

Visualisation de données

Introduction

Raphaël Fournier-S'niehotta

CNAM Paris, fournier@cnam.fr

HTT-FOD
RCP216
2021-2022

le **cnam**

Plan

1 | Introduction

- 1 – Motivation
- 2 – Histoire, exemples

2 | Principes

- 1 – Données
- 2 – Composants de la visualisation

3 | Représentations

4 | Interaction

5 | Bonnes pratiques

- 1 – Choix du type de graphique
- 2 – Pièges à éviter

6 | Sujets avancés

Introduction

Motivation

Data visualisation

À quoi sert la visualisation de données ?

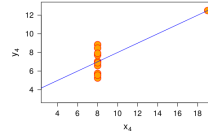
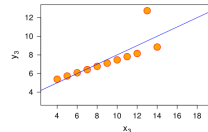
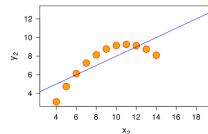
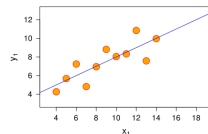
- à comprendre
 - à apprendre
 - à communiquer
 - à décider
 - à convaincre
-
- Data Viz
 - Info[rmation] Viz

Pourquoi visualiser?

■ Quartet, de Francis Anscombe

I		II		III		IV	
x	y	x	y	x	y	x	y
10,0	8,04	10,0	9,14	10,0	7,46	8,0	6,58
8,0	6,95	8,0	8,14	8,0	6,77	8,0	5,76
13,0	7,58	13,0	8,74	13,0	12,74	8,0	7,71
9,0	8,81	9,0	8,77	9,0	7,11	8,0	8,84
11,0	8,33	11,0	9,26	11,0	7,81	8,0	8,47
14,0	9,96	14,0	8,10	14,0	8,84	8,0	7,04
6,0	7,24	6,0	6,13	6,0	6,08	8,0	5,25
4,0	4,26	4,0	3,10	4,0	5,39	19,0	12,50
12,0	10,84	12,0	9,13	12,0	8,15	8,0	5,56
7,0	4,82	7,0	7,26	7,0	6,42	8,0	7,91
5,0	5,68	5,0	4,74	5,0	5,73	8,0	6,89

Propriété	Valeur
Moyenne des x	9,0
Variance des x	10,0
Moyenne des y	7,5
Variance des y	3,75
Coefficient de corrélation entre les x et les y	0,816
Équation de la droite de régression linéaire	$y = 3 + 0,5x$
Somme des carrés des erreurs relativement à la moyenne	110,0



La datasaurus dozen



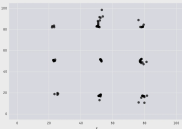
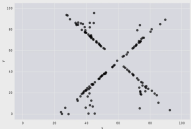
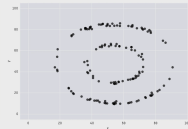
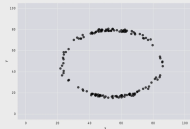
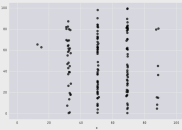
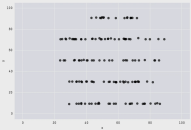
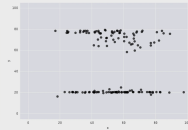
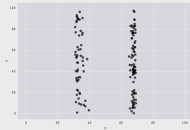
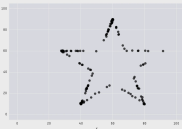
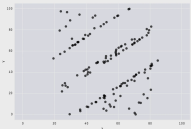
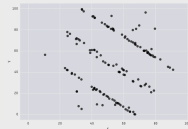
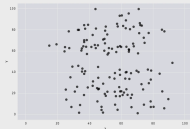
X Mean: 54.26

Y Mean: 47.83

X SD : 16.76

Y SD : 26.93

Corr. : -0.06



Autres motivations

- la vision est notre sens dominant
- on est forts pour reconnaître des motifs visuels
- on a souvent besoin de voir pour expliquer, raisonner, décider

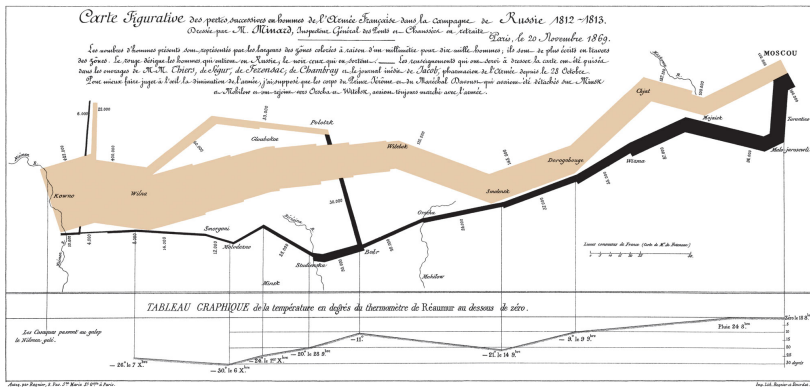
Mais aussi :

- aider notre mémoire (en la déchargeant au profit du cortex visuel)
- faciliter la surveillance d'événements potentiellement nombreux
- permettre l'exploration d'un espace de paramètres
- interface : limiter la quantité de pages à parcourir

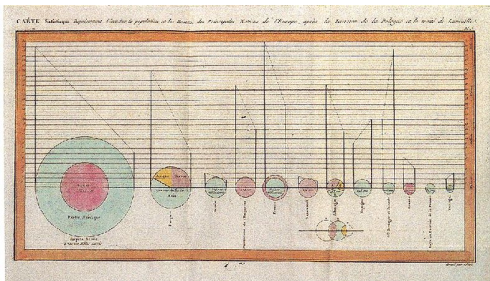
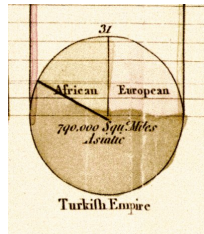
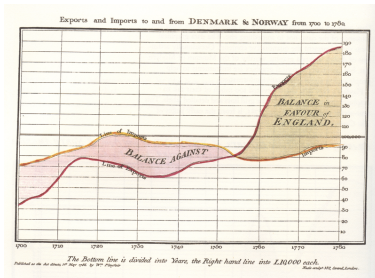
Histoire, exemples

La carte de Minard (1867)

- Charles Minard (1781–1870)
- statistiques sur la taille de l'armée, la température, le temps



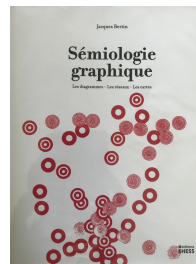
Fondateur : William Playfair (1759–1823)



Jacques Bertin

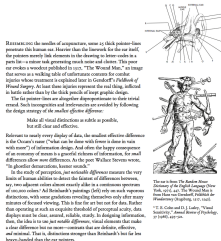
- Jacques Bertin (1918–2010).
Père de la Sémiologie graphique (1967)

L'EFFICACITÉ est définie par la proposition suivante :
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.

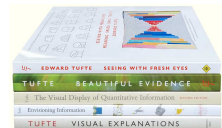


Edward Tufte

- Edward Tufte (1942–, t-œuf-ti)
- nombreux ouvrages, critique de Powerpoint



- sparklines



Aujourd'hui

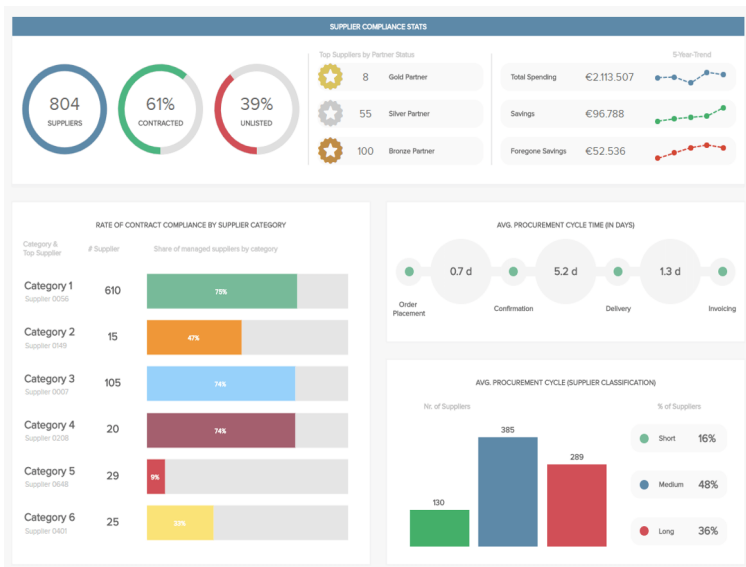
- la visualisation de données est omniprésente
- nombreux contextes, spécialisés
 - centre de contrôle (cockpit, salle, trading)
 - sciences (toutes!)
 - visualisation logicielle



Exemples



Exemples

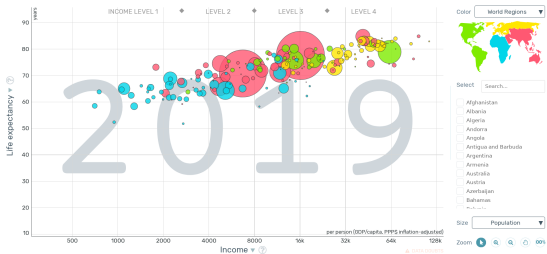


Examples



Hans Rosling

- 1948-2017, médecin, statisticien suédois
- <https://www.gapminder.org/tools/>



TrashTrack

- <http://senseable.mit.edu/trashtrack>

Principes

Définition

Visual representations of the semantics, or meaning, of information. In contrast to scientific visualization, information visualization typically deals with nonnumeric, nonspatial, and high- dimensional data.
(Chen, 2005)

Information visualization (InfoVis) produces (interactive) visual representations of abstract data to reinforce human cognition and perception; thus enabling the viewer to gain knowledge about the internal structure of the data and causal relationships in it.
<http://www.infovis-wiki.net/index.php/>

Qu'est-ce qu'une représentation ?

