Visualisation d'information (1)
historique, applications, outils

Pierre Cubaud <cubaud@cnam.fr>

mars 2022

le cnam

Plan de l'exposé

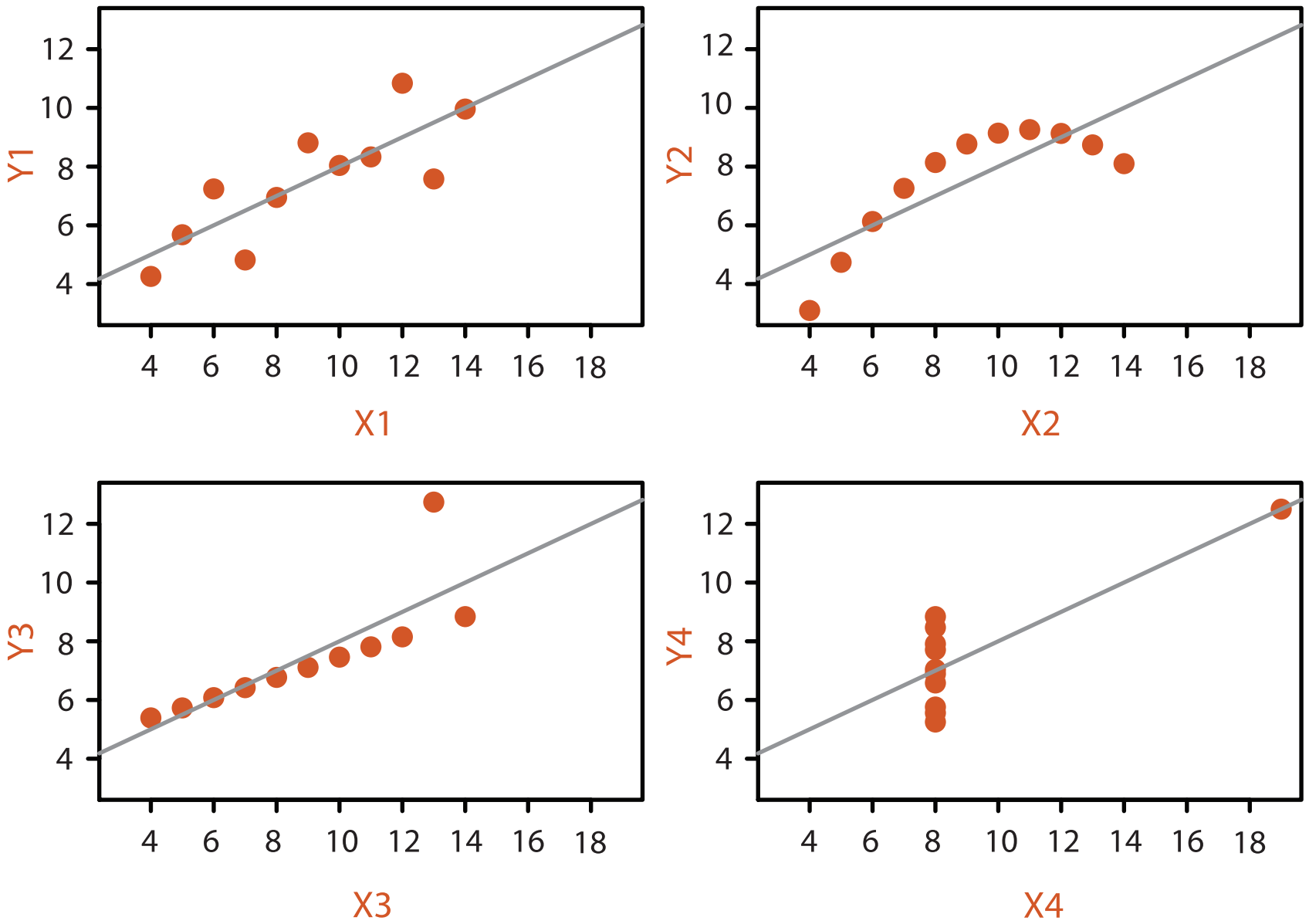
1. Exemples d'applications de la VI
2. Définition du sujet
3. Bertin et Tufte
4. Le Xerox PARC et le HCI lab
5. aujourd'hui : "Information is beautifull"
6. Logiciels

