



Visualisation des données

PRO1036

Analyse de données scientifiques en R

Tim Bollé

03.11.2025

Visualisation des données

« *The purpose of visualization is insight, not pictures* »

- Ben Shneiderman

DataViz

Communiquer

- › Transmettre de l'information aux autres

Explorer

- › Détecter des régularités, relations ou anomalies

Raisonner

- › Comprendre et réfléchir sur les données
- › Tester des hypothèses par une mise en perspectives des observations/faits pertinents

Mémoriser

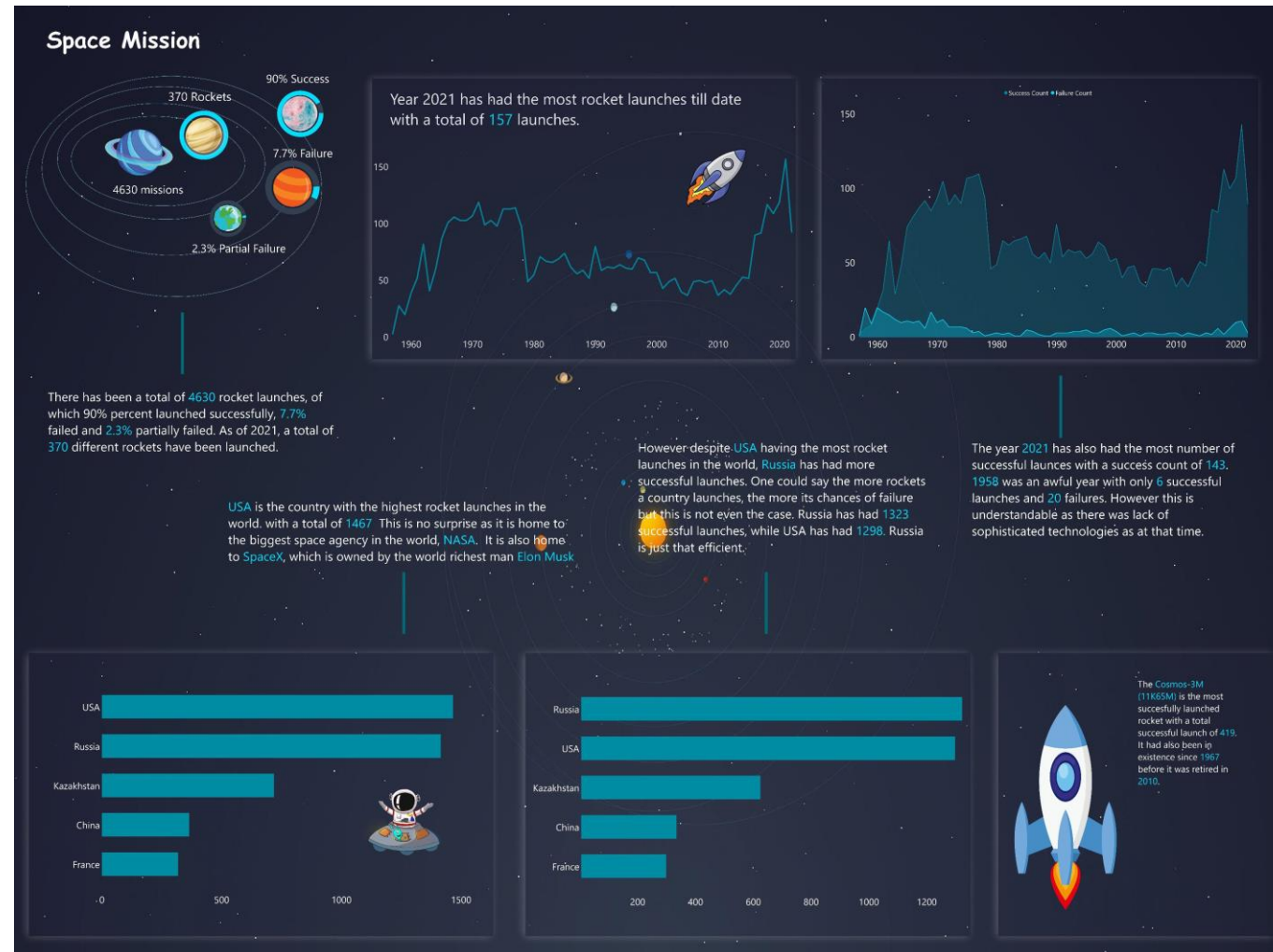
- › Regrouper des informations non homogènes sur un support visuel

Définition

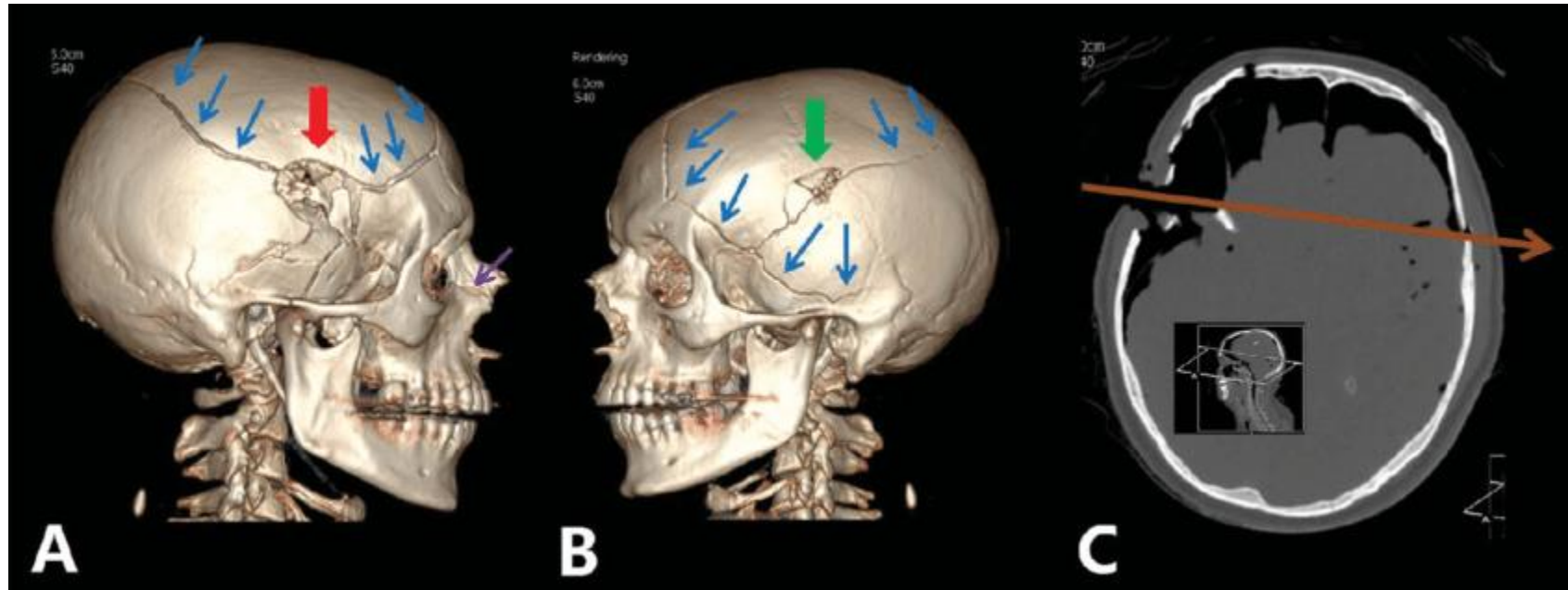
« La représentation graphique fait partie des systèmes de signes que l'homme a construits pour retenir, comprendre et communiquer les observations qui lui sont nécessaires. »

- Bertin 1967

Infographies



Imagerie médicale



Grabherr, Silke & Egger, Coraline & Vilarino, Raquel & Campana, Lorenzo & Jotterand, Melissa & Dedouit, Fabrice. (2017). Modern post-mortem imaging: an update on recent developments. *Forensic sciences research*. 2. 52-64. 10.1080/20961790.2017.1330738.

Perception



Perception vs cognition

Perception

- › Suite à un stimulus sensoriel
- › Vue, ouïe, odorat, touché et gout
- › Dépend de l'expérience et de ce que l'on attend

Cognition

- › Souvenir, savoir, penser, juger, résoudre des problèmes

Peuvent être en conflit:

Rouge Vert Bleu Orange Violet Rose Vert Rouge Bleu

Perception



TOBLERONE

Perception



Perception

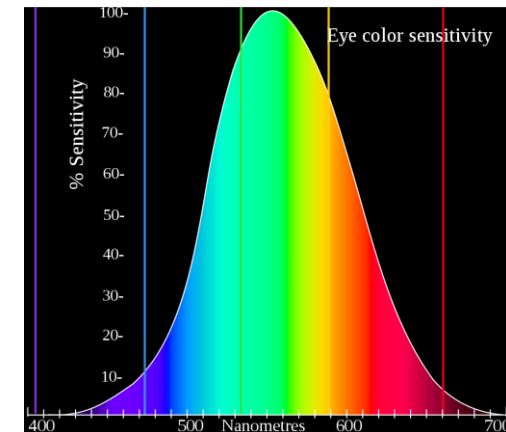
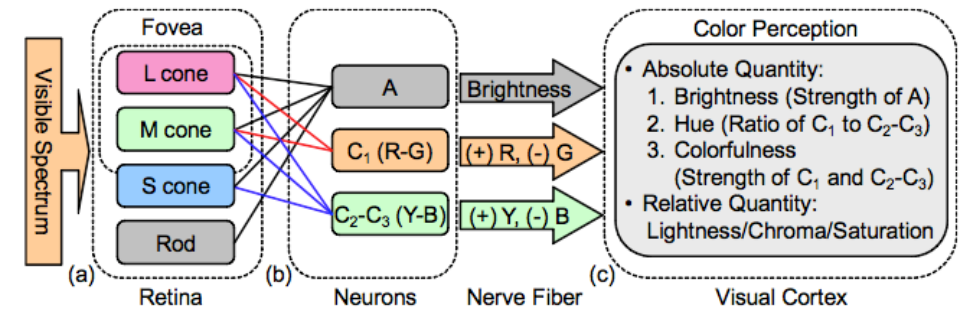
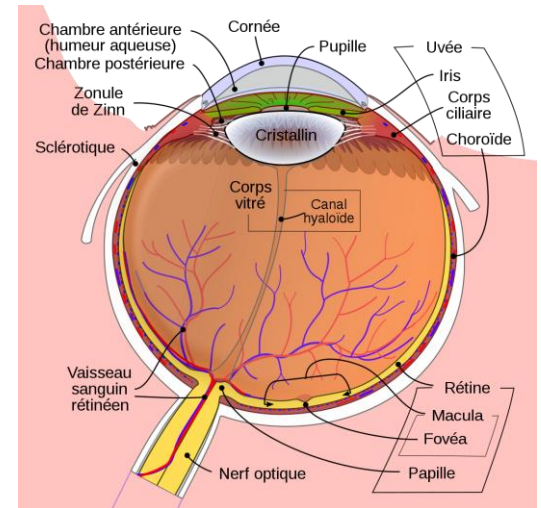
La vision est sélective

- › L'œil s'attache à certains endroits
- › Perception préattentive

La perception de la couleur n'est pas uniforme

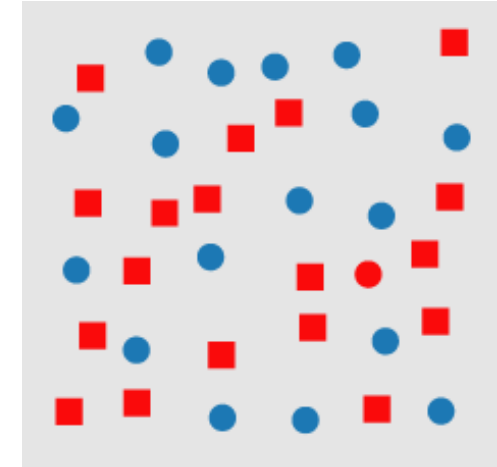
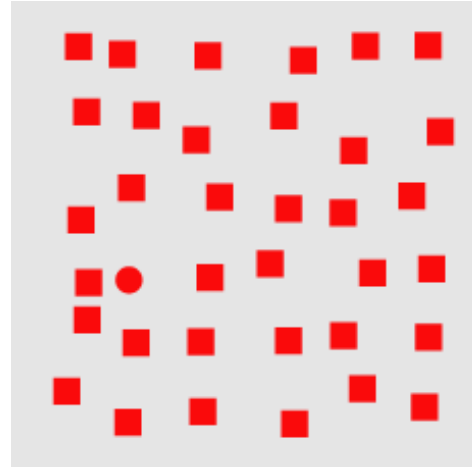
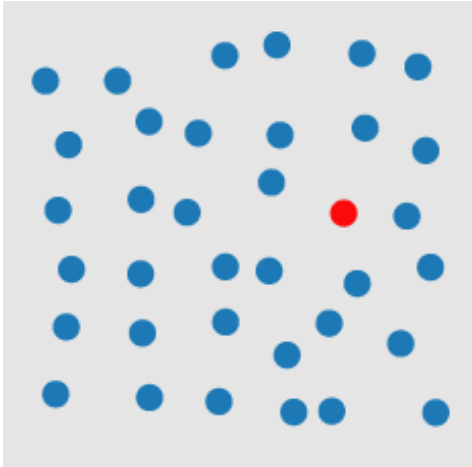
Nous avons tendance à grouper des éléments

- › Principes de Gestalt



Perception pré-attentive

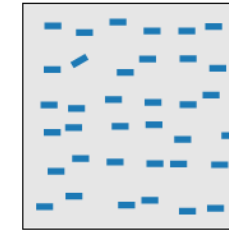
Où est le point rouge ?