- un système formel, avec lequel l'information peut être spécifiée
- un système de signes, qui signifient autre chose qu'eux-mêmes
- exemple : le nombre 34

34

decimal

100010

binary

XXXIV

roman

Présentation

- chaque représentation révèle des aspects différents de l'information
 - décimal : compter, informer sur les puissances de 10
 - binaire : compter, informer sur les puissances de 2
 - romains : impressionner les amis

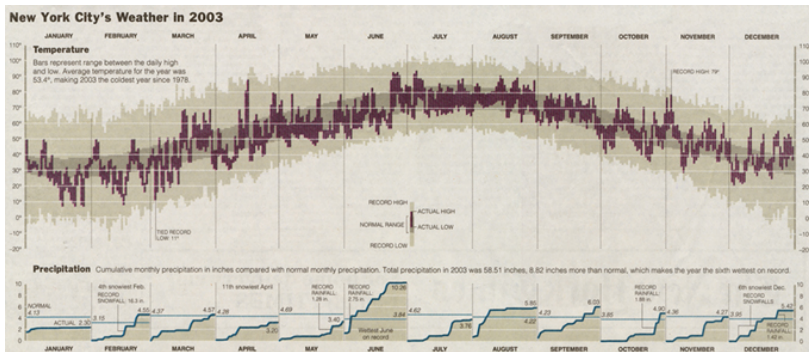
34, 34, 34

Principes d'excellence graphique

- présentation bien conçue de données intéressantes (contenu, design)
- idées complexes communiquées avec clarté, précision, efficacité
- grand nombre d'idées communiquées en peu de temps et avec une économie d'encre et d'espace
- (en général) multiples variables
- ne trompe pas sur les données

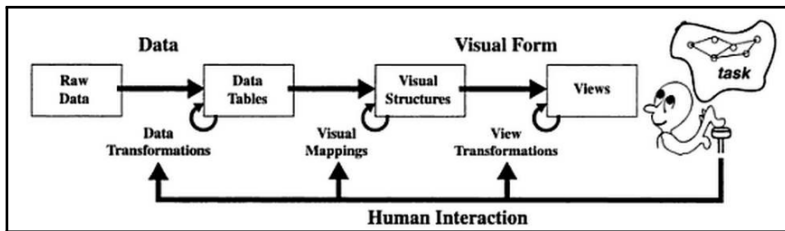
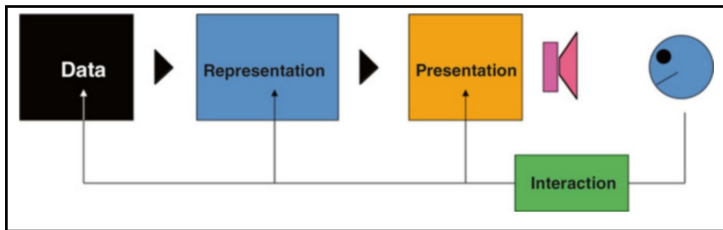
(Tufte)

Météo New York (Tufté)

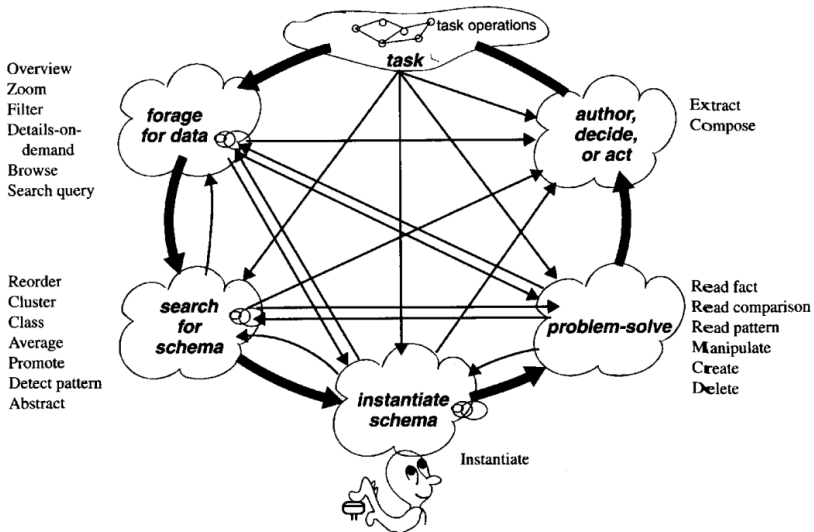


It shows daily high and low temperatures for 2003, for a normal year, and for record days, along with cumulative monthly precipitation. Some interesting inferences about average, variation, and rates of change over the seasons can be made visually. Note the flat part of the cycle through mid-December to mid-February and then a fairly rapid rise in temperature mid-February to May. And so on.

Pipeline



Pipeline



Risques

- choisir des données fausses/de mauvaise qualité
- prendre une structure de données inadaptée
- éliminer des données importantes
- montrer les mauvaises choses
- mauvaise représentation
- mauvaise présentation
- interactions inadaptées à l'exploration

Données

Données

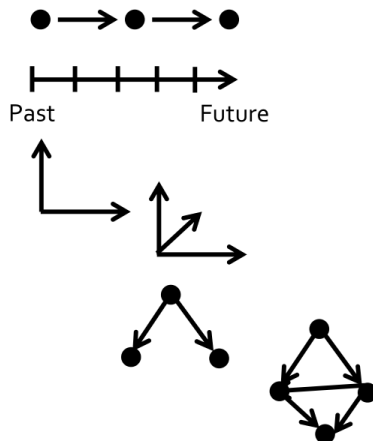
- Les données sont le cœur de la visualisation
- il faut impérativement :
 - comprendre les propriétés des données
 - savoir quelles méta-données sont disponibles
 - savoir ce que l'on veut extraire des données

Types de données

- Catégorielles
 - fruits : pommes, oranges
- Ordonnées
 - qualité : A, AA,
 - peut être compté et ordonné, mais pas mesuré
- Quantitatives
 - intervalles; fractions
 - on peut faire des calculs dessus

Axes de données

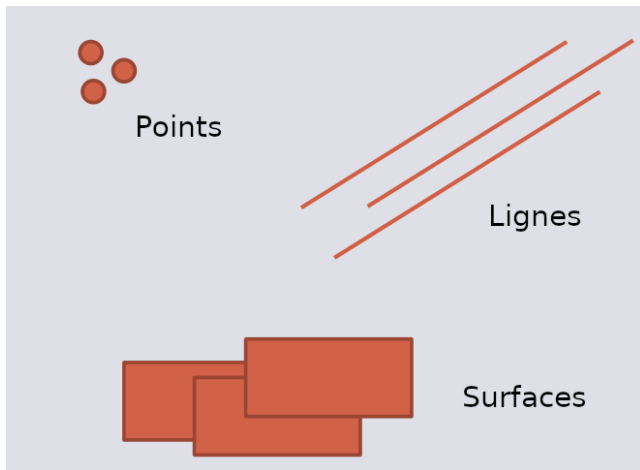
- 1D (linéaire)
- temporelles
- 2D (cartes)
- 3D
- nD
- arbres, graphes



- multiples manières de représenter!
- quelques contraintes (temps, arbres)