+ exemple récent des élections américaines

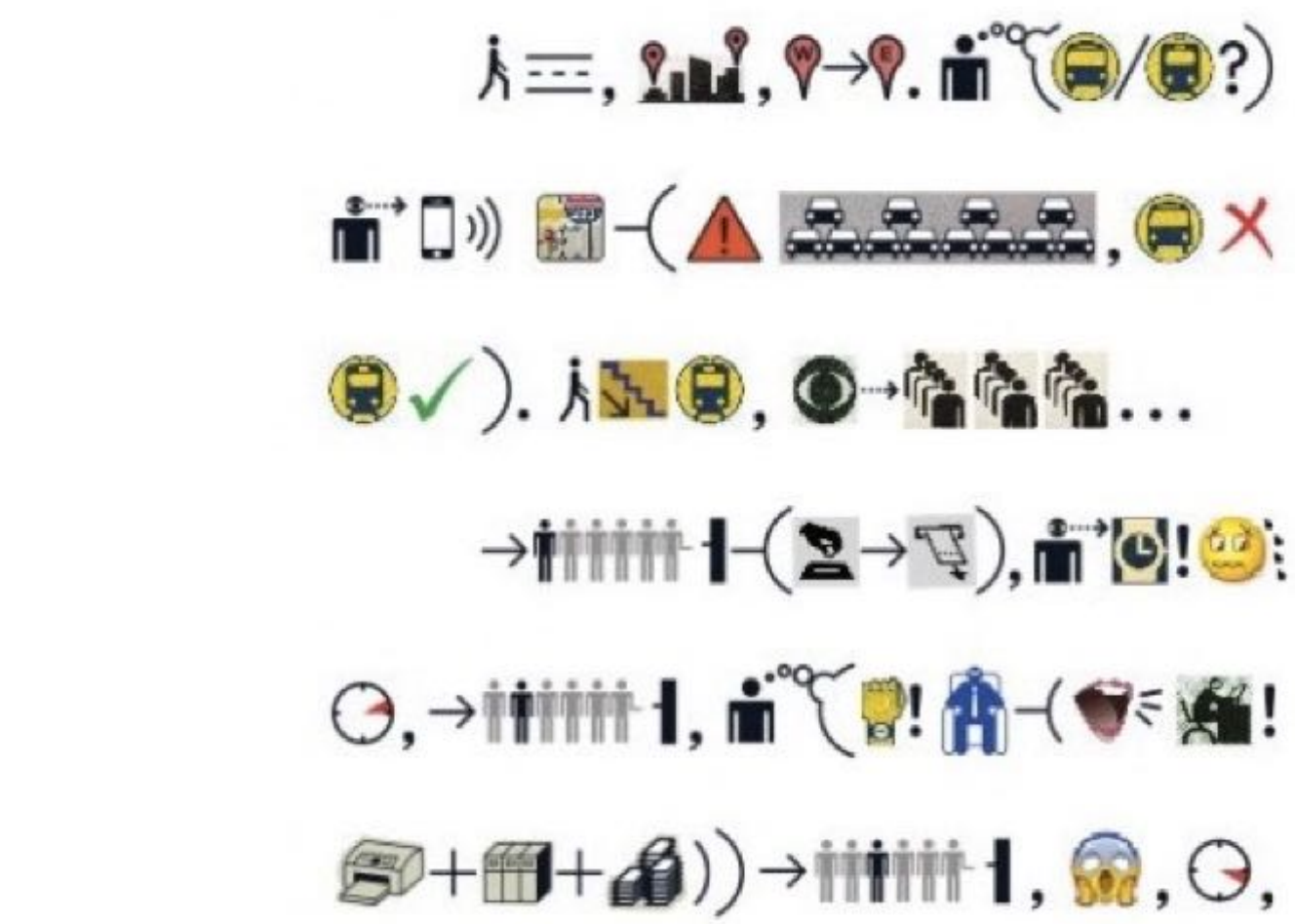
Pourquoi faire des graphiques ?