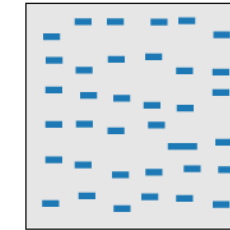
Perception pré-attentive

Aussi appelé « Visual popout »

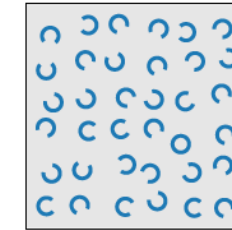
- › Propriétés détectées par le système visuel «primaire»
- › Rapide et précis



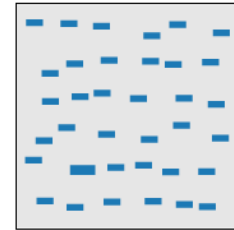
line (blob) orientation
Júlész & Bergen 83; Sagi & Júlész 85a, Wolfe et al. 92; Weigle et al. 2000



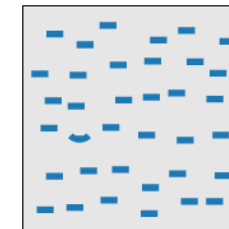
length, width
Sagi & Júlész 85b; Treisman & Gormican 88



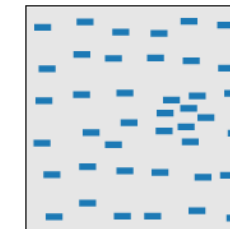
closure
Júlész & Bergen 83



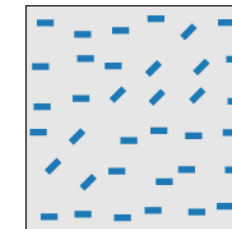
size
Treisman & Gelade 80; Healey & Enns 98; Healey & Enns 99



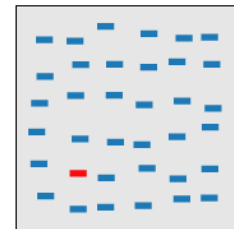
curvature
Treisman & Gormican 88



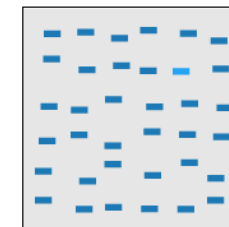
density, contrast
Healey & Enns 98; Healey & Enns 99



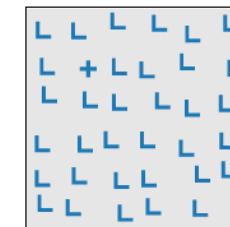
number, estimation
Sagi & Júlész 85b; Healey et al. 93; Trick & Pylyshyn 94



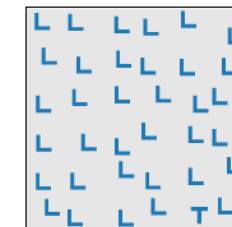
colour (hue)
Nagy & Sanchez 90; Nagy et al. 90; D'Zmura 91; Kawai et al. 95; Bauer et al. 96; Healey 96; Bauer et al. 98; Healey & Enns 99



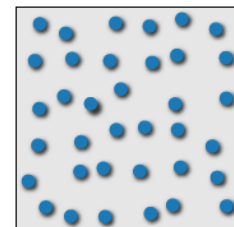
intensity, binocular lustre
Beck et al. 83; Treisman & Gormican 88; Wolfe & Franzel 88



intersection
Júlész & Bergen 83



terminators
Júlész & Bergen 83

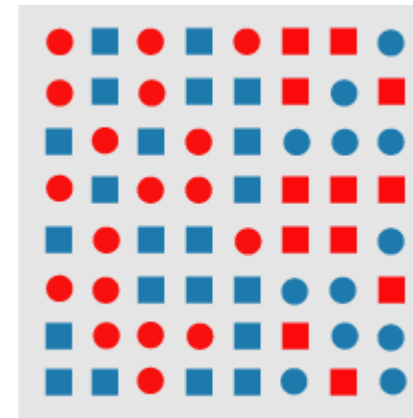
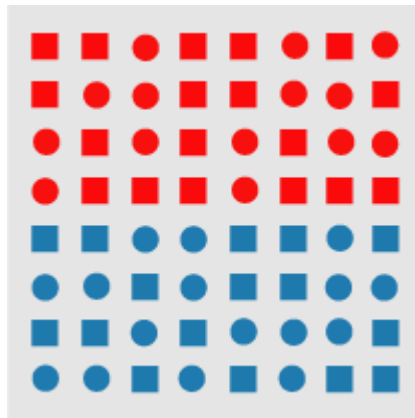


3D depth cues
Enns 90b; Nakayama & Silverman 86

Perception pré-attentive

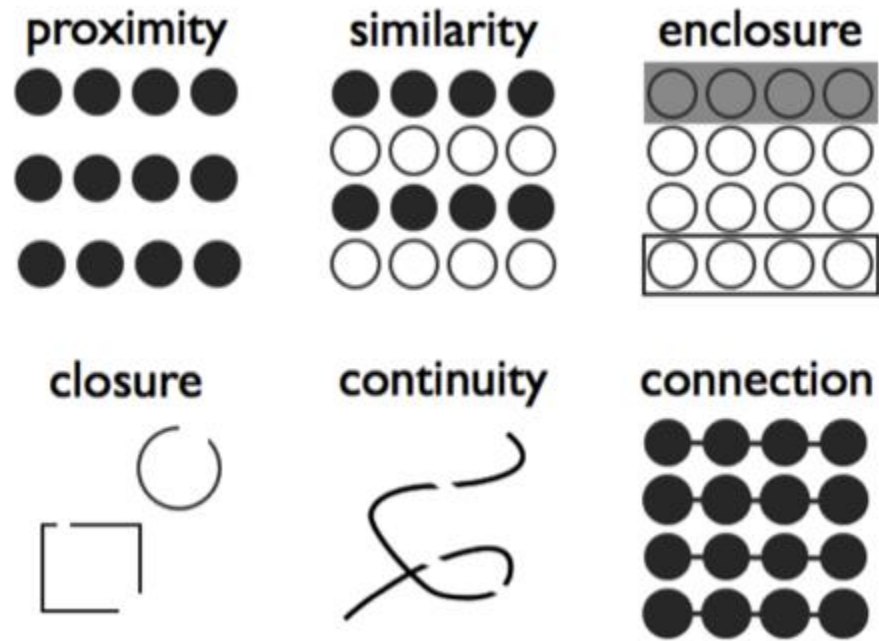
Séparation de zones

› Regroupement d'éléments



Principes de Gestalt

Notre vision tend à regrouper certaines informations selon certains paramètres



Choix de représentations

Principes

La création d'une visualisation est guidée par différents principes:

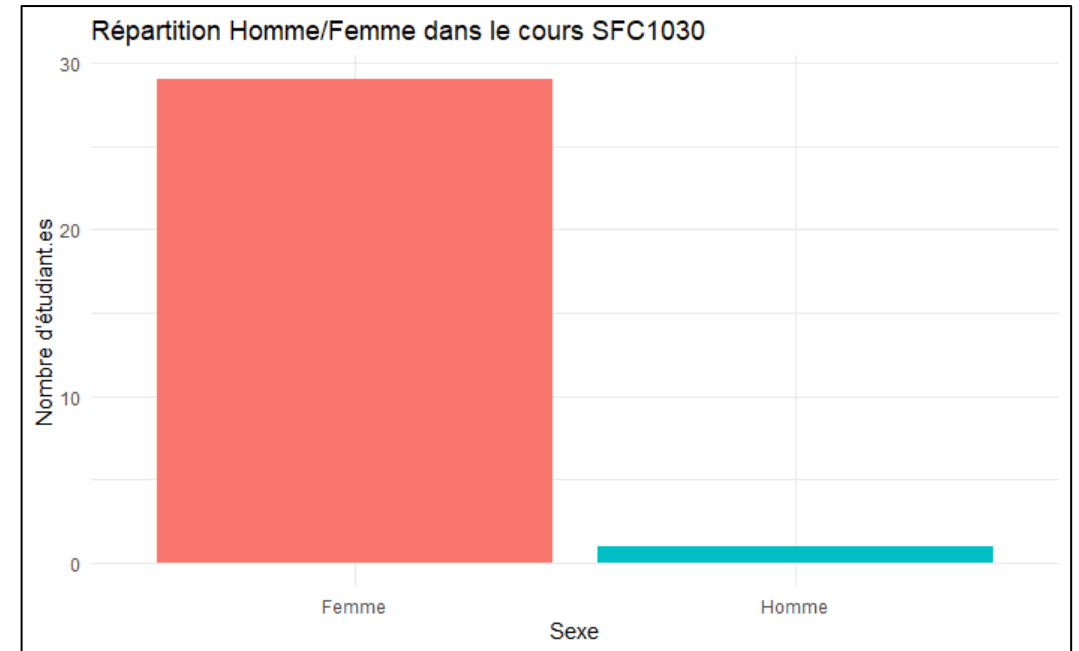
- › Principe d'utilité
- › Principe de dimensionnalité
- › Principe d'expressivité
- › Principe d'intégrité
- › Principe d'efficacité

Principe d'utilité

On utilise une visualisation lorsque c'est nécessaire

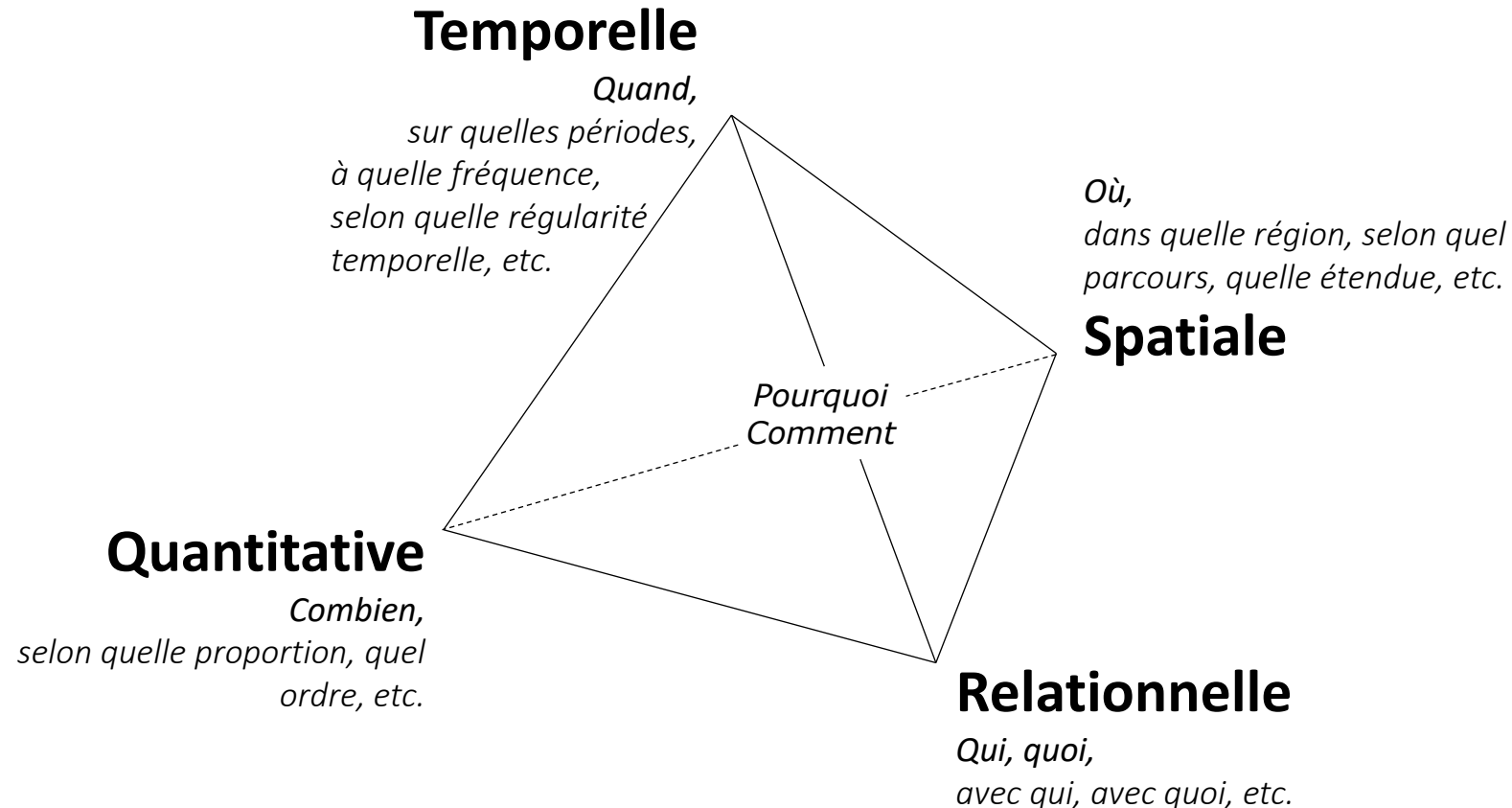
- › Souvent pour travailler avec des données présentant un certain degré de complexité

