrie de l'analyse des mouvements oculaires de lecteurs devant de nombreuses images souligne l'importance :

- en premier lieu, des titres, sous-titres et annotations ;
- ensuite de la présence d'icônes, de figures significatives (un dinosaure, un drapeau...)
- enfin, de la redondance de l'information (encodée de différentes façons).

Source : Michelle Borkin - Beyond Memorability: Visualization Recognition and Recall



# Pourquoi est-ce, sans doute, le meilleur graphique jamais construit ?



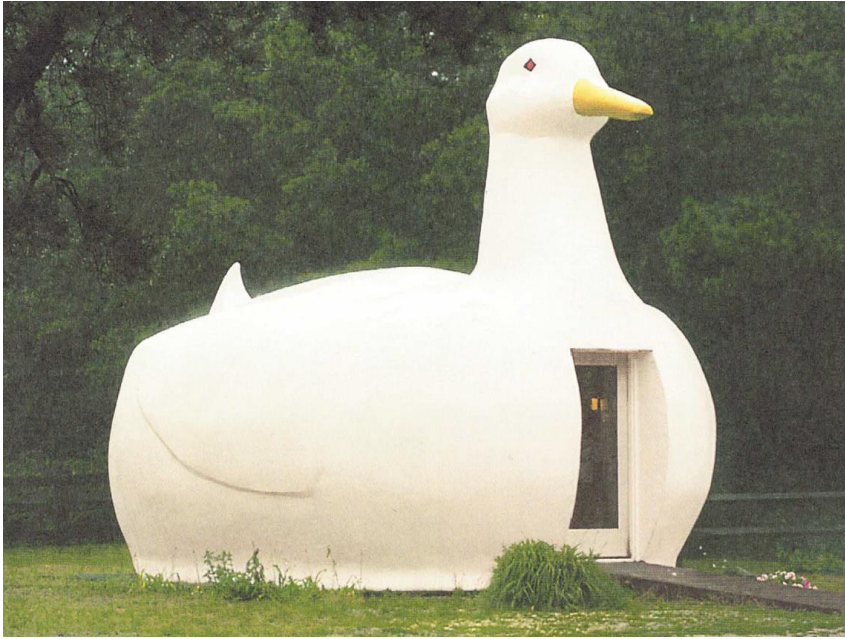
Source : Charles-Joseph Minard

# Alberto Cairo

« Start with a **strong focus**,  
**do as much research as you can**,  
Organize, summarize  
then deliver your conclusions  
in a **structured**  
and **visually appealing** manner ».

*The functional art, 2013*

# Le mirage du « Duck » : une pure exhibition de technicité, sans grande efficacité fonctionnelle ou expressive

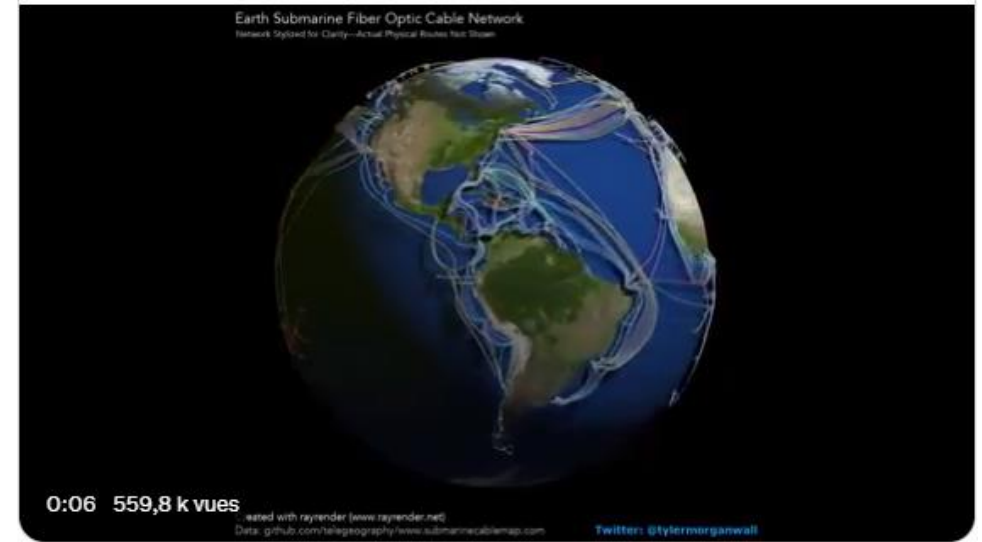


Origine : Edward Tufte – Visual display of quantitative information : Big Duck, Flanders

*Quand le designer semble davantage soucieux d'exhiber sa maîtrise d'une nouvelle technologie (et d'en attendre félicitations et buzz sur les réseaux) que de construire la représentation la plus efficace des données qu'il a à sa disposition.*

Tufte, dont l'ironie voire la méchanceté sont légendaires, n'est pas tendre avec ce qu'il dénomme les « ducks ».