Composants de la visualisation

Composants de la visualisation

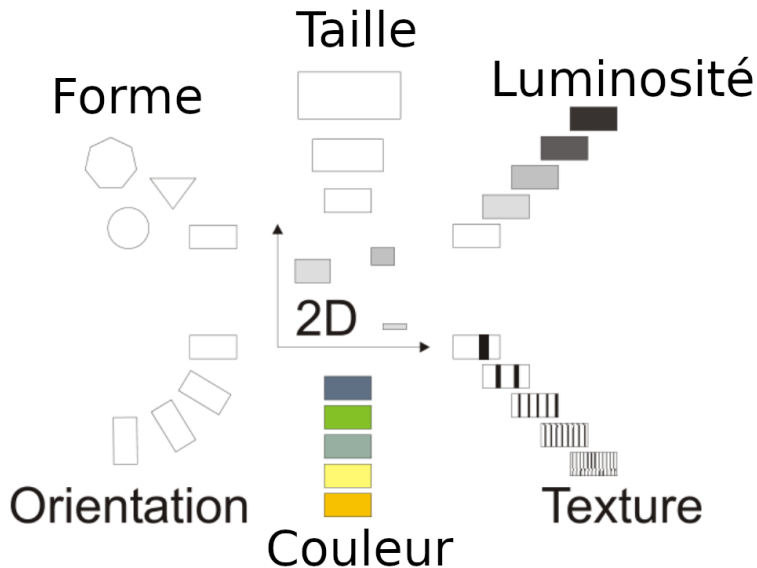


La sémiologie graphique, J. Bertin (1965)

Points – Lignes – Aires

- Un point représente une **position** dans le plan qui n'a pas de longueur ou de surface, sa signification est indépendante de la taille et de la forme du signe qui le rend visible.
- Une ligne représente un phénomène qui a une longueur mesurable mais pas une aire. Sa signification est indépendante de l'épaisseur et des caractéristiques du trait qui la rend visibles. (**lien, frontière**)
- Une surface représente quelque chose qui a une taille mesurable, indiquée par l'aire couverte. Une aire peut changer de position, mais pas de taille, d'orientation ou de forme sans que cela affecte la signification perçue.

Variables



Variables

Classification de Bertin (1965)

Variables rétiniennes :

but du graphique :

Taille 

Valeur 

Grain 

Couleur 

Orientation 

Forme 

association
de catégories  G C O F

selection
d'une catégorie  T V G C

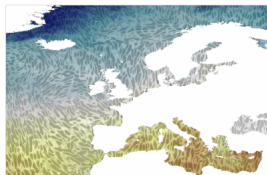
comparaison
d'ordres  T V G

comparaison
de quantités  T

Variables étendues (ordinateurs)

Saturation ○ ● ● ● ● ● + Transparence
Luminosité ● ● ● ● ● ●

Champs texturés



(Klebnikov et al, 2012)

Flou



(Kosara et al., 2002)

+ clignotement + animation en général

Variables étendues (ordinateurs)

- flicker
 - frequency, rhythm, appearance
- depth? 'quasi' 3D
 - depth, occlusion, aerial perspective, binocular disparity
- Illumination
- transparency



Caractéristiques des variables

	POINTS	LIGNES	ZONES				
XY 2 DIMENSIONS DU PLAN				Q	O	≠	≡
Z TAILLE				Q	O	≠	≠
VALEUR					O	≠	≠
LES VARIABLES DE SÉPARATION DES IMAGES							
GRAIN				O		≠	≡
COULEUR						≠	≡
ORIENTATION						≠	≡
FORME						≠	≡

Q	Quantitative (proportionnelle)
O	Ordonnée
≠	Sélective (différentielle et instantanée)
≡	Associative (visibilité inchangée des autres Variables Visuelles)
≠	Dissociative (visibilité variable des autres Variables Visuelles)

Représentations

Représentation de données multivariées

- Multivariées : ≥ 3 facteurs
- Exemple : automobile = (prix, places, conso., vitesse, ...)
- 1 ou 2 facteurs : tableur + grapheur (Excel, Calc, Sheets, etc.)

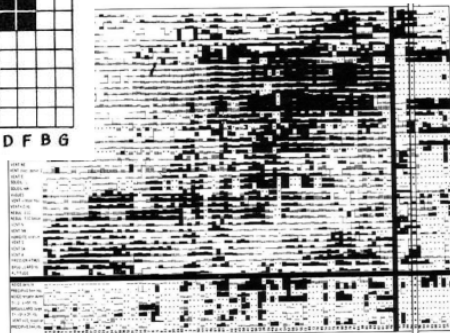
Tables

l							
m							
n							
o							
p							
q							
r							
s							
t							
u							
v							
w							
	A	B	C	D	E	F	G



n							
s							
t							
v							
l							
q							
u							
m							
o							
w							
p							
r							
	C	H	A	E	D	F	B

Bertin



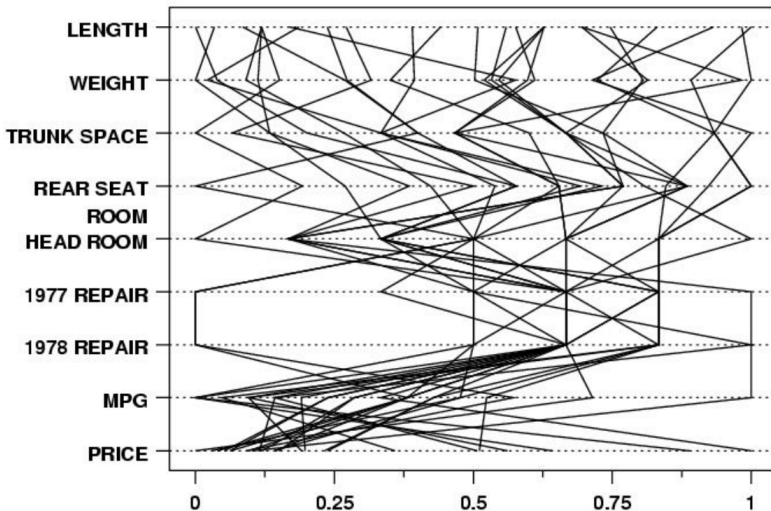
3

95

[illegible]

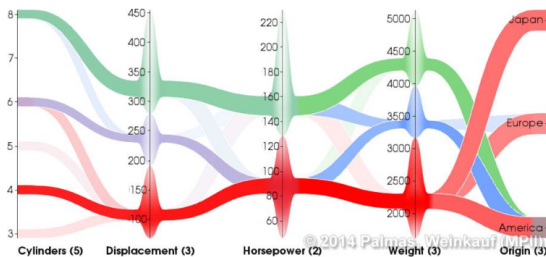
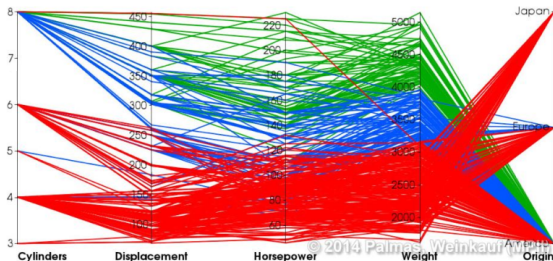
Hsu & Cubaud, HCI' 2009

Parallel coordinate plot



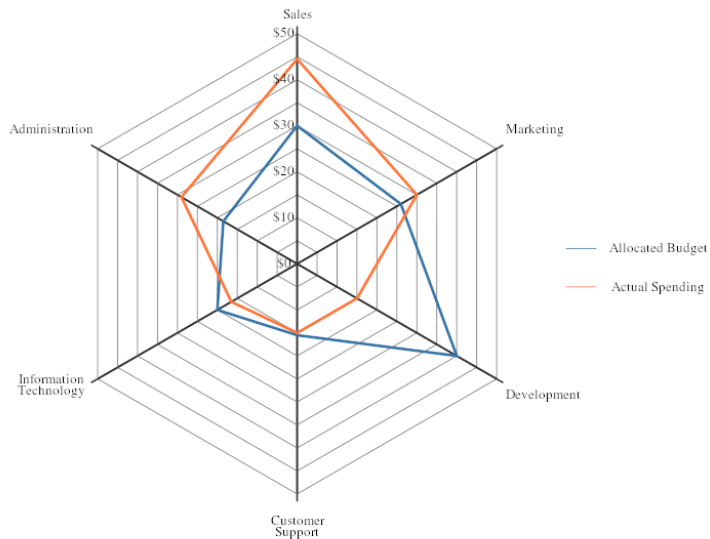
A. Inselberg, 1959

Parallel coordinate plot : groupement



<http://www.csc.kth.se/~weinkauff/>

Diagramme en étoile



Petits multiples

Série de graphes similaires, avec des axes et échelles communs, permettant une comparaison simple

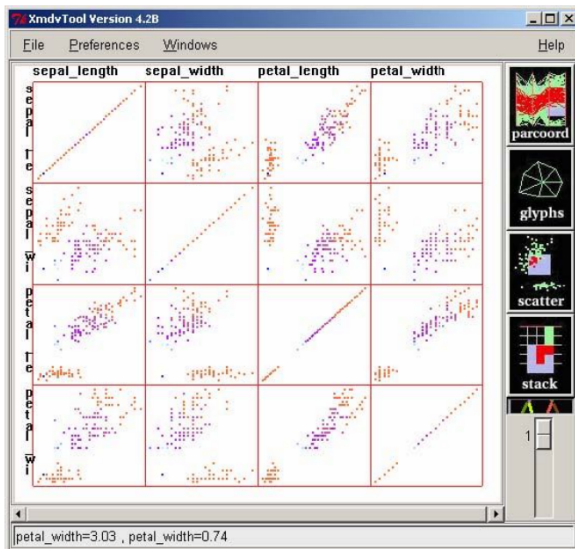
Small multiples reflect much of the theory of data graphics:

For non-data-ink, less is more.