Anscombe's Quartet: Raw Data

	1		2		3		4	
	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
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	



https://fr.wikipedia.org/wiki/Quartet_d%27Anscombe



Xu Bing - Une histoire sans mot - Grasset, 2013

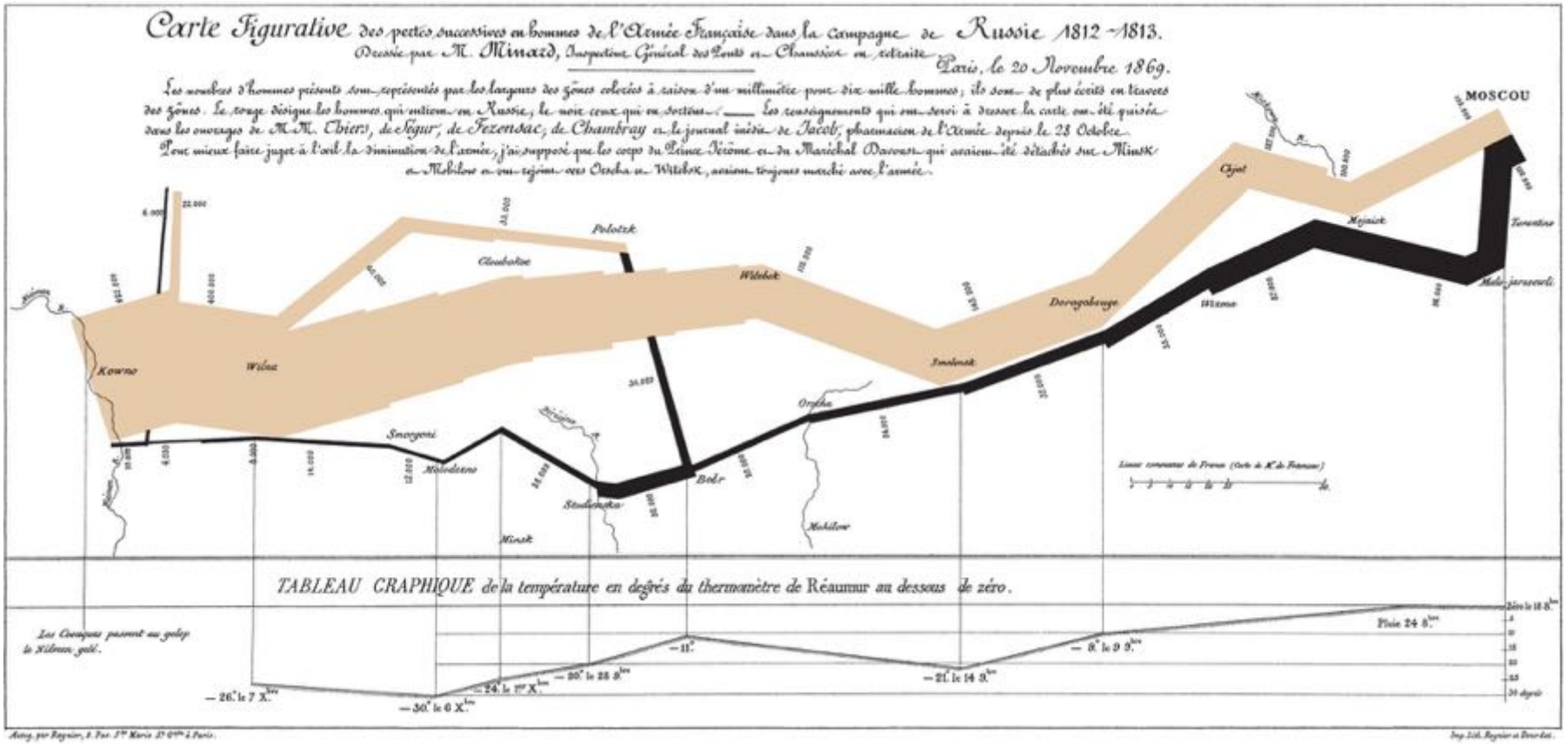
1. Exemples

choisir dans une grande masse d'informations

Pub dans Metro, nov 2007

« cover flow » Apple Ipod Touch

Autre exemple : comprendre un phénomène



Edward R. Tufte (1992). The Visual Display of Quantitative Information

Minard (1869) : carte augmentée par des flots
« graphics reveals data » (Tufte)