Rayrender Github:  
[github.com/tylormorganwal...](https://github.com/tylormorganwall)  
[Afficher cette discussion](#)



1:36 PM · 23 sept. 2021 · Twitter Web App

19 Retweets 41 Tweets cités 101 J'aime

Source : [Twitter](#)



# Un assortiment choisi de Ducks récents...

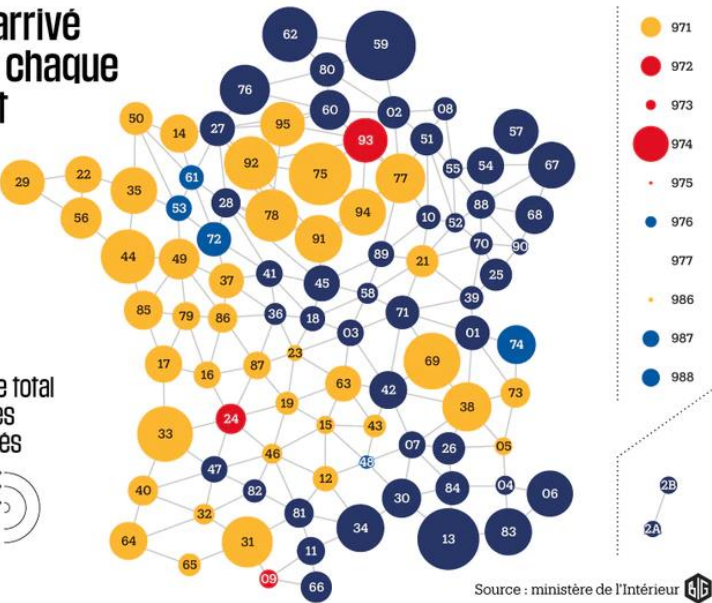


Le candidat arrivé en tête dans chaque département

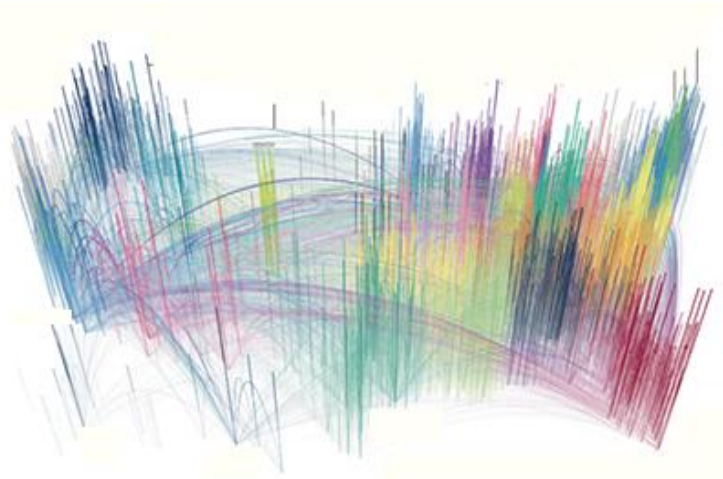
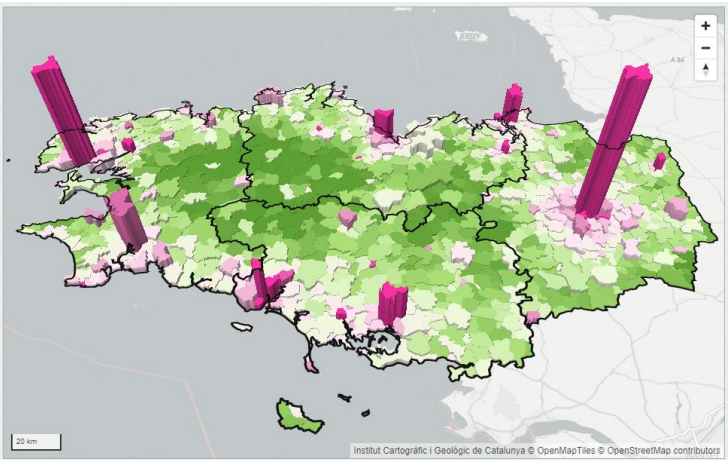
- Emmanuel MACRON
- Marine LE PEN
- François FILLON
- Jean-Luc MÉLENCHON

Nombre total de votes exprimés

1 353 802 >  
500 000 >  
26 834 >

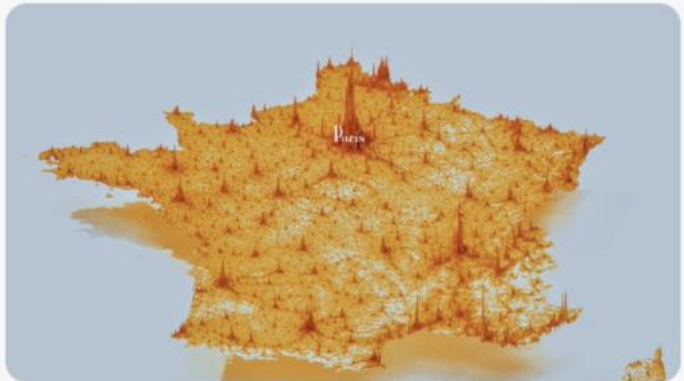


Source : ministère de l'Intérieur



Population density of France.