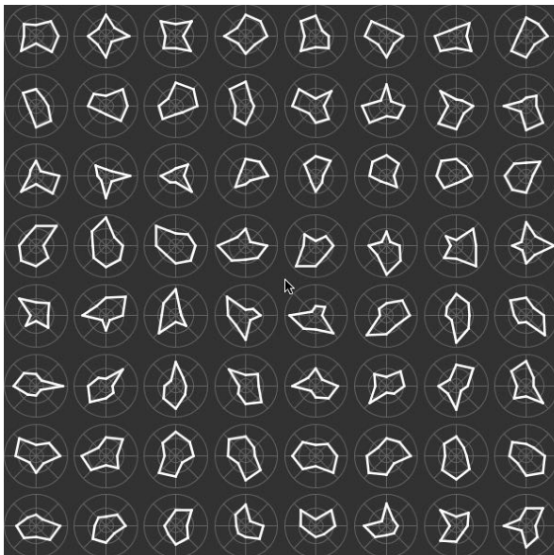
For data-ink, less is a bore.⁶

⁶The two aphorisms on the meaning of “less” are, respectively, credited to Ludwig Mies van der Rohe and to Robert Venturi, *Complexity and Contradiction in Architecture* (New York, second edition, 1977), p. 17.

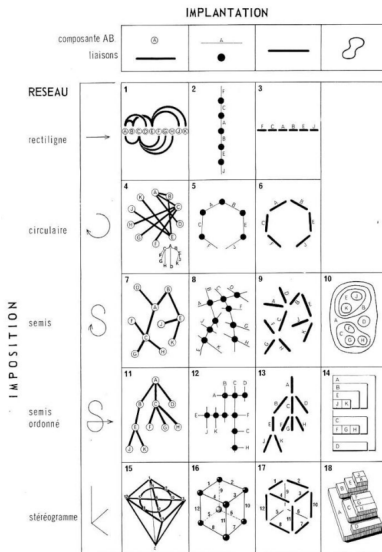
Petits multiples



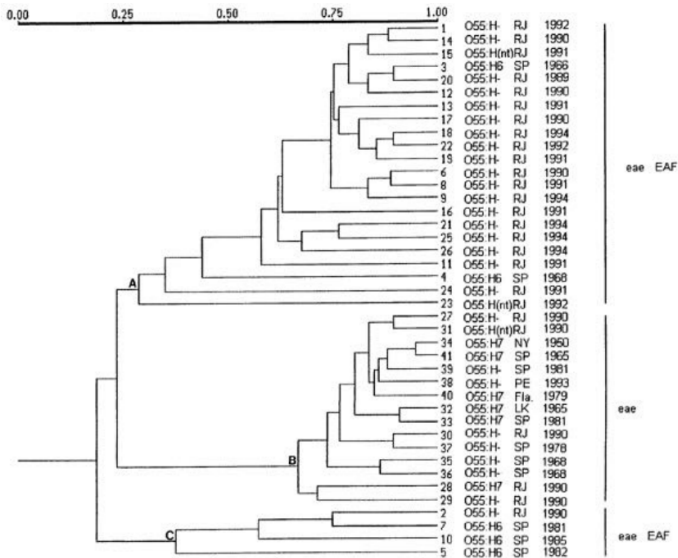
Petits multiples



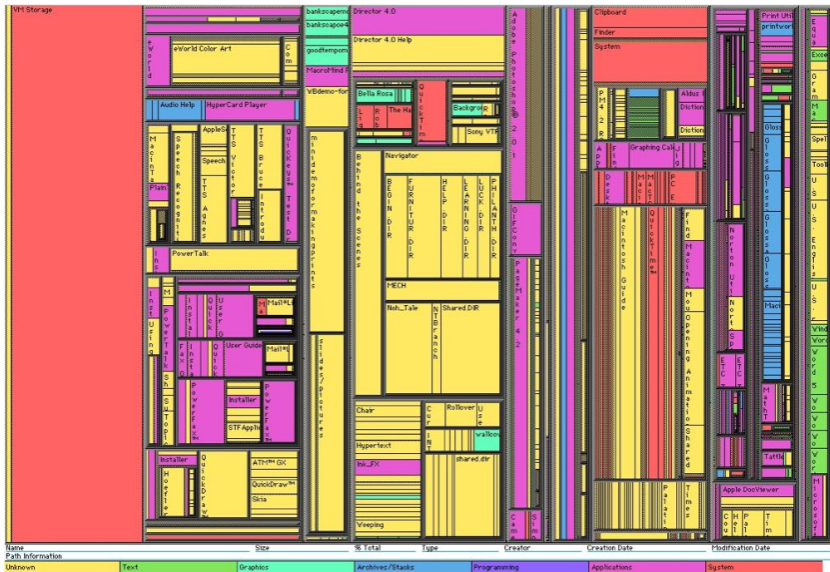
Graphes (Bertin)



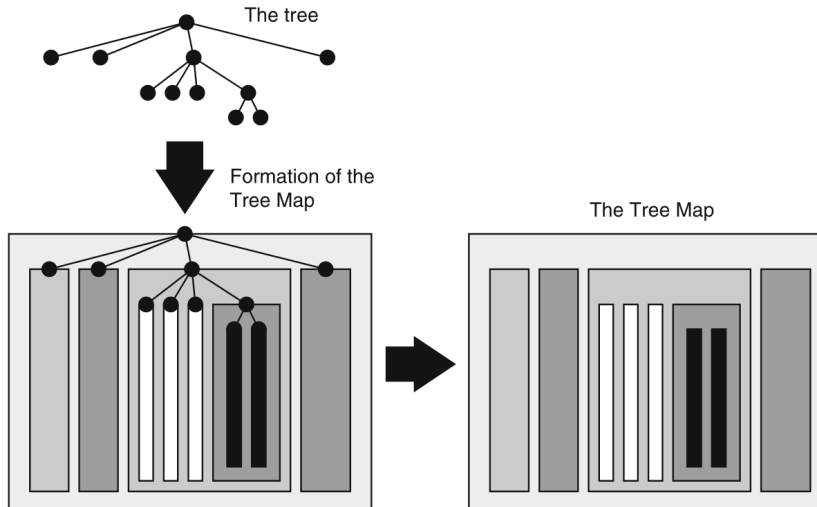
Dendrogrammes



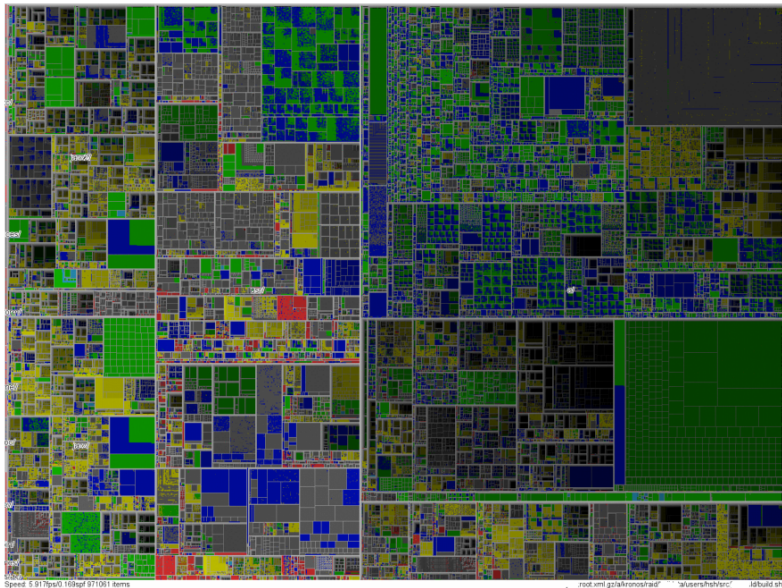
Treemaps



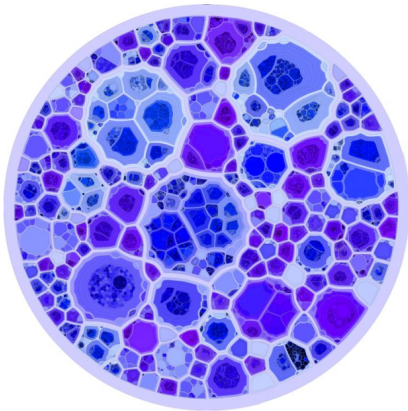
Treemaps : principe



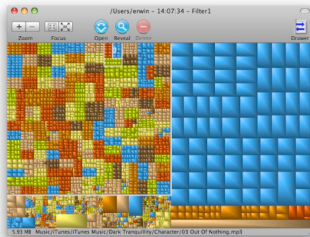
Treemaps



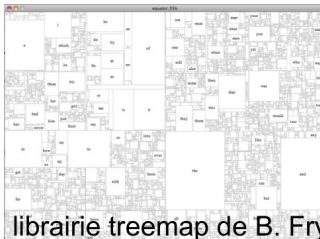
Treemaps



<http://www.informatik.uni-konstanz.de/deussen/mitarbeiter/oliver-deussen/olivers-kunstversuche/>