Une image vaut mieux de 10 000 mots

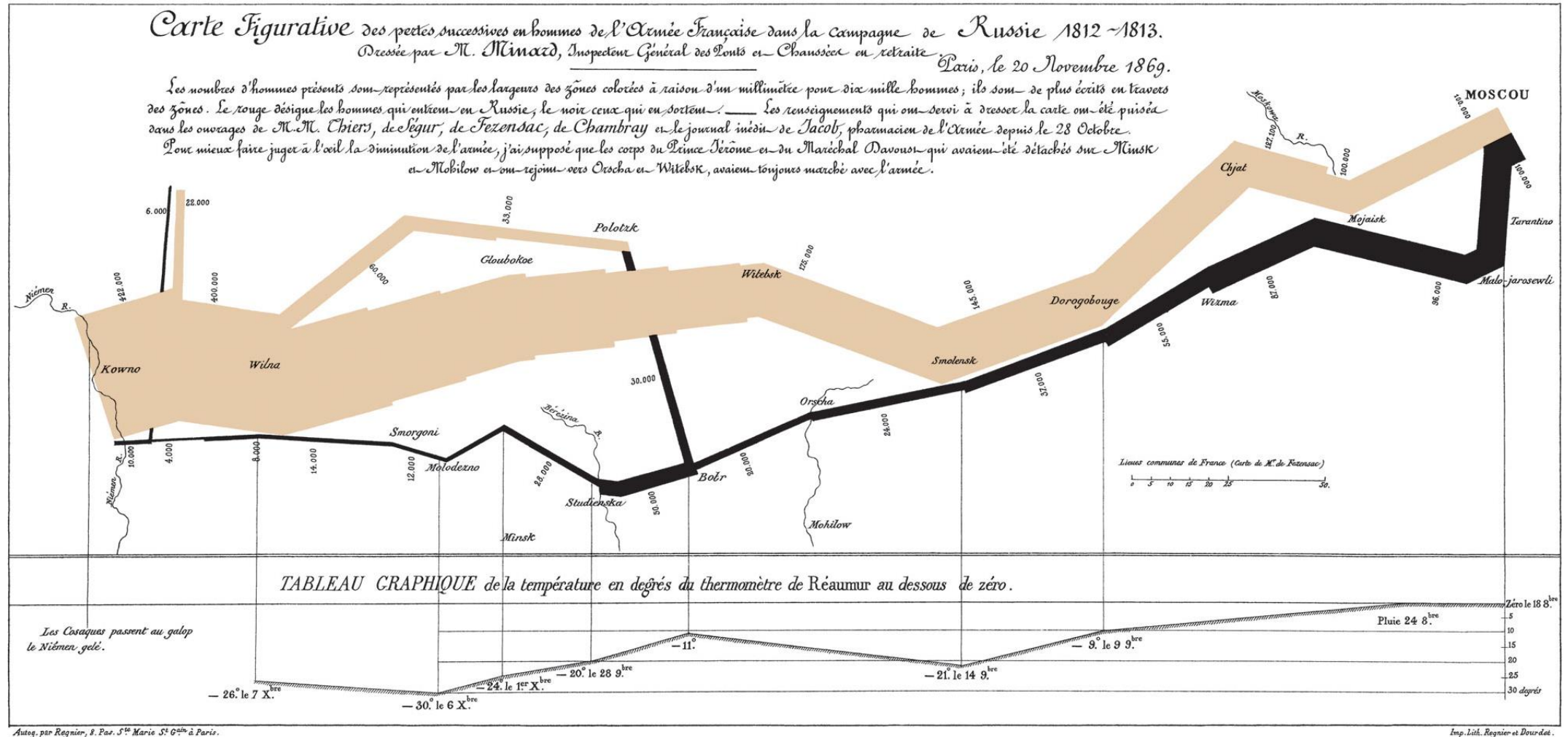


Principe de dimensionnalité

Le choix d'une représentation doit être fait en fonction de la dimension d'analyse dominante



Exemples



Nature des données

		Opérations possibles	Exemples
Qualitative	Nominale	= ≠	Types de ... (cas, lieux, personnes, rôles, etc.)
	Ordinale	= ≠ < >	Gravité, importance, hiérarchie
Quantitative	Intervalle	= ≠ < > + -	Dates et localisations
	Ratio	= ≠ < > + - · /	Mesures et dénombrements (tailles, distances, coûts, ...)

Principe d'expressivité

Un ensemble de faits est bien exprimé dans un langage visuel si les phrases (les éléments de représentation) du langage expriment

- > la **totalité** des faits
- > **uniquement** les faits contenus dans les données

Une communication est basée sur le fait que tous les participants partagent des conventions qui déterminent

- > Comment **construire** le message
- > Comment **interpréter** le message

Monosémie

« Un système est monosémique quand la connaissance de la signification de chaque signe précède l'observation de l'assemblage des signes »

Besoin de lever les ambiguïtés

- › Documentation des sources
 - › **Fiabilité**: Auteur, version, date/lieu de création/modification
 - › **Validité**: Sources des données
- › Sur la représentation
 - › **Titre**: simple et informatif
 - › **Légende**: données, sources, définir les conventions
 - › **Échelles**
 - › **Repères**: conventions visuelles

Variables visuelles

Marques – Les briques de base de la visualisation

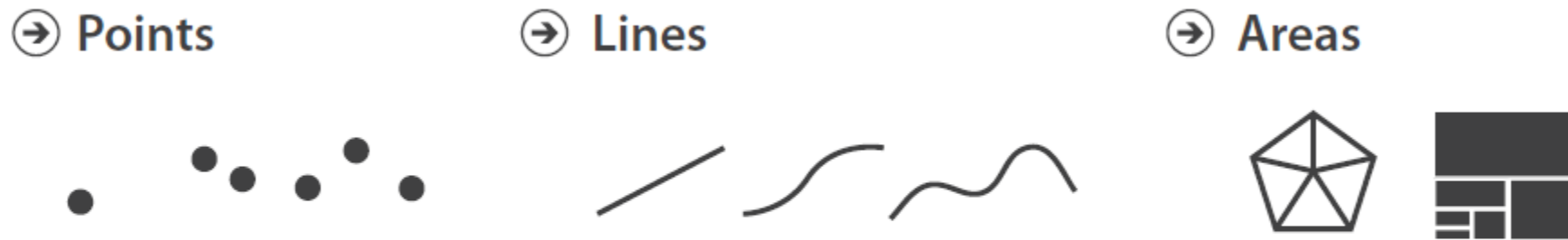


Figure 5.2. Marks are geometric primitives.

Variables visuelles

Channels – Organisation et apparence des marques

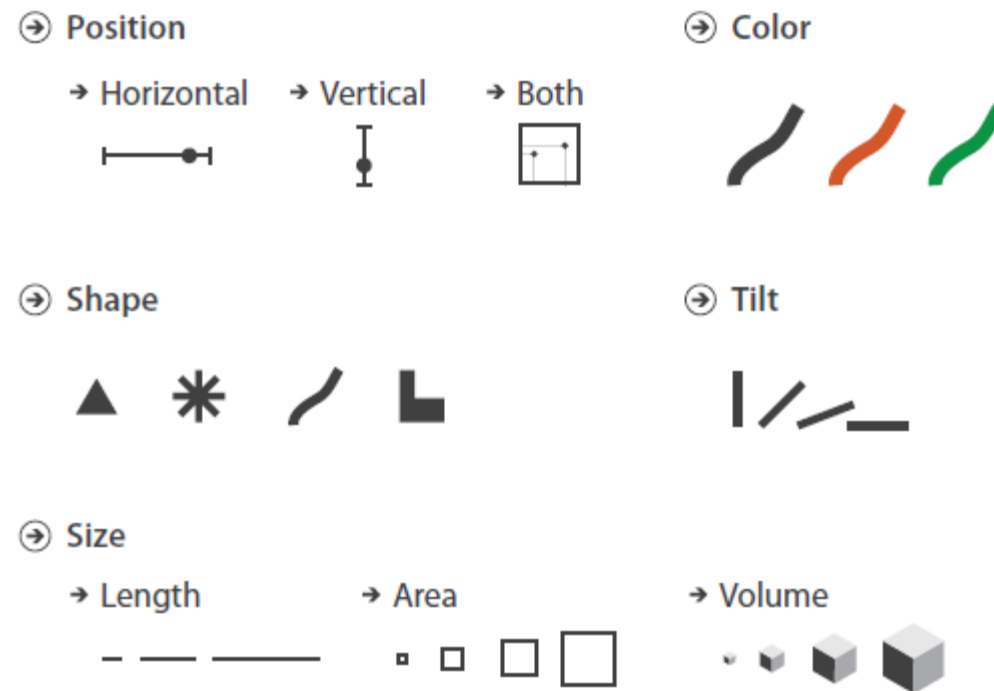


Figure 5.3. Visual channels control the appearance of marks.

Variables visuelles

LES VARIABLES DE L'IMAGE										
			POINTS			LIGNES			ZONES	
XY										
2 DIMENSIONS DU PLAN										
Z										
TAILLE										
VALEUR										
LES VARIABLES DE SÉPARATION DES IMAGES										
GRAIN										
COULEUR										
ORIENTATION										
FORME										

Choix des variables

Marks as Items/Nodes

➔ Points



➔ Lines



➔ Areas



Marks as Links

➔ Containment



➔ Connection



Figure 5.5. Marks can represent individual items, or links between them.




























Choix des variables

Plus précis

↑

↓

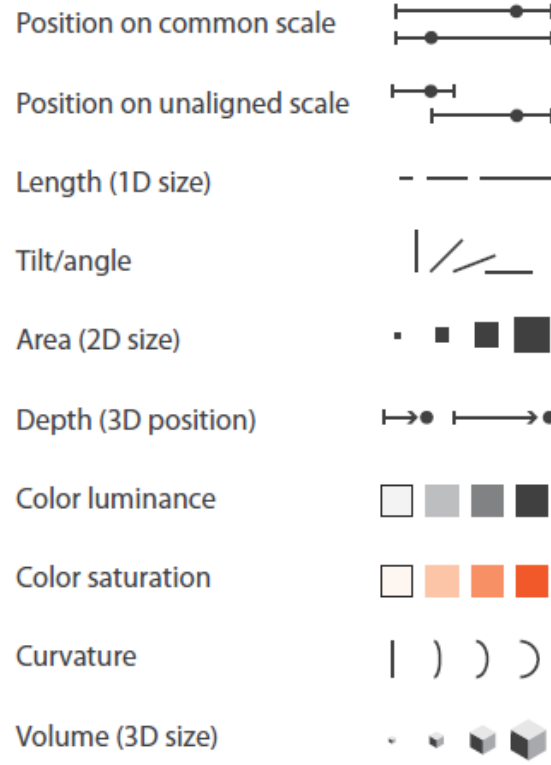
Moins précis

Quantitatif	Ordinal	Nominal
Position 	Position 	Position 
Longueur 	Intensité 	Teinte 
Angle 	Saturation 	Intensité 
Pente 	Teinte 	Saturation 
Surface 	Longueur 	Forme 
Intensité 	Angle 	Longueur 
Saturation 	Pente 	Angle 
Teinte 	Surface 	Pente 
Forme 	Forme 	Surface 

Choix des variables

Channels: Expressiveness Types and Effectiveness Ranks

➔ Magnitude Channels: Ordered Attributes



➔ Identity Channels: Categorical Attributes

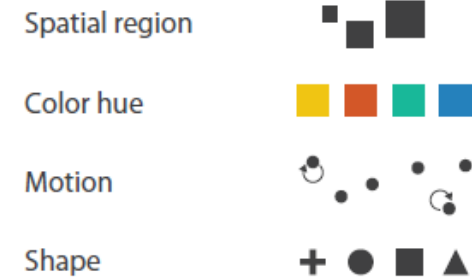


Figure 5.6. Channels ranked by effectiveness according to data and channel type. Ordered data should be shown with the magnitude channels, and categorical data with the identity channels.