Le fondateur ? W. Playfair (1800)

ÉLÉMENS

DE STATISTIQUE,

Où l'on démontre, d'après un principe entièrement neuf, les ressources de chaque Royaume, État et République de l'Europe ; suivis d'un état sommaire des principales Puissances et Colonies de l'Indostan.

Orné de Cartes coloriées, représentant, d'un coup-d'œil, les forces physiques de toutes les Nations Européennes.

TRADUIT de l'anglais de Wm. PLAYFAIR ;

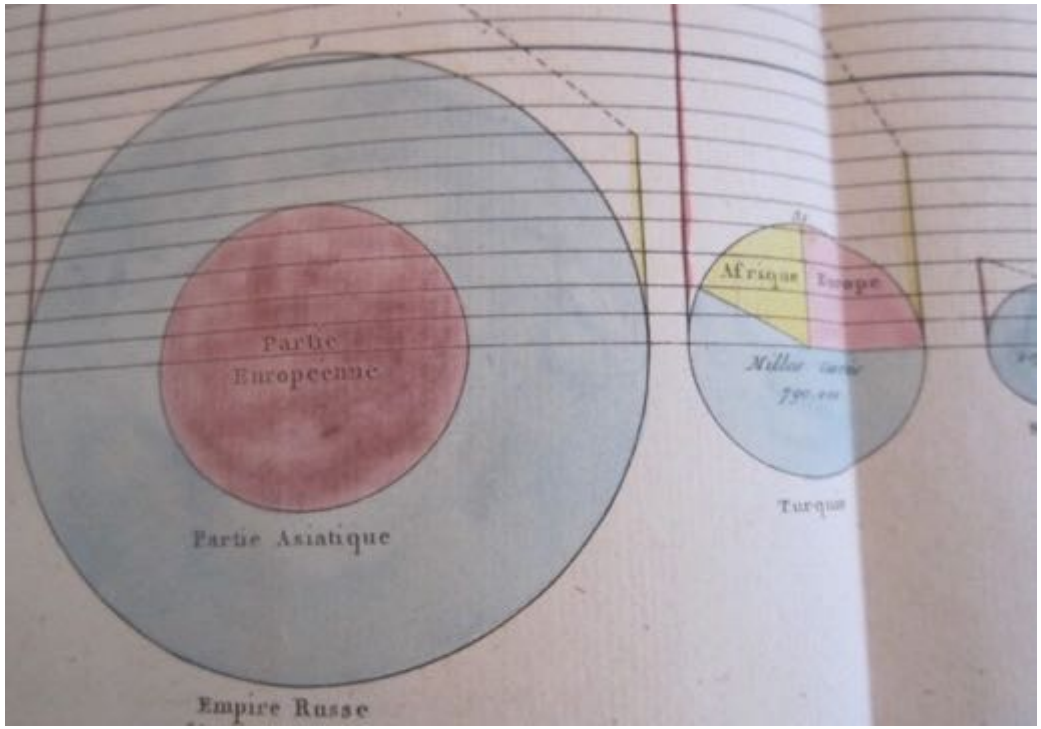
PAR Dlo. Fois. DONNANT, de l'Athénée des Arts, ci-devant interprète dans les États-Unis d'Amérique, traducteur des Élémens de l'organisation sociale.

On y a ajouté un Tableau comparatif de l'étendue et de la population de tous les Départemens de la France, un Précis statistique des États-Unis d'Amérique, un Essai sur la navigation intérieure de ce pays, un Tableau des principales divisions du nouveau continent, un aperçu des marchandises et denrées qui conviennent le mieux au commerce des Républiques Française et Américaine, etc., etc.