<http://grandperspectiv.sourceforge.net/>



librairie treemap de B. Fry
pour Processing

Interaction

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Overview first, zoom and filter, then details-on-demand

Visual information seeking mantra

The seven tasks are at a high level of abstraction. More tasks and refinements of these tasks would be natural next steps in expanding this table. The seven tasks are:

Overview: Gain an overview of the entire collection.

Zoom : Zoom in on items of interest

Filter: filter out uninteresting items.

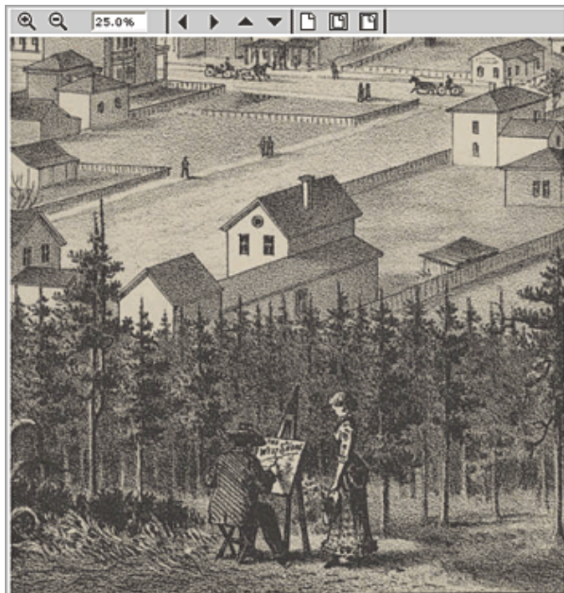
Details-on-demand: Select an item or group and get details when needed.

Relate: View relationships among items.

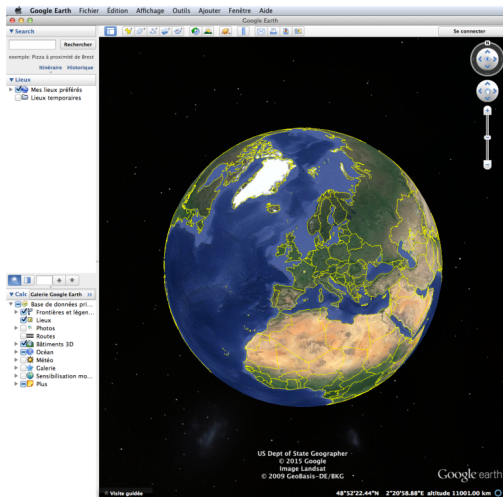
History: Keep a history of actions to support undo, replay, and progressive refinement.

Extract: Allow extraction of sub-collections and of the query parameters.

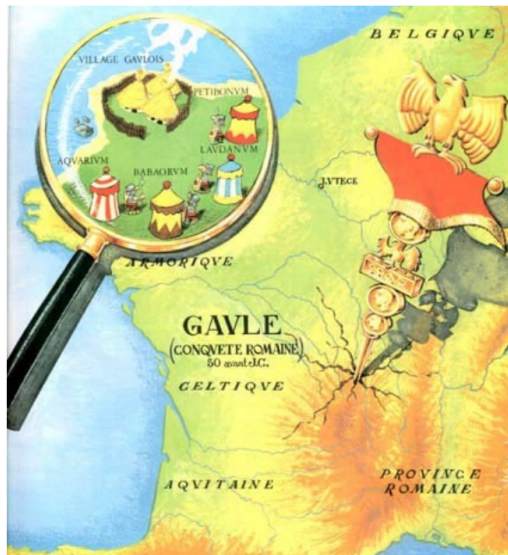
Vue globale et détails



Google earth



Distorsion



Distorsion : principe

Principe

A Review and Taxonomy of Distortion-Oriented Presentation Techniques

Y. K. LEUNG
Swinburne University of Technology
and
M. D. APPERLEY
Massey University

One of the common problems associated with large computer-based information systems is the relatively small window through which an information space can be viewed. Increasing interest

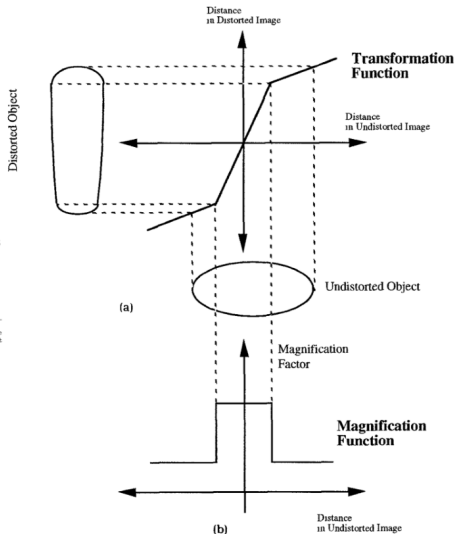
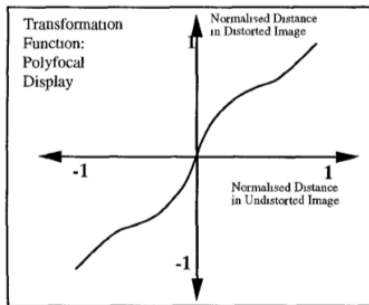
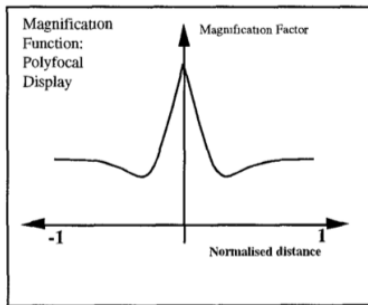


Fig. 3. (a) The transformation of an elliptic object by applying the transformation function of a Bifocal Display in one dimension; (b) the corresponding magnification function of the Bifocal Display.

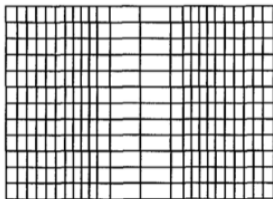
Distorsion : principe



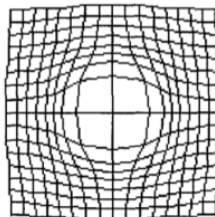
(a)



(b)

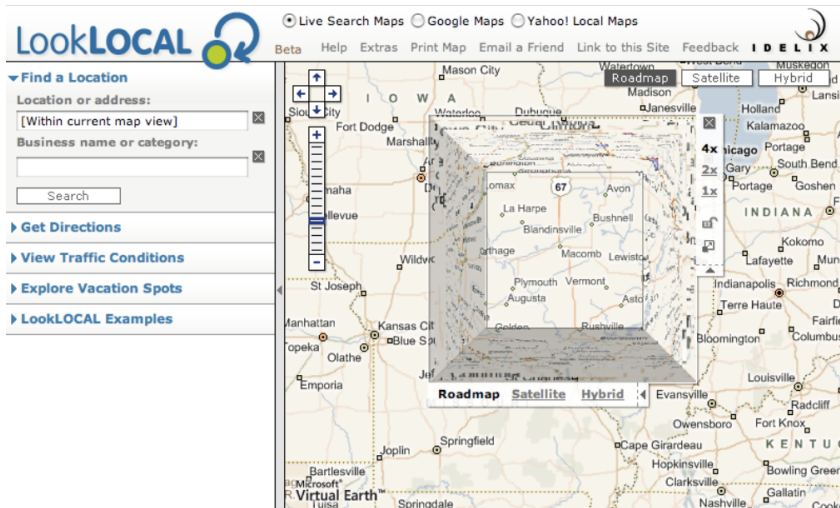


(c)