Exemple



Choix des couleurs

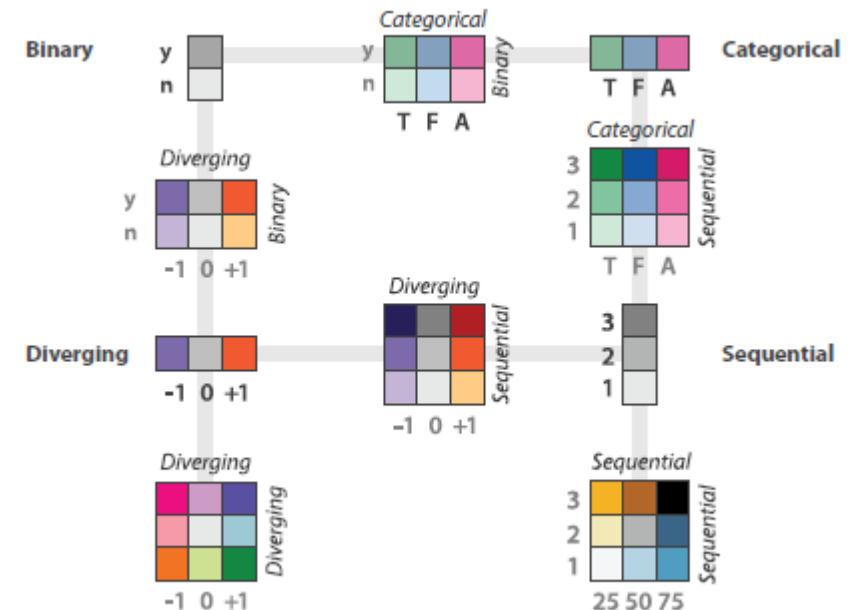
On utilise des *color maps*

- › Choix de la plage de couleur va dépendre de ce que l'on souhaite représenter

Attention avec les couleurs...

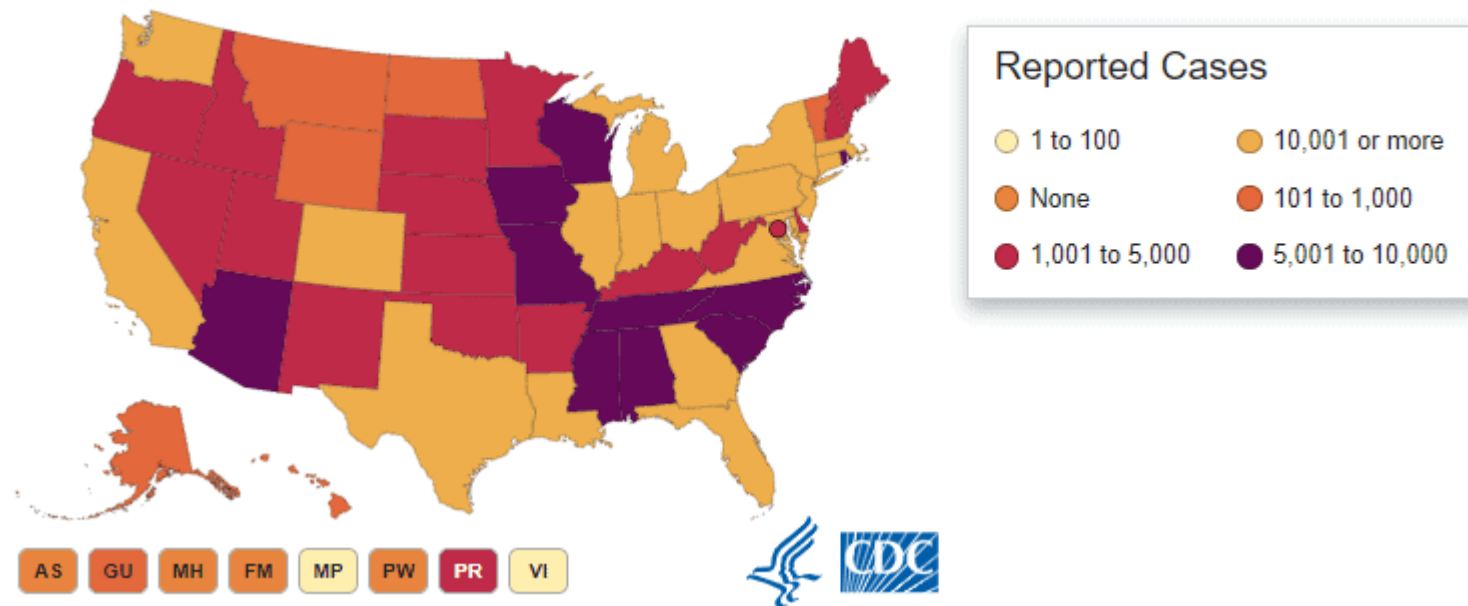
Certaines personnes sont daltoniennes

- › 8% des hommes – 1% des femmes



Example

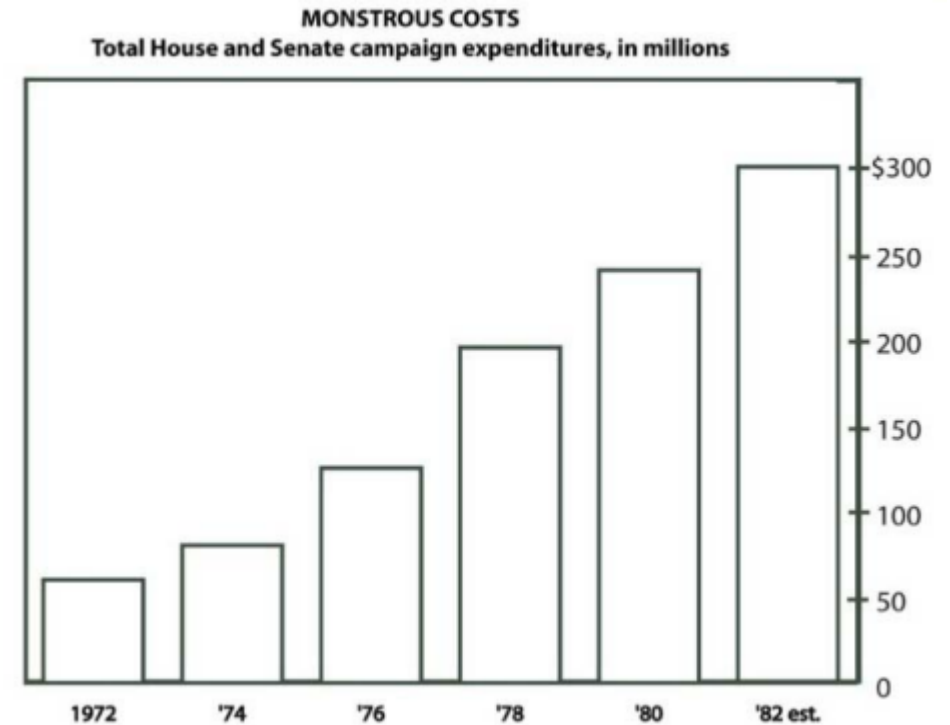
18 states report more than 10,000 cases of COVID-19.



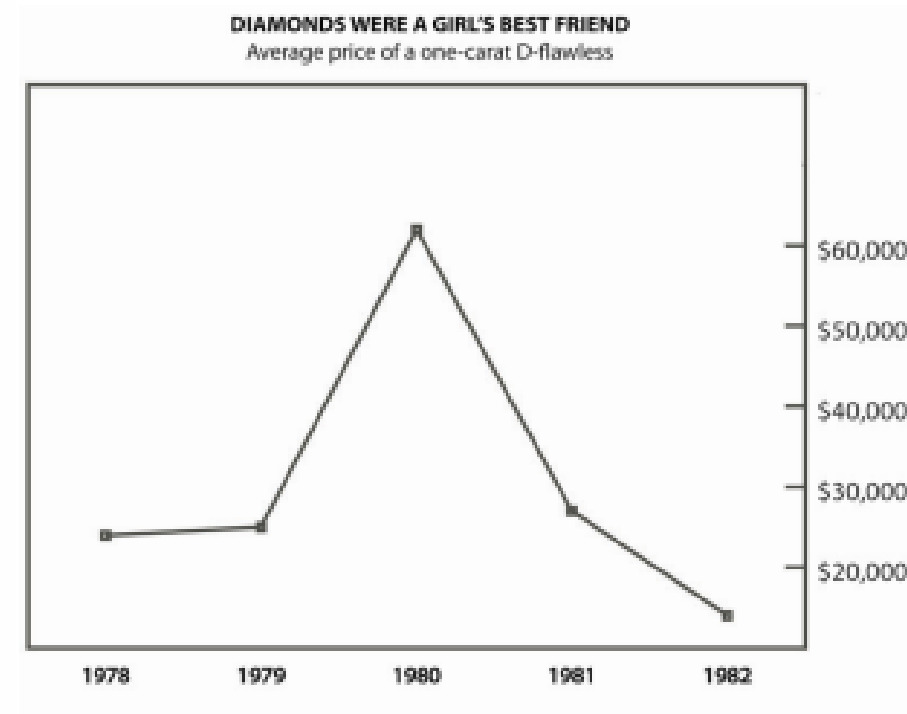
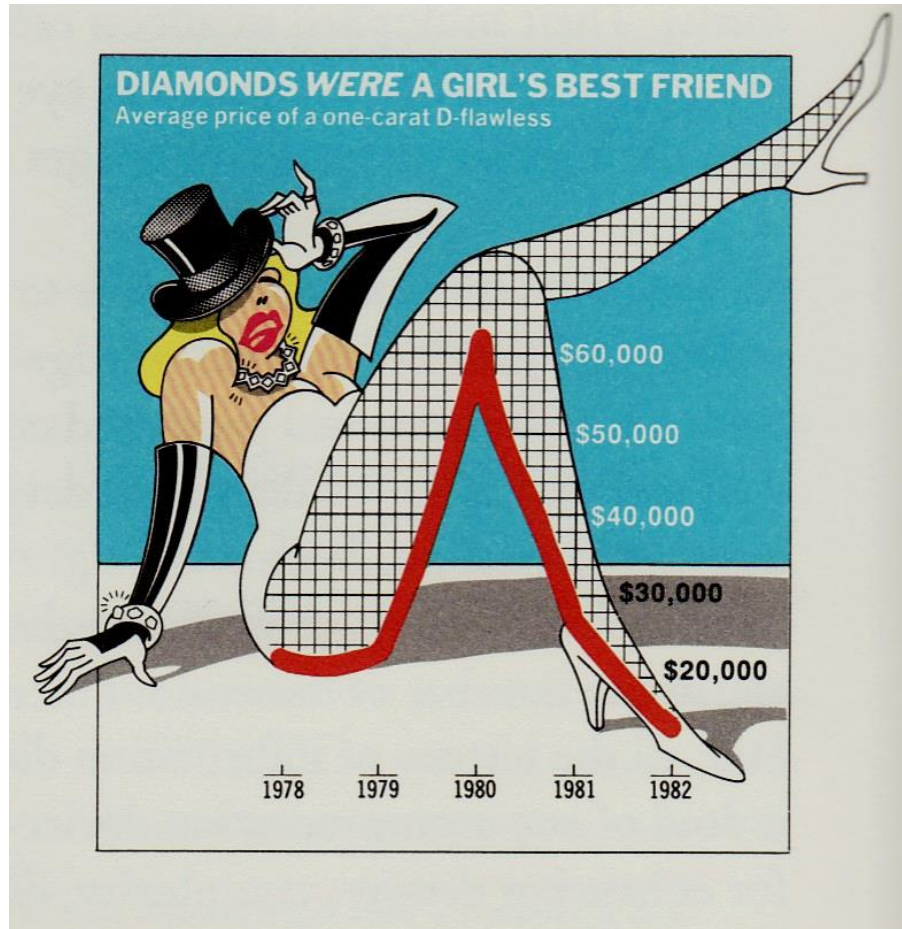
Principe d'efficacité