#rayshader adventures, an #rstats tale



5 27 204



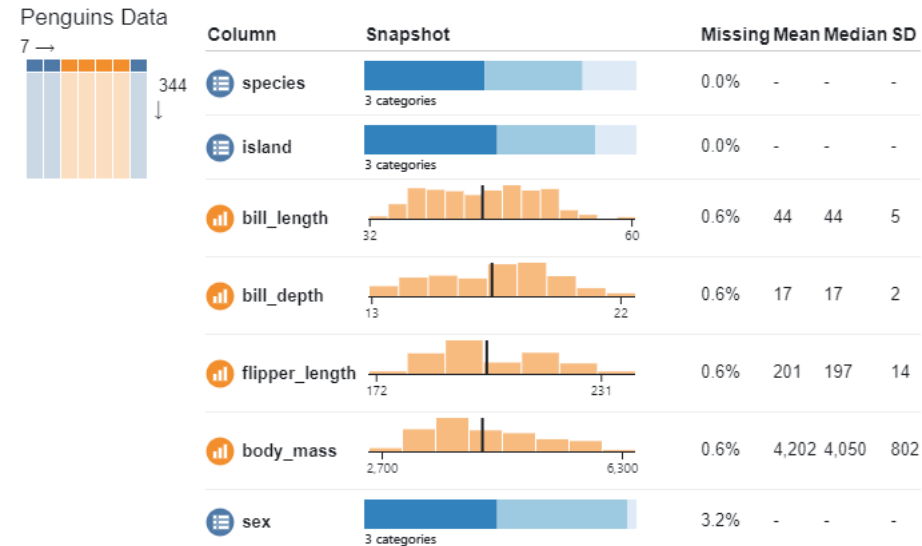
# Ma palette d'outils de travail en analyse et datavisualisation

Construire une dataviz requiert de bons outils de travail, et en pratique une chaine d'outils complémentaires ou gradués selon l'intensité des traitements et le besoin de documenter les étapes de construction.

## Analyse exploratoire

*Excel, R/tidyverse, Arquero*

*En ligne : Observable/Mike Freeman  
avec Data Wrangler et Table Summary*



## Graphiques

*Excel, Datawrapper, Plot/Vega lite*

## Peaufinage

*D3/CSS, Illustrator (ou Inkscape)*

# Ouvrages de référence (classement personnel)

1. Sémiologie graphique, Jacques Bertin, 1967
2. La graphique, Jacques Bertin, 1977
3. The Visual Display of Quantitative Information, Edward Tufte, 1983
4. Visual explanations, Edward Tufte, 1997
5. Beautiful evidence, Edward Tufte, 2006
6. Datavisualisation handbook, Juuso Koponen, Jonathan Hilden, 2019
7. Information Visualization, Colin Ware, 2021
8. Storytelling with Graphs, Bruce Gabrielle, 2018
9. Manuel de cartographie, Nicolas Lambert, Christine Zanin, 2017
10. Visualisation Analysis & Design, Tamara Munzner, 2015
11. Storytelling with data, Cole Nussbaumer Knaflitz, 2015
12. The functional art, Alberto Cairo, 2013

# Liens documentaires en relation avec cette présentation

[La Sémiologie graphique de Jacques Bertin a cinquante ans !, Gilles Palsky, 2017](#)

[Beyond Memorability: Visualization Recognition and Recall, Michelle Borkin & alii, 2015](#)

[Evaluation of Artery Visualizations for Heart Disease Diagnosis , Michelle Borkin & alii, 2011](#)

[Visualizing the Titanic disaster, Michael Friendly & alii, 2019](#)

[Catholic priest who saved Irish passengers on Titanic could become a saint](#)

[Titanic – Parallel Sets, Mike Bostock, 2019](#)

[Representation and Misrepresentation: Tufte and the Morton Thiokol Engineers on the Challenger, Wade Robinson & alii, 2002](#)

[30 Years After Explosion, Challenger Engineer Still Blames Himself, Wbur, 2016](#)

[Les Misérables co-occurrence, Mike Bostock, 2012](#)

[Table Summary, Mike Freeman, 2021](#)

[Espace Mondial, l'Atlas, Sciences Po, 2018](#)

[La « brutale éloquence » de Minard, icem7, 2020](#)