(d)

Distorsion : cartes



Distorsion : cartes



(a)



(b)

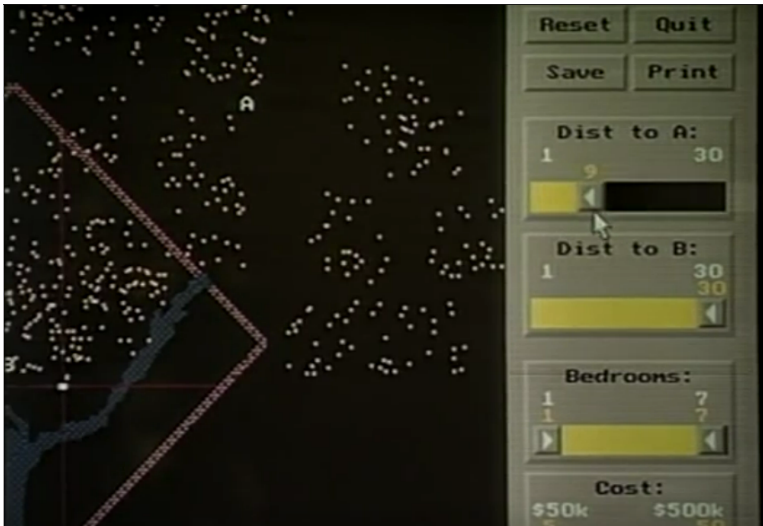


(c)



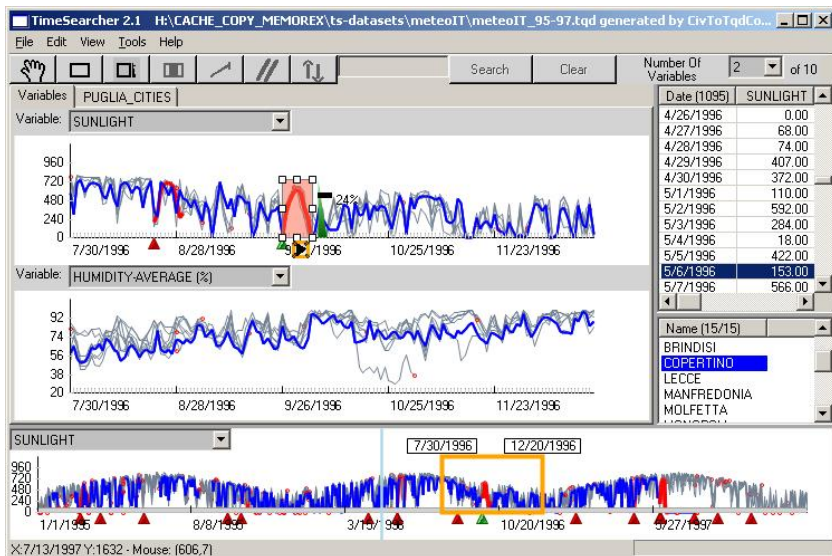
(d)

Filtrage



- vidéo : <https://www.youtube.com/watch?v=5X8XY9430fM>

Filtrage



Filtrage

Insight Table Lens - [foremost.txt]

File Edit View Tools Options Window Help

inxight

	Year	Quarter	Product	Channel	Region	Saleperson	Units	Revenue	Profits
126	1993	2	ForeCode Pro	Direct Sales	Southwest	Kevin Polen	1029	439898	171561
444	1993	4	ForeCode Pro	VAR	West	Tom Tuttle	302	122310	51371
445	1993	4	ForeCode Pro	VAR	West	Ann Thomas	302	122310	51371
446	1993	3	ForeMost S...	Direct Sales	Midwest	Sal Vitatone	301	2.8595e+006	929338
447	1993	3	ForeMost S...	VAR	South	Gary Copper	301	2.709e+006	948150

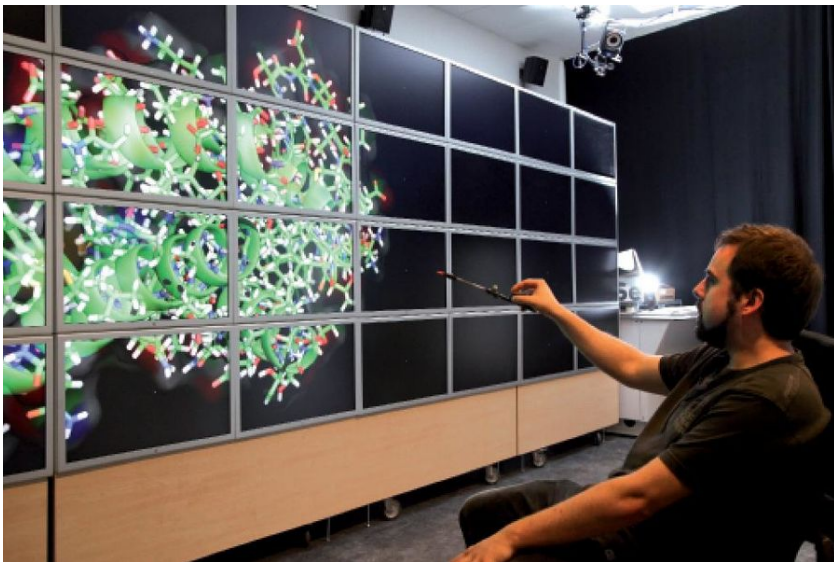
299

Key: 1992 Col: Profits Median: 30813.00000 Row: 1967, Col: 9

video

<https://www.youtube.com/watch?v=qWqTrRAC52U>

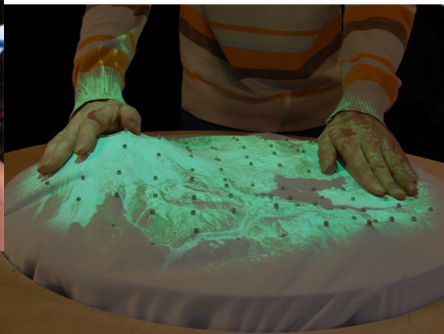
Grands écrans



Autres affichages



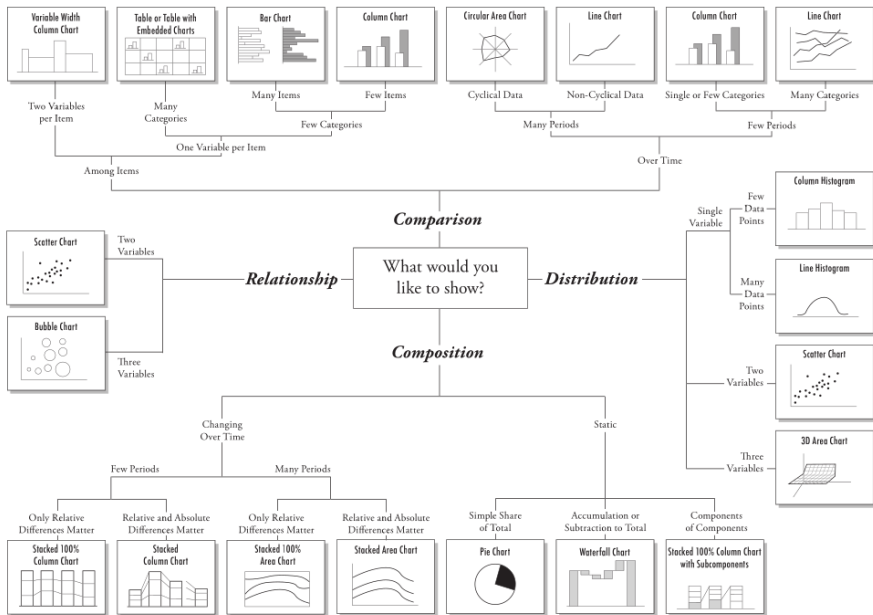
<http://www.aviz.fr/phys>



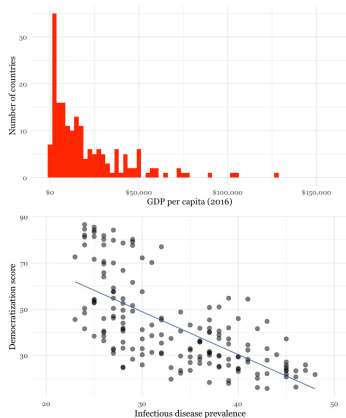
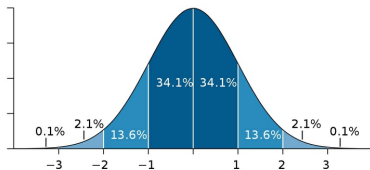
<http://tangible.media.mit.edu/project/relief/>