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

Exemple



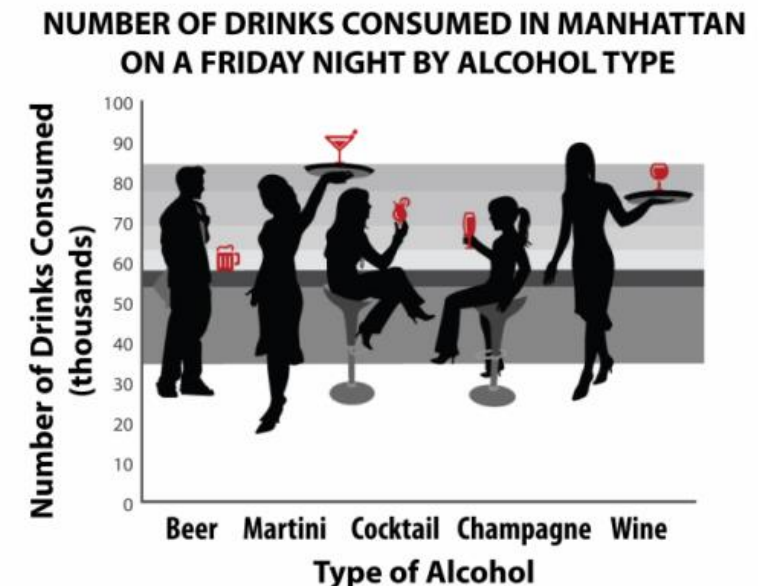
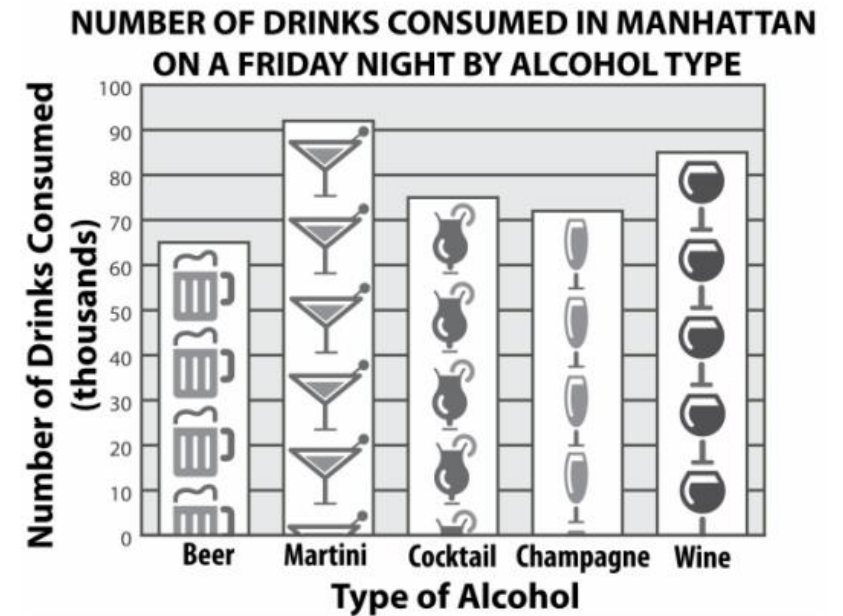
Exemple



Chartjunk - intérêt

Résultats de l'étude de Bateman *et al.* (2010)

- > Ne semble pas y avoir de problème d'interprétation
- > Permet de se souvenir plus longtemps de la visualisation (sujet et détails)
- > Pros
 - > Agréable, attire l'œil, plus d'impact
- > Cons
 - > Lisibilité, sérieux



Principe d'intégrité

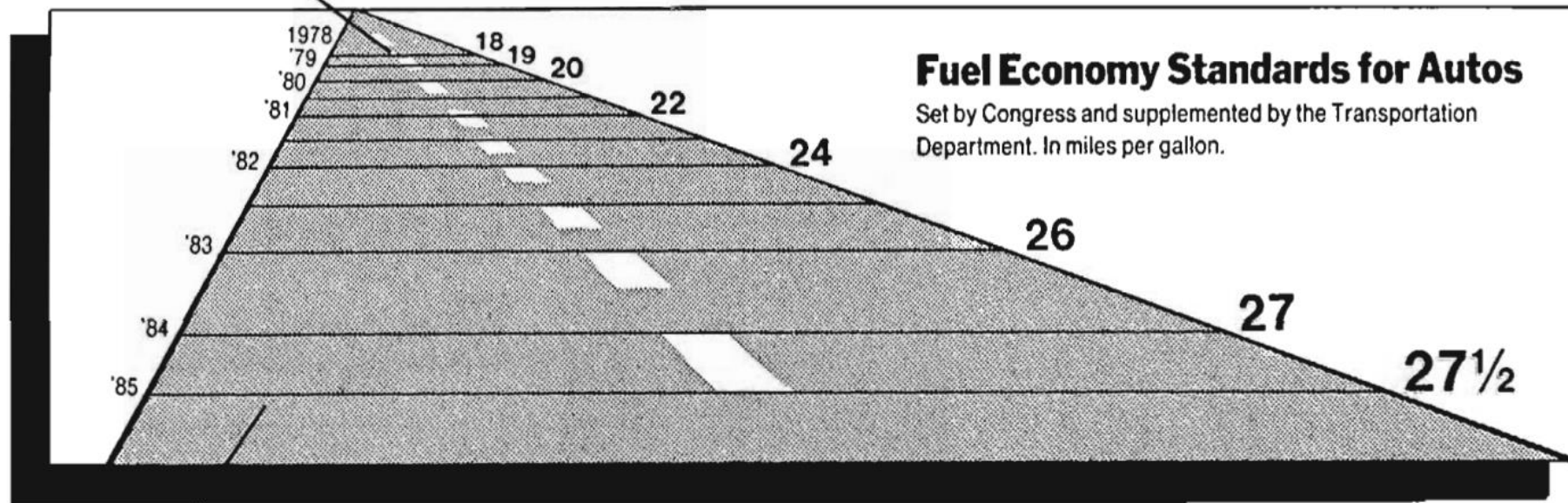
Toute variation de grandeur, telle que mesurée sur le graphique doit être directement proportionnelle à la variation dans les données

Éviter les distorsions et ambiguïtés visuelles
› Éviter des biais de perceptions

$$\text{Lie Factor} = \frac{\text{Size of effect in graphic}}{\text{size of effect in data}}$$

Exemples - Distorsions

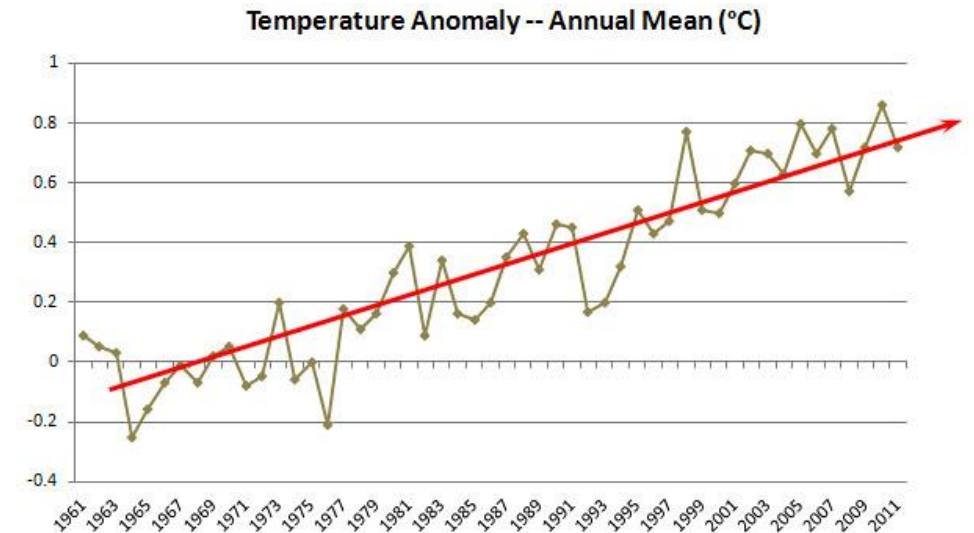
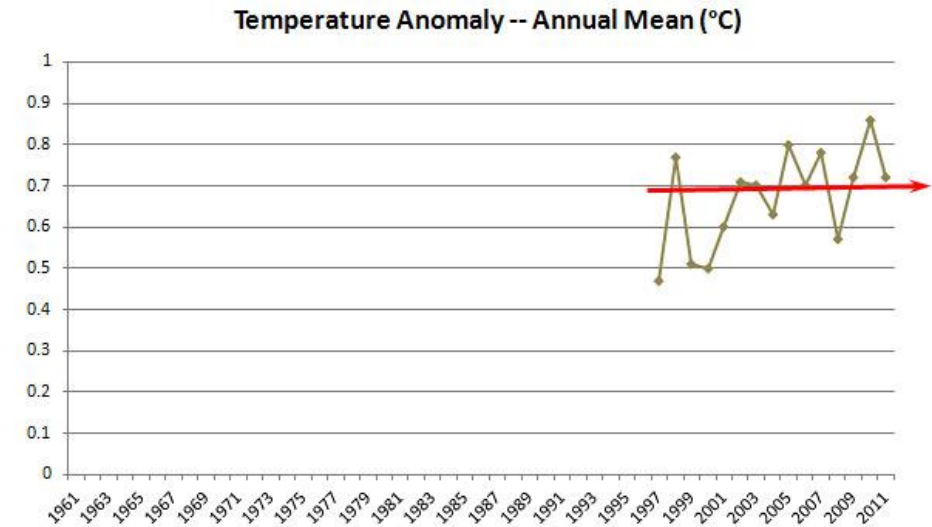
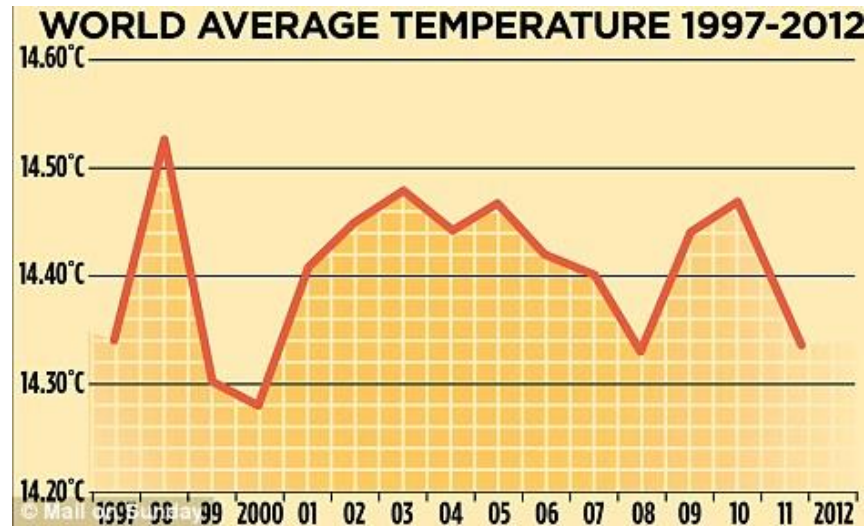
This line, representing 18 miles per gallon in 1978, is 0.6 inches long.



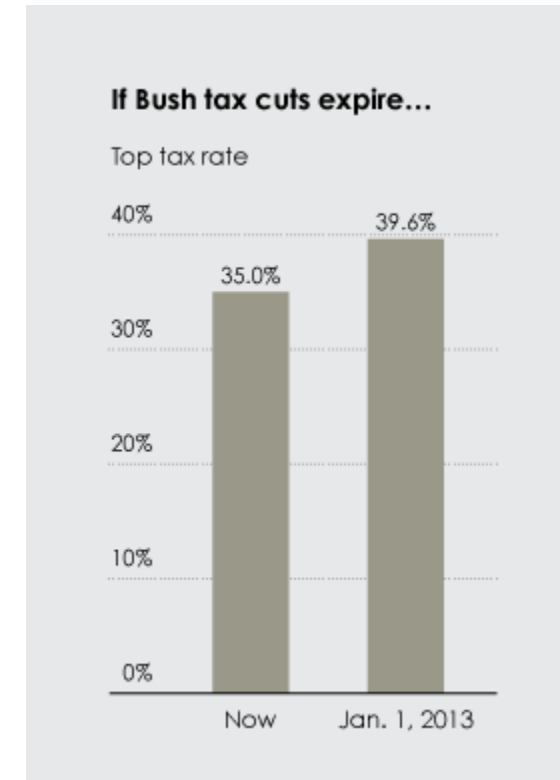
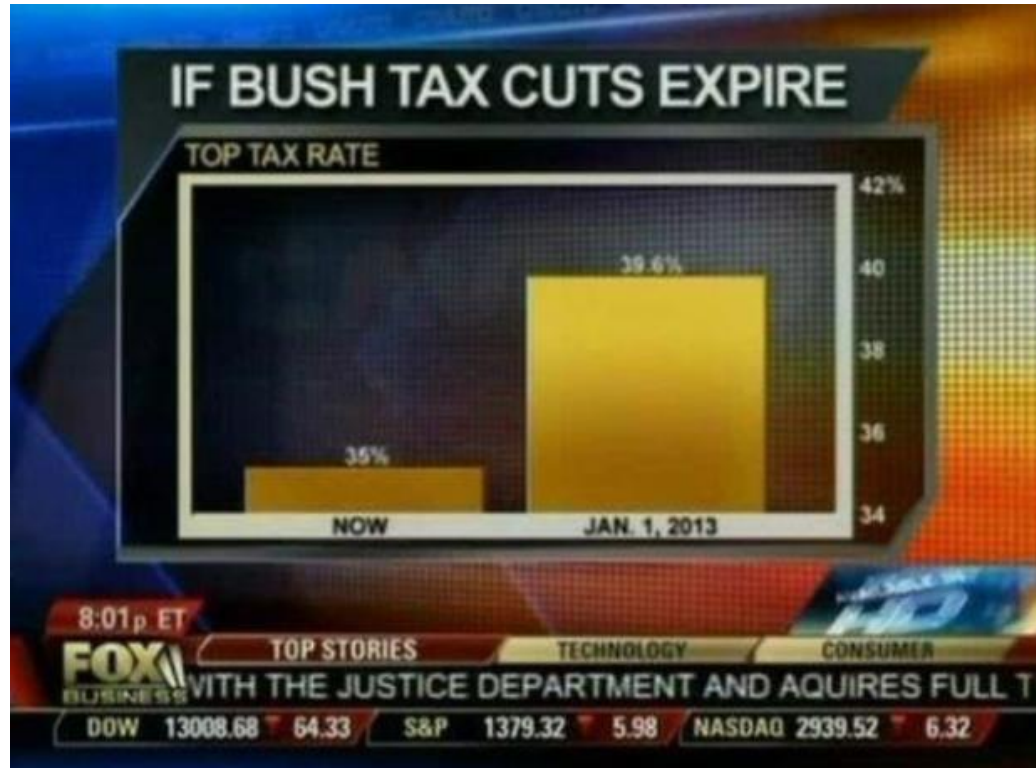
This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

New York Times, August 9, 1978, p. D-2.