A PARIS,

Chez { BAILLIOT jne., Libraire, rue Hautefeuille, N°. 34.
GENETS jne., Libraire, rue de Thionville, N°. 1846, près le Pont-Neuf.

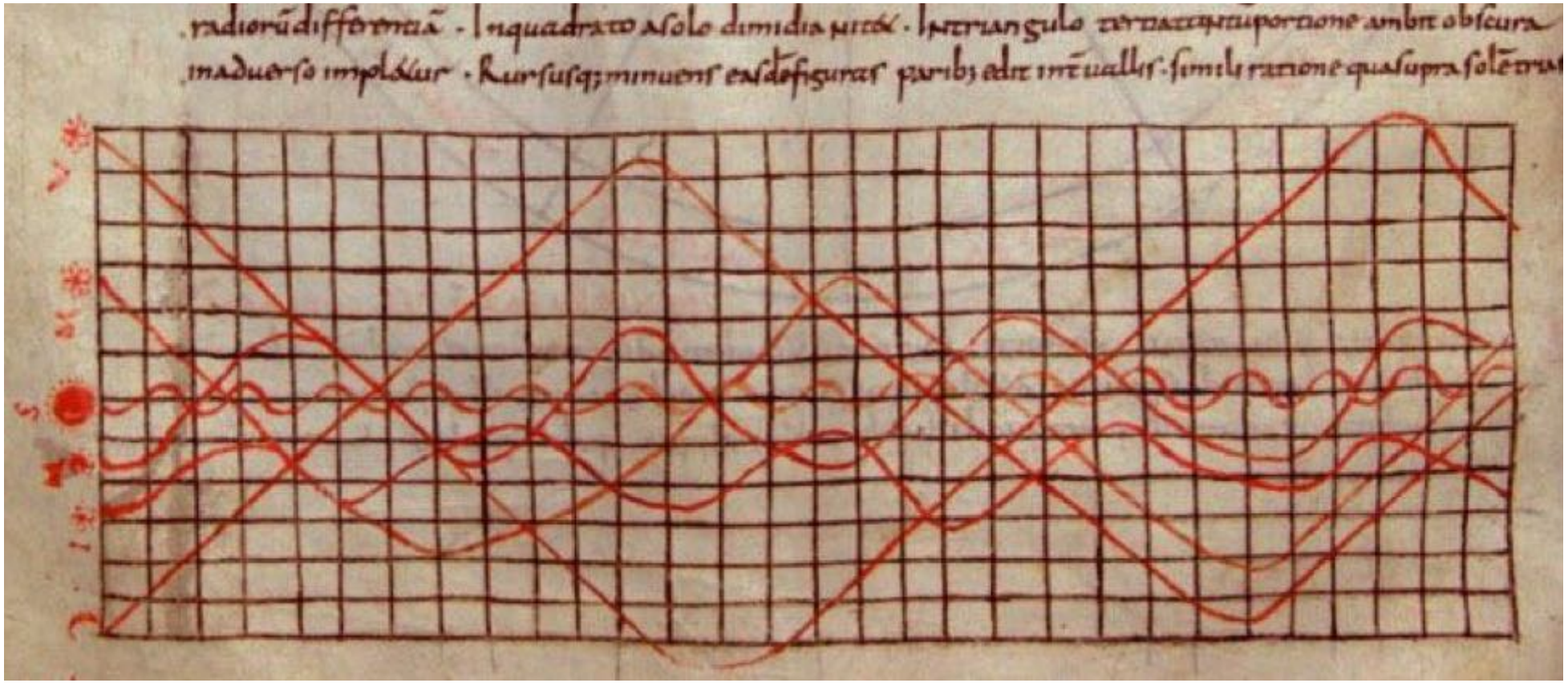
An XI — (1802.)



Playfair, Elémens de statistique, Economie, EO,
Mathématiques, Etats-Unis, les premiers diagrammes !