Bonnes pratiques

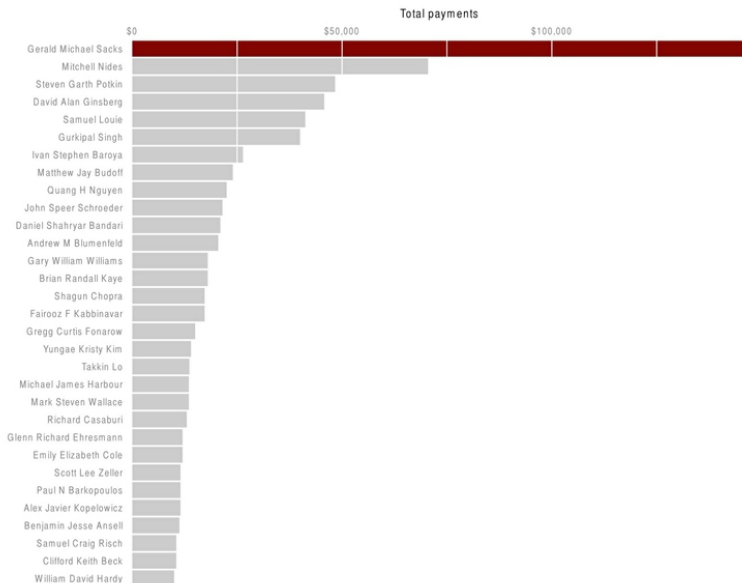
Choix du type de graphique



Distribution (d'une variable continue)

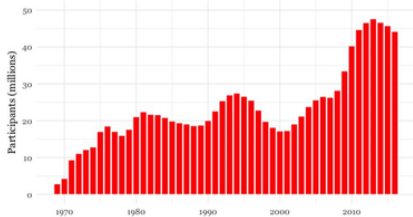


Comparaisons simples



Comparaisons temporelles

Column chart



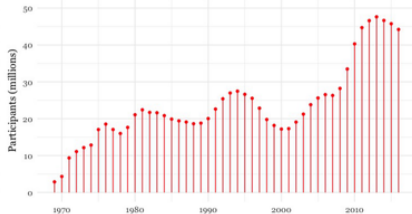
Line chart



Dot and line chart

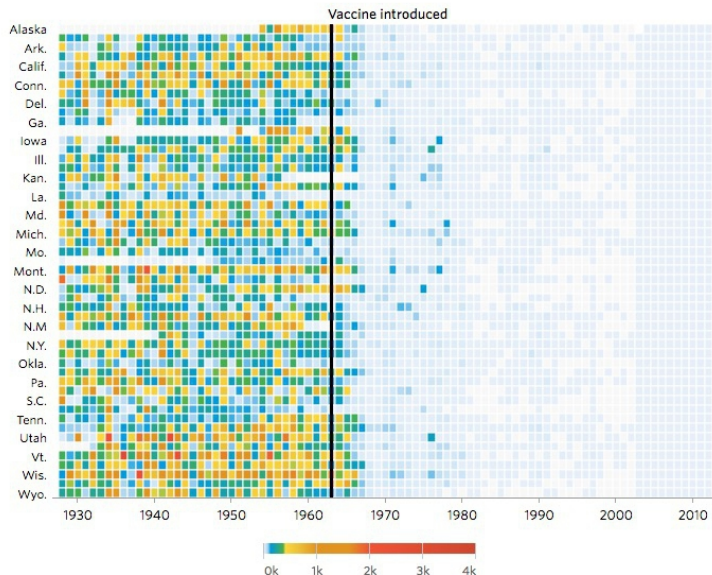


Dot and column chart

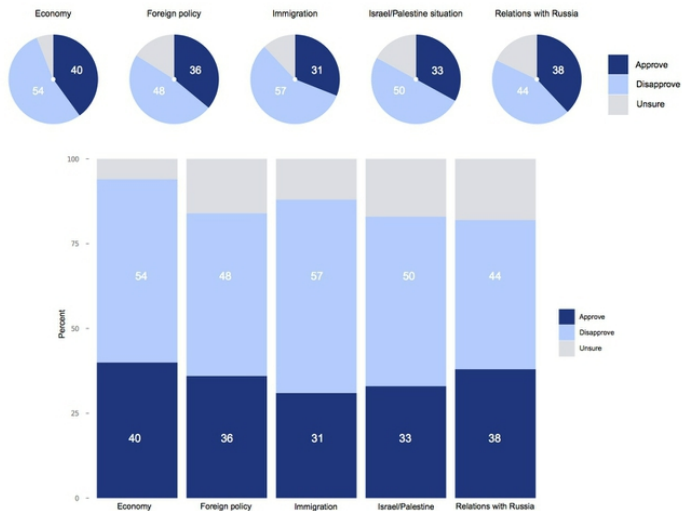


Comparaisons temporelles

Measles



Composition : parties d'un tout

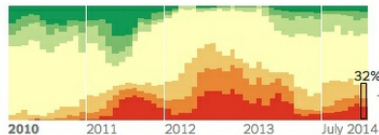


Composition : évolution

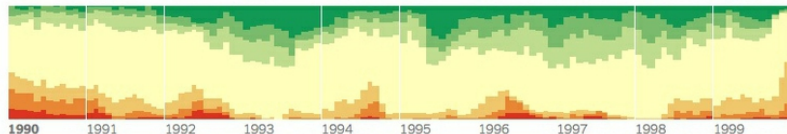
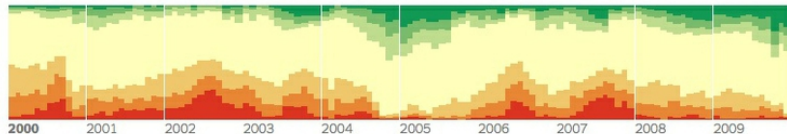
PALMER DROUGHT INDEX

1895-now

Wetter Drier

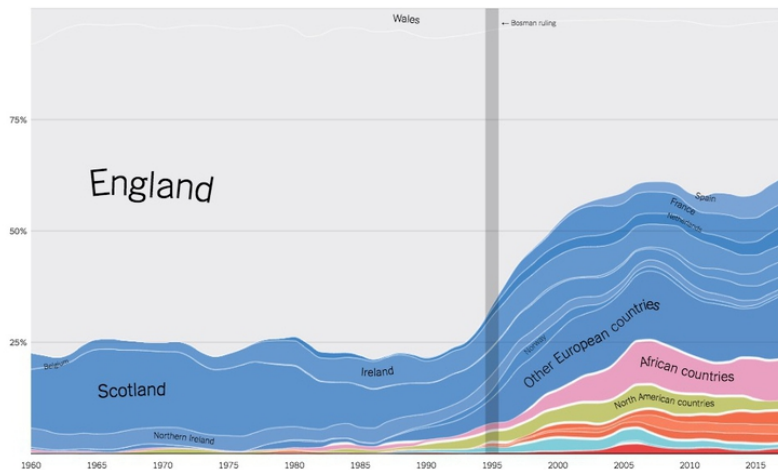


During July, 32 percent of the contiguous U.S. was in moderate to extreme drought.



Composition : évolution

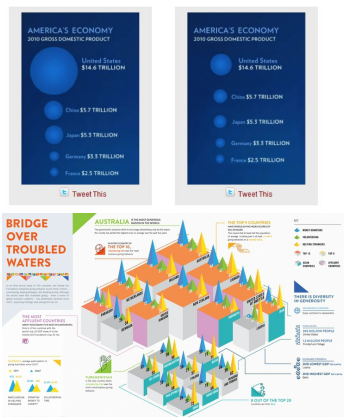
Where players in the **Premier League** have come from



Pièges à éviter

Quelques pièges

- si un carré de côté 1 représente une valeur de 1, une valeur de 4 doit être représentée par un carré de côté 2
- attention aux disques : il faut mettre à l'échelle selon la racine du rayon ($\pi \times R^2$)
- éviter la 3D pour des graphiques simples
- Clarté et simplicité
- Série de diagrammes (étapes)



"Everything should be made as simple as possible, but not simpler." (A. Einstein)

Erreurs humaines

- inférer des tendances à partir du bruit
- voir des motifs
- signification d'une corrélation (http://tylervigen.com/view_correlation)
- "significatif"
- être graphiste est (aussi) un métier!