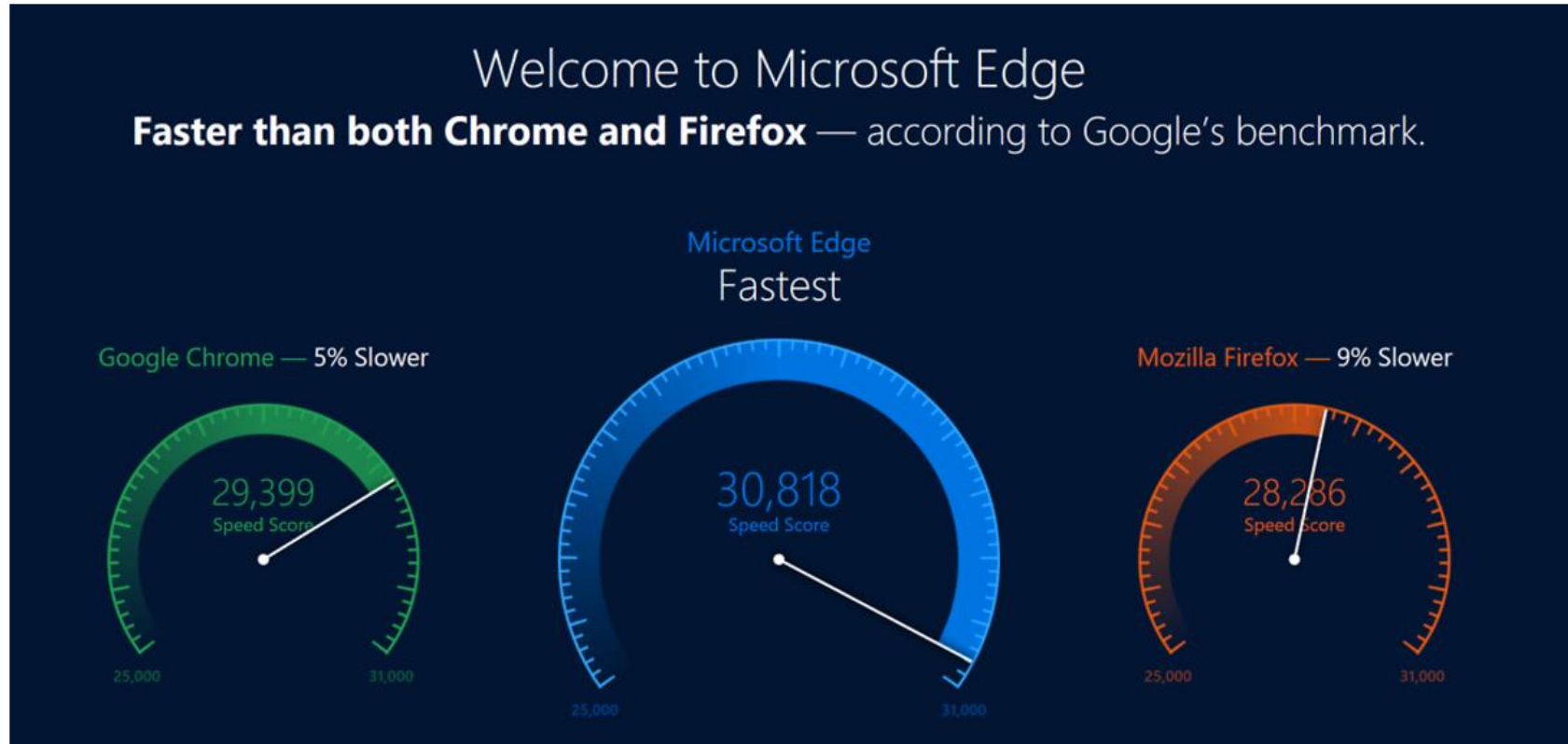
Exemples - Décontextualisation



Exemples - Changement d'échelle

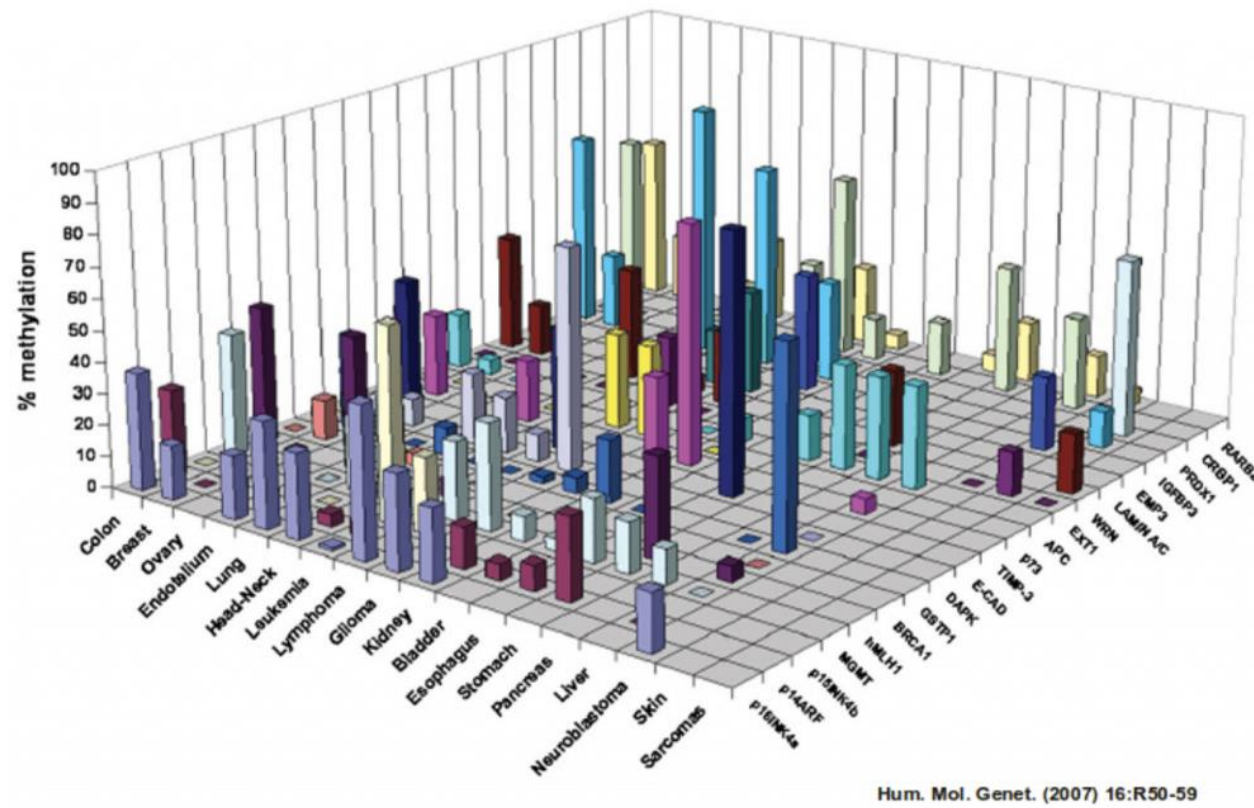


Exemples - Changement d'échelle

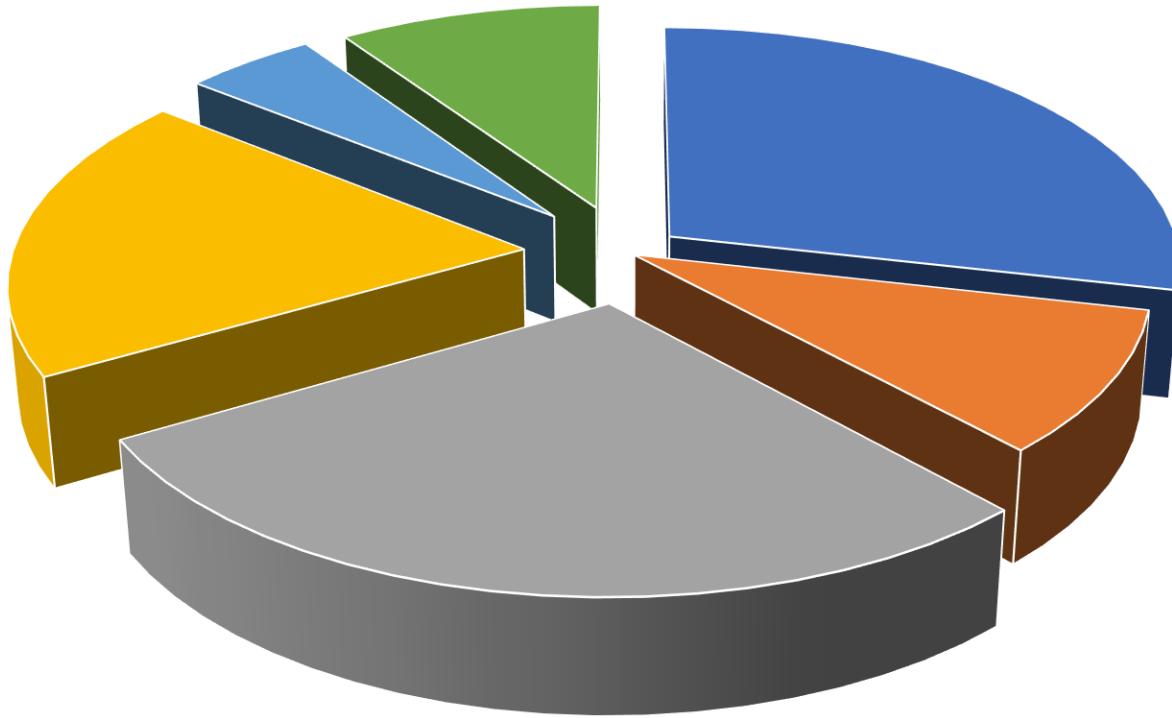


Exemple - 3D

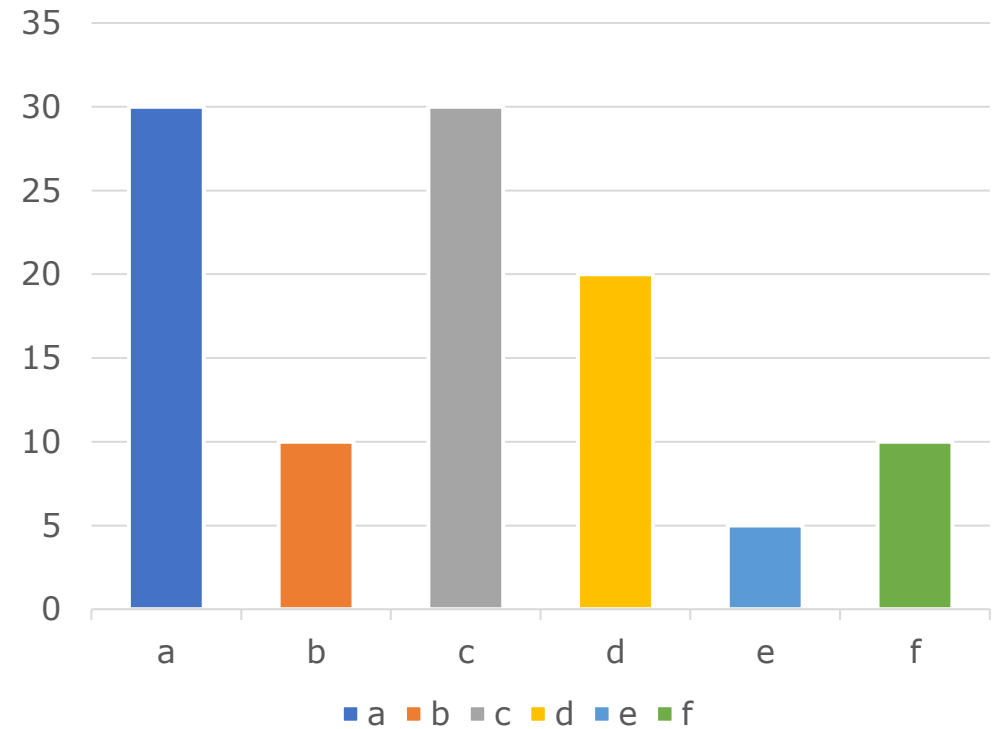
A CpG Island Hypermethylation Profile of Human Cancer