Etat de l'objet : –
Vente terminée : 24 sept. 2013 21:16:30 Paris
Meilleure enchère : 755,00 EUR [21 enchères]
Ajouter à la liste
Livraison : 10,00 EUR Standard | Détails

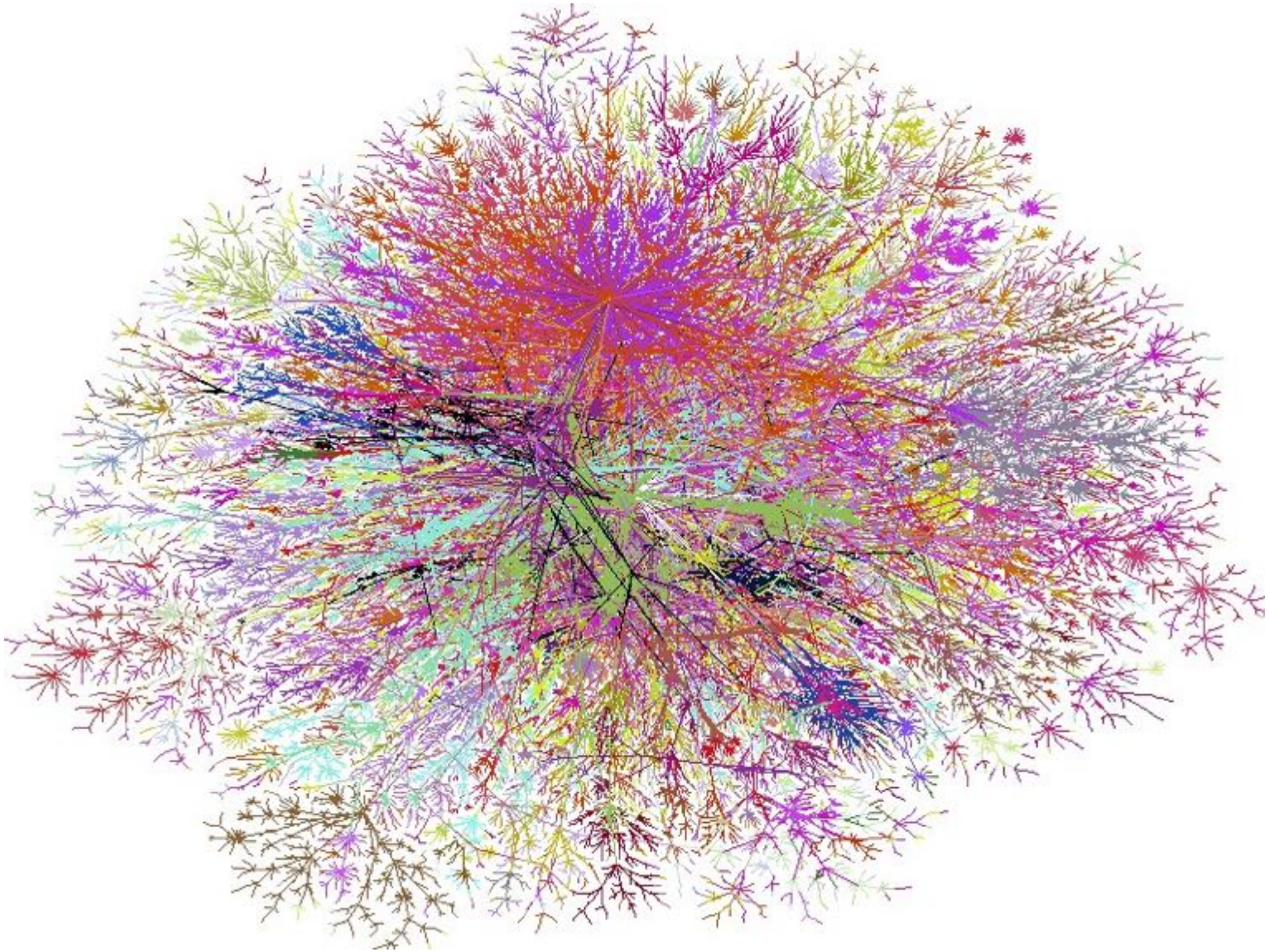
Manuscrit du Xeme s.
chronogramme des latitudes des planètes et du soleil