Sujets avancés

Visualiser du texte

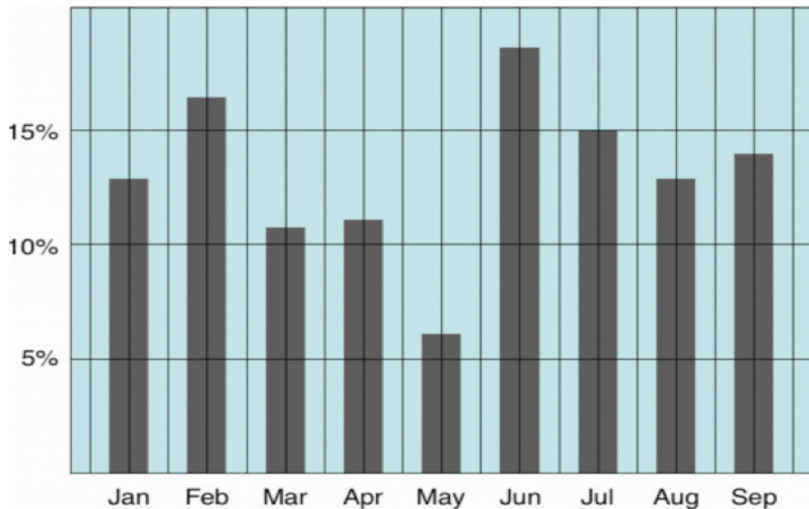
- pour comprendre (un doc), grouper (des docs), comparer
 - beaucoup de données, bruitées (polysémie, généré)
 - pas pré-attentif
-
- beaucoup de données prêtes pour les machines
 - des techniques simples donnent de bons résultats (tag cloud, word count, graphes de mots)

`http://textvis.lnu.se`

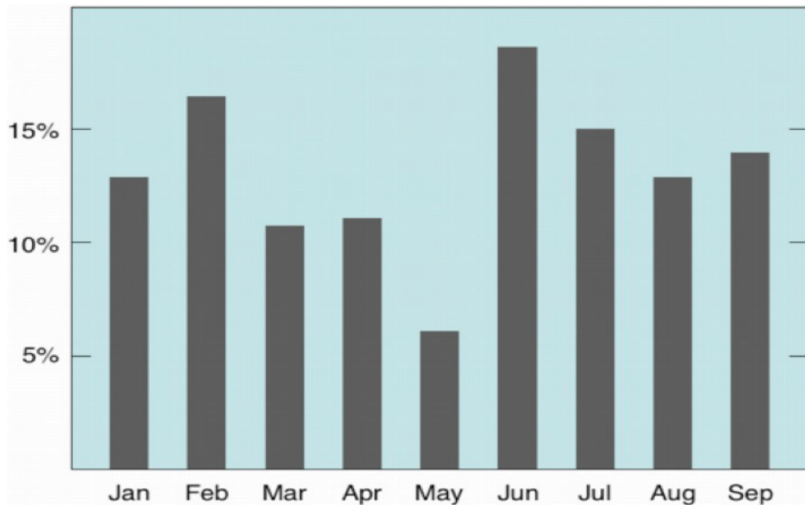
Couleurs

- Avoir toujours un contraste fort entre arrière et avant-plan
- éviter 12 couleurs, 5 est recommandé
- utiliser des couleurs distinctes
- utiliser des dégradés de couleurs
- <http://colorbrewer2.org/>
- attention au daltonisme

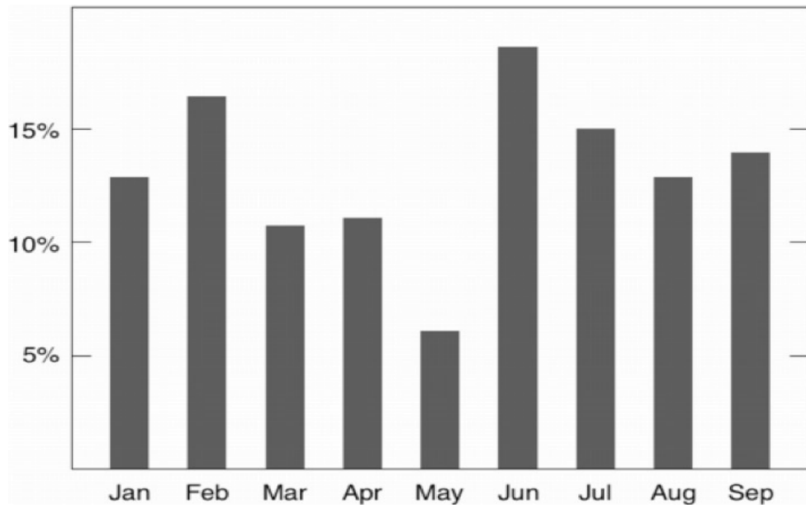
Data to ink ratio



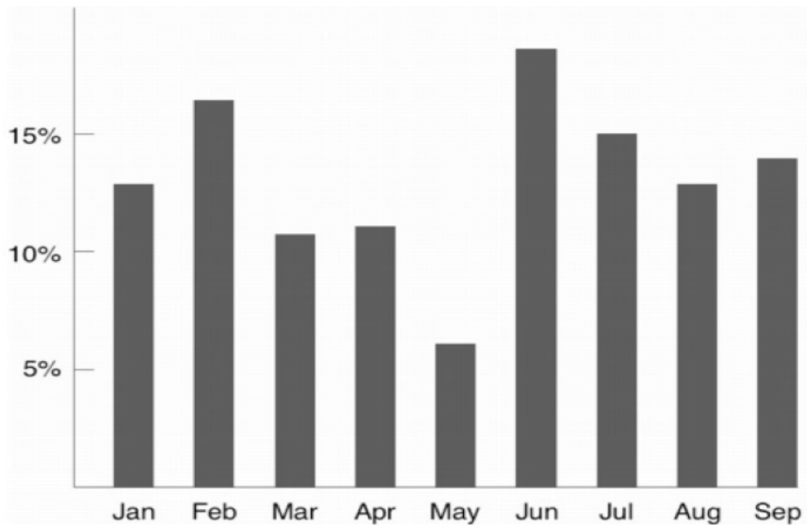
Data to ink ratio



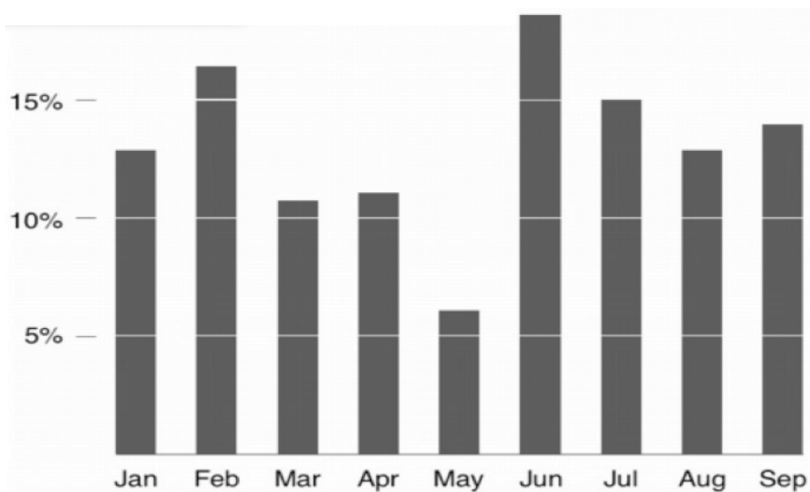
Data to ink ratio



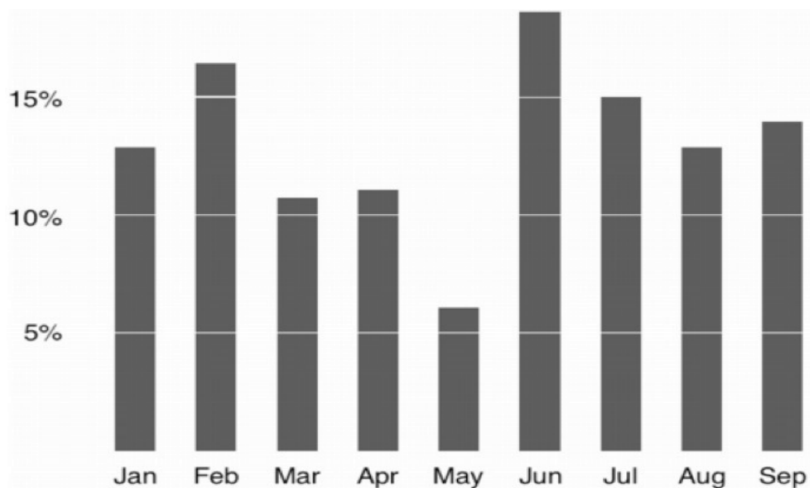
Data to ink ratio



Data to ink ratio



Data to ink ratio



Et aussi

- typographie
- agencements

Références

Références

Ce cours repose notamment sur les travaux de :

- Pierre Cubaud, Professeur au CNAM Paris
- Petra Isenberg & Anastasia Bezerianos, Inria/LRI (Univ. Paris-Sud/Saclay)
- Peter Aldhous <https://www.peteraldhous.com/>