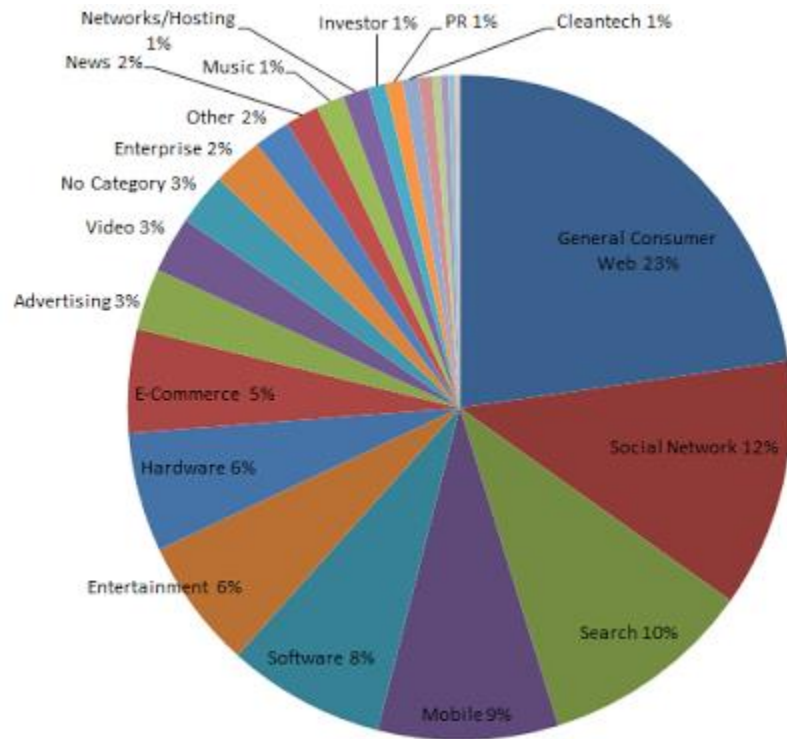
Exemple - 3D



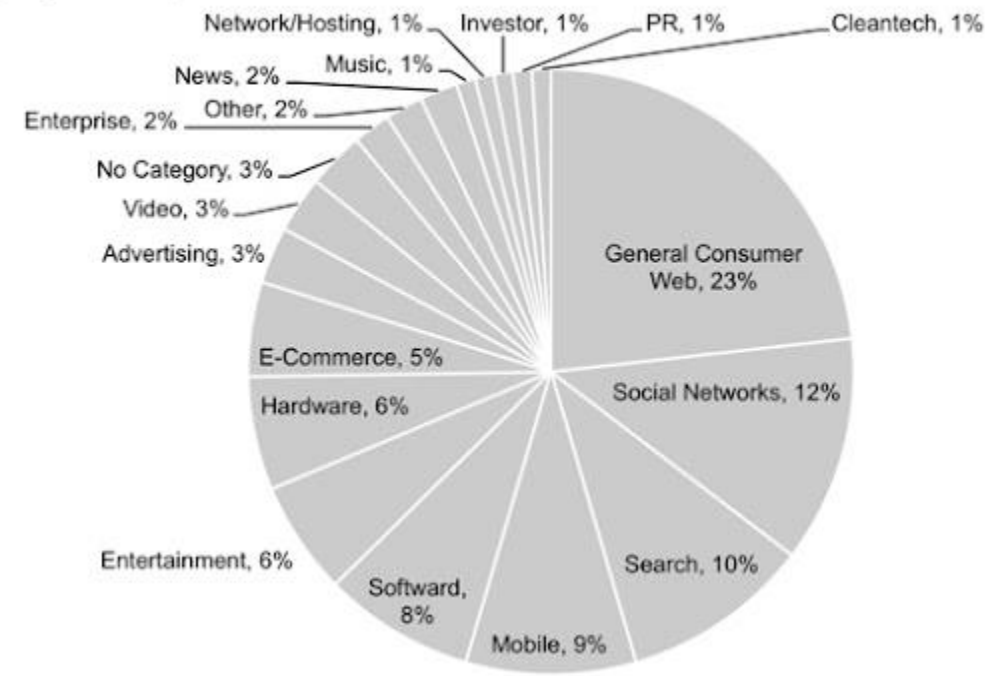
■ a ■ b ■ c ■ d ■ e ■ f



Exemple - Pie Chart

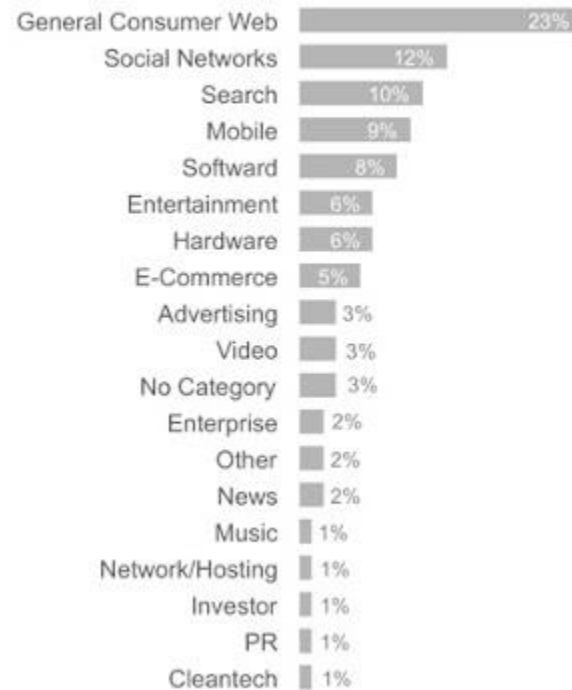


TechCrunch Coverage: 2005 - 2011
A slightly better pie?



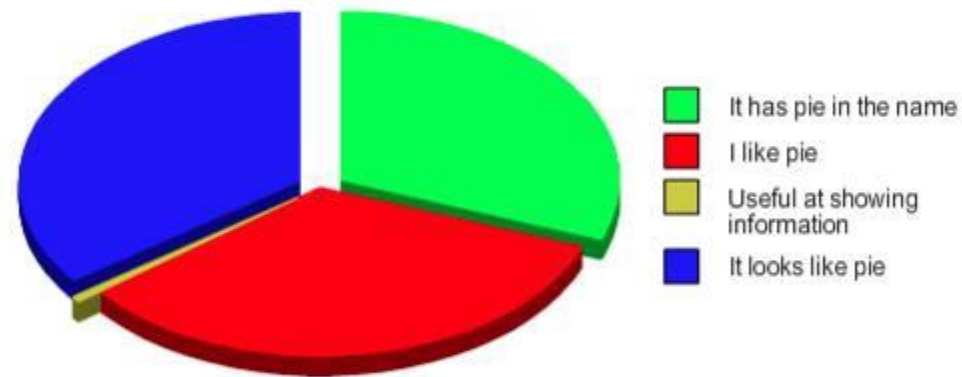
Exemple - Pie Chart

TechCrunch Coverage: 2005 - 2011
Bars are best!



Exemple - Pie Chart

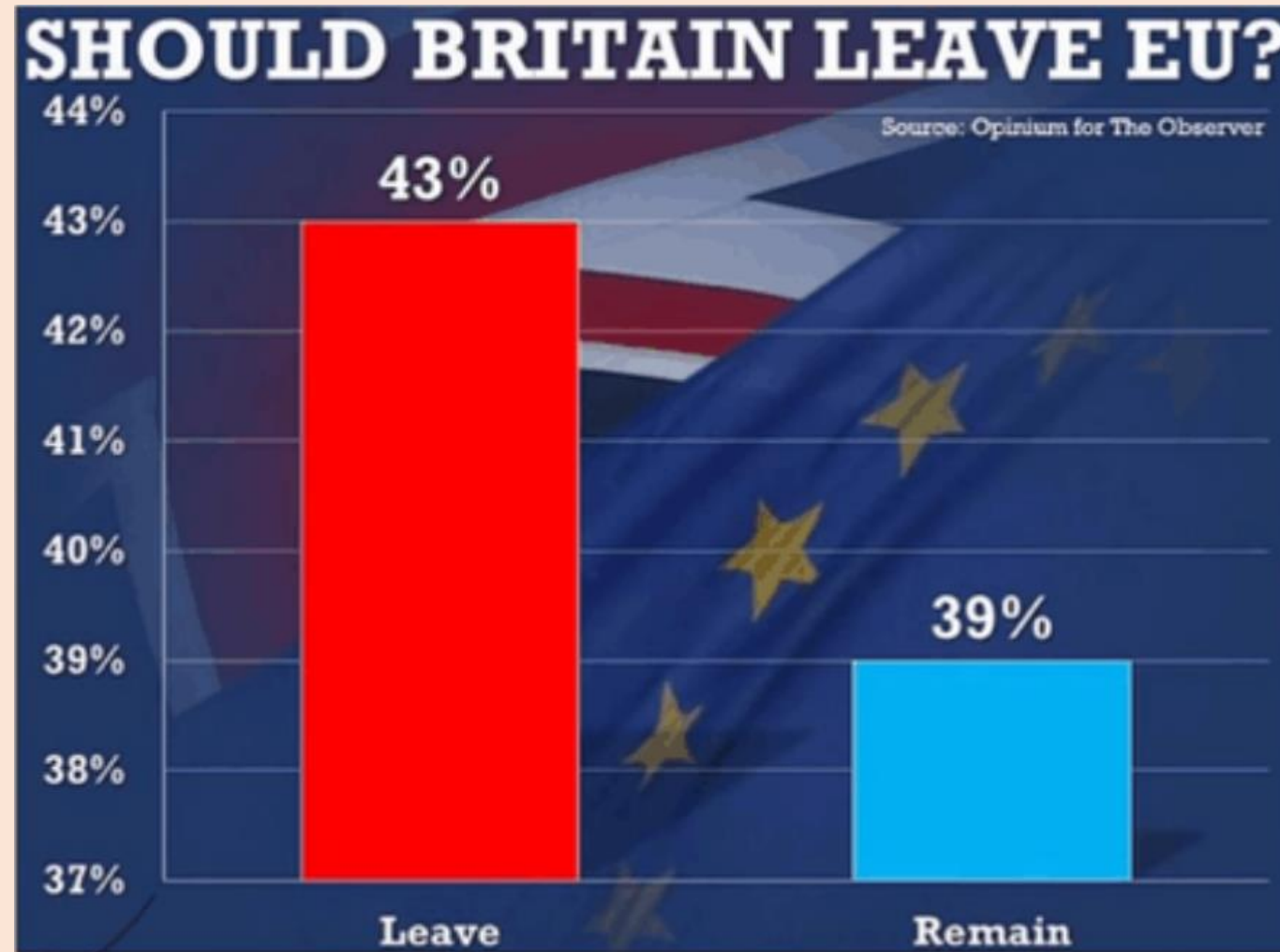
Why I like pie charts



Better pie charts:

<https://blog.datawrapper.de/pie-charts/>

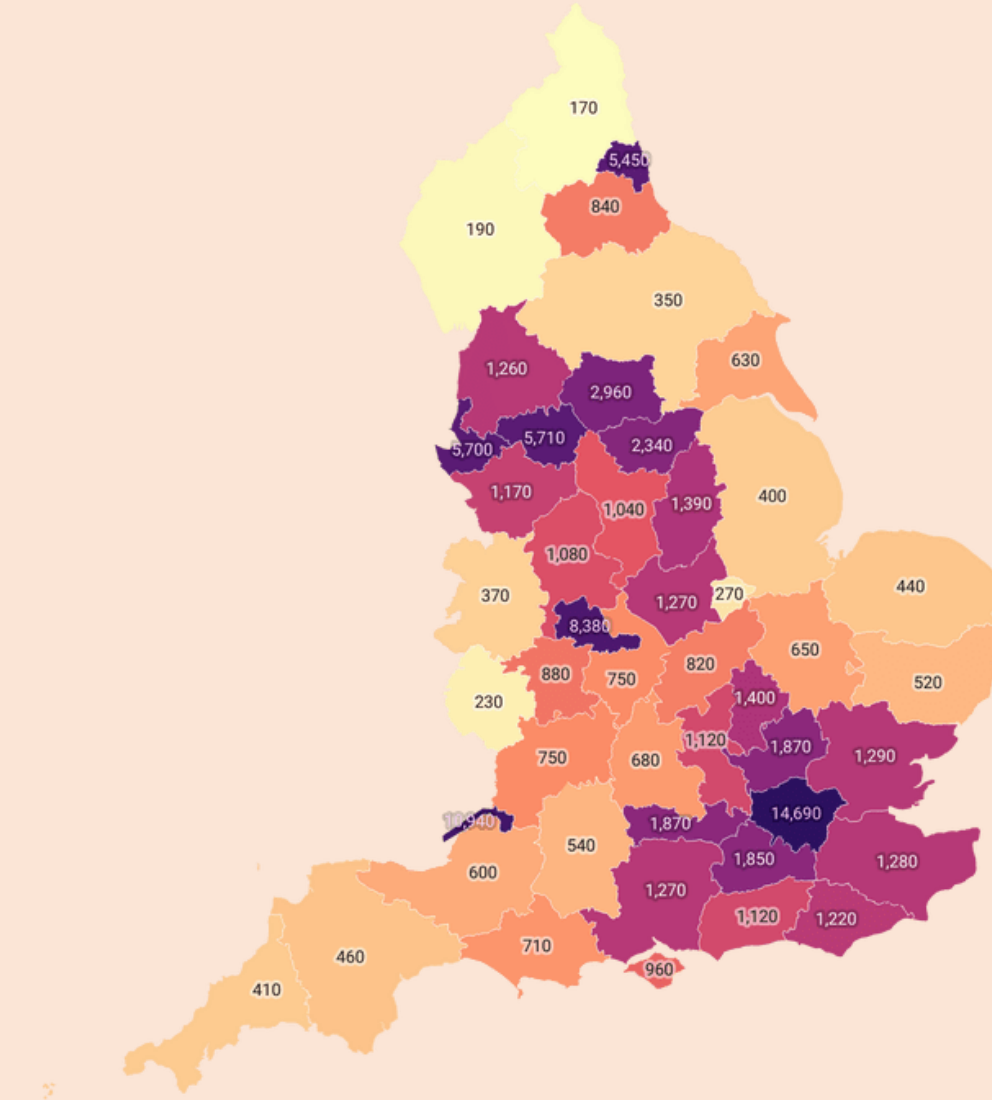
Exercise



Exercise

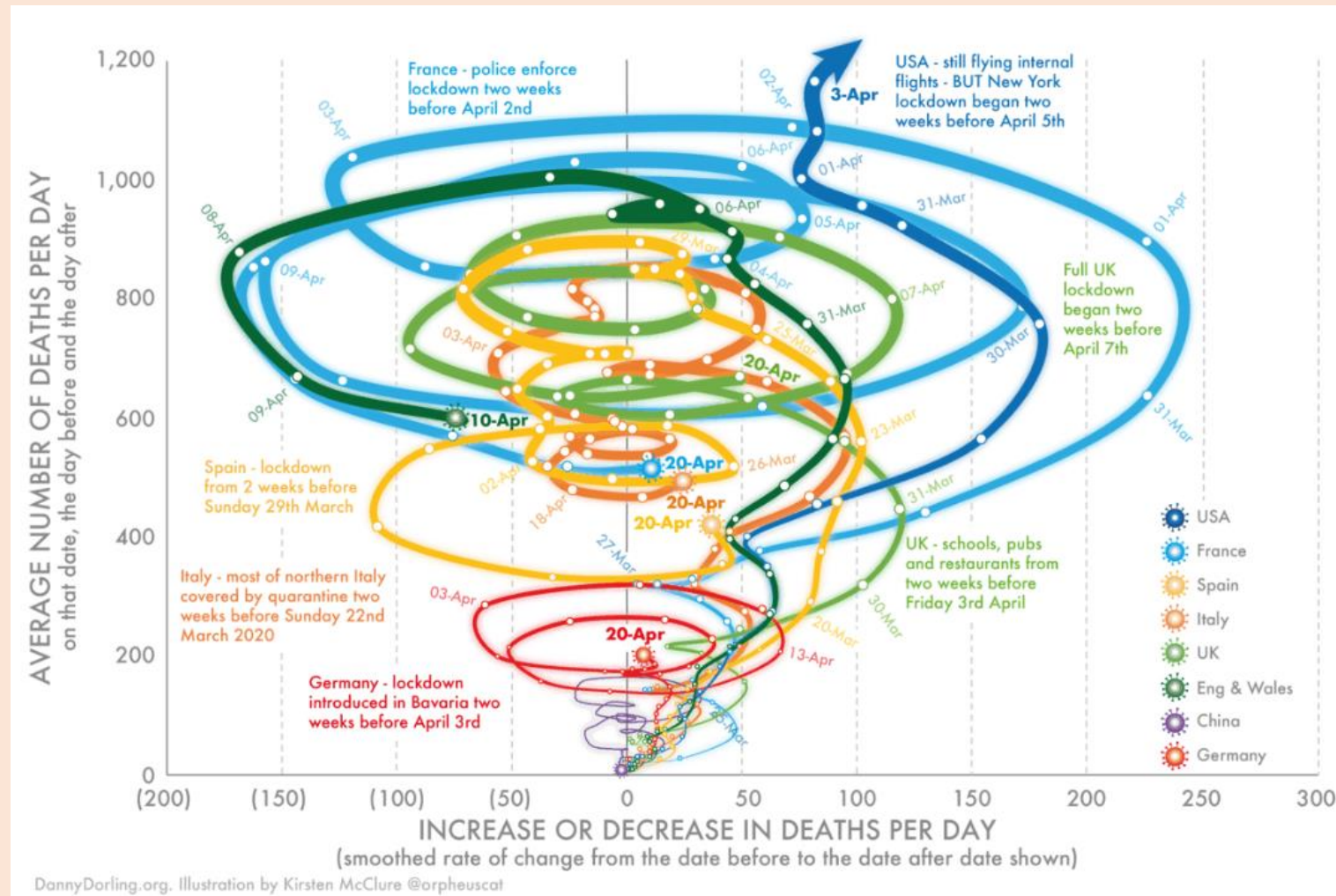
Population density of each county in England

In Square Miles



Created with Datawrapper

Exercise



Exercise

The Woodlands area age breakdowns

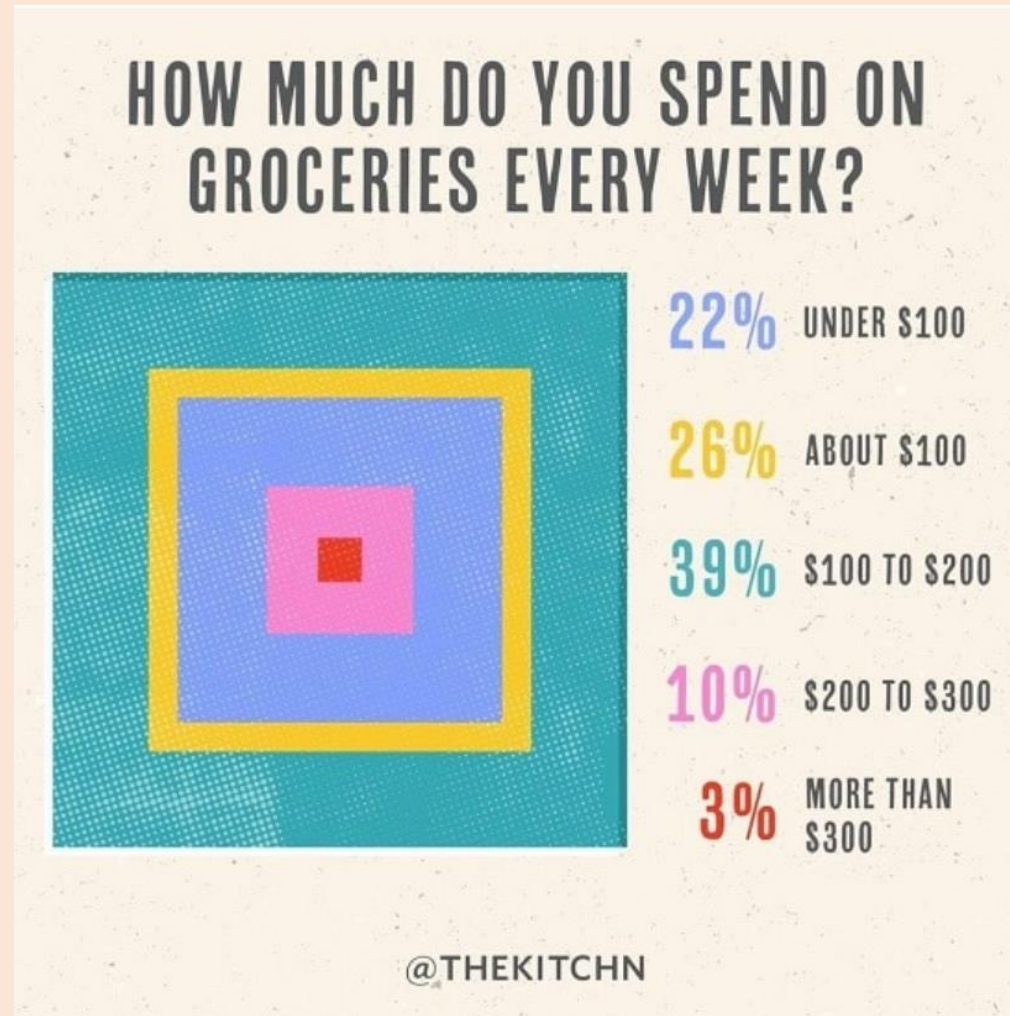
2015-19 (2019 American Community Survey 5-year estimates)

0-19 20-39 40-59 60-79 80+

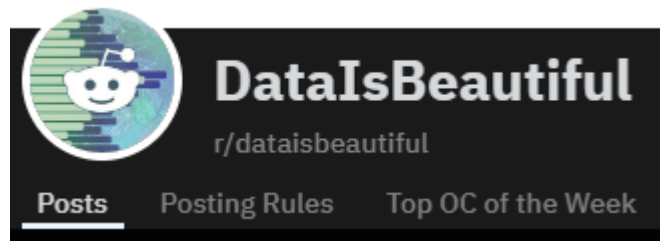


Source: U.S. Census Bureau

Exercise



Références



Quelques galleries:

- > <http://euclid.psych.yorku.ca/SCS/Gallery/noframes.html>
- > https://www.visual-literacy.org/periodic_table/periodic_table.html
- > <https://r-graph-gallery.com/>
- > <https://d3js.org/>
- > <https://d3-graph-gallery.com/>

Références

- › Munzner T (2014) *Visualization Analysis and Design*. New York: A K Peters/CRC Press. DOI : [10.1201/b17511](https://doi.org/10.1201/b17511).
- › Bertin J (2005) *Sémiologie graphique: les diagrammes - les réseaux - les cartes*. 4e éd. Ré-impression. Paris: École des Hautes Études en Sciences Sociales (1967 – 1^{ère} éd).
- › Tufte ER (2001) *The Visual Display of Quantitative Information*. 2e édition. Cheshire, Conn: Graphics Press.
- › Healey CG (s. d.) Perception in Visualization. Disponible à l'adresse : <https://www.csc2.ncsu.edu/faculty/healey/PP/> [Consulté le 20 février 2023].
- › Stevens SS (1946) On the Theory of Scales of Measurement. *Science* 103(2684). American Association for the Advancement of Science: 677-680. DOI : [10.1126/science.103.2684.677](https://doi.org/10.1126/science.103.2684.677).
- › Mackinlay J (1986) Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics* 5(2): 110-141. DOI : [10.1145/22949.22950](https://doi.org/10.1145/22949.22950).
- › Fekete J-D (2011) La visualisation analytique, pour comprendre des données complexes. Disponible à l'adresse : <https://interstices.info/la-visualisation-analytique-pour-comprendre-des-donnees-complexes/> [Consulté le 21 février 2023].
- › Farooq A (2012) A Thousand Words. Disponible à l'adresse : <http://www.softviscollection.org/intro/a-thousand-words/> [Consulté le 21 février 2023].
- › Rossy, Q, Ribaux O, Boivin R et Fortin F (2019) Le traitement de l'information dans l'enquête criminelle. In: Cusson M, Ribaux O, Blais É, et al. (éds) *Nouveau traité de sécurité: sécurité intérieure et sécurité urbaine*. Nouvelle édition. Montréal: Hurtubise, p. 428-447.
- › Bateman S, Mandryk RL, Gutwin C, Genest A, McDine D et Brooks C (2010) Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 10 avril 2010, p. 2573-2582. CHI '10. Association for Computing Machinery. DOI : [10.1145/1753326.1753716](https://doi.org/10.1145/1753326.1753716).
- › Berry C (2021) When Data Visualization Really Isn't Useful (and When It Is) - Old Street Solutions. Disponible à l'adresse : <https://www.oldstreetsolutions.com/good-and-bad-data-visualization> [Consulté le 22 février 2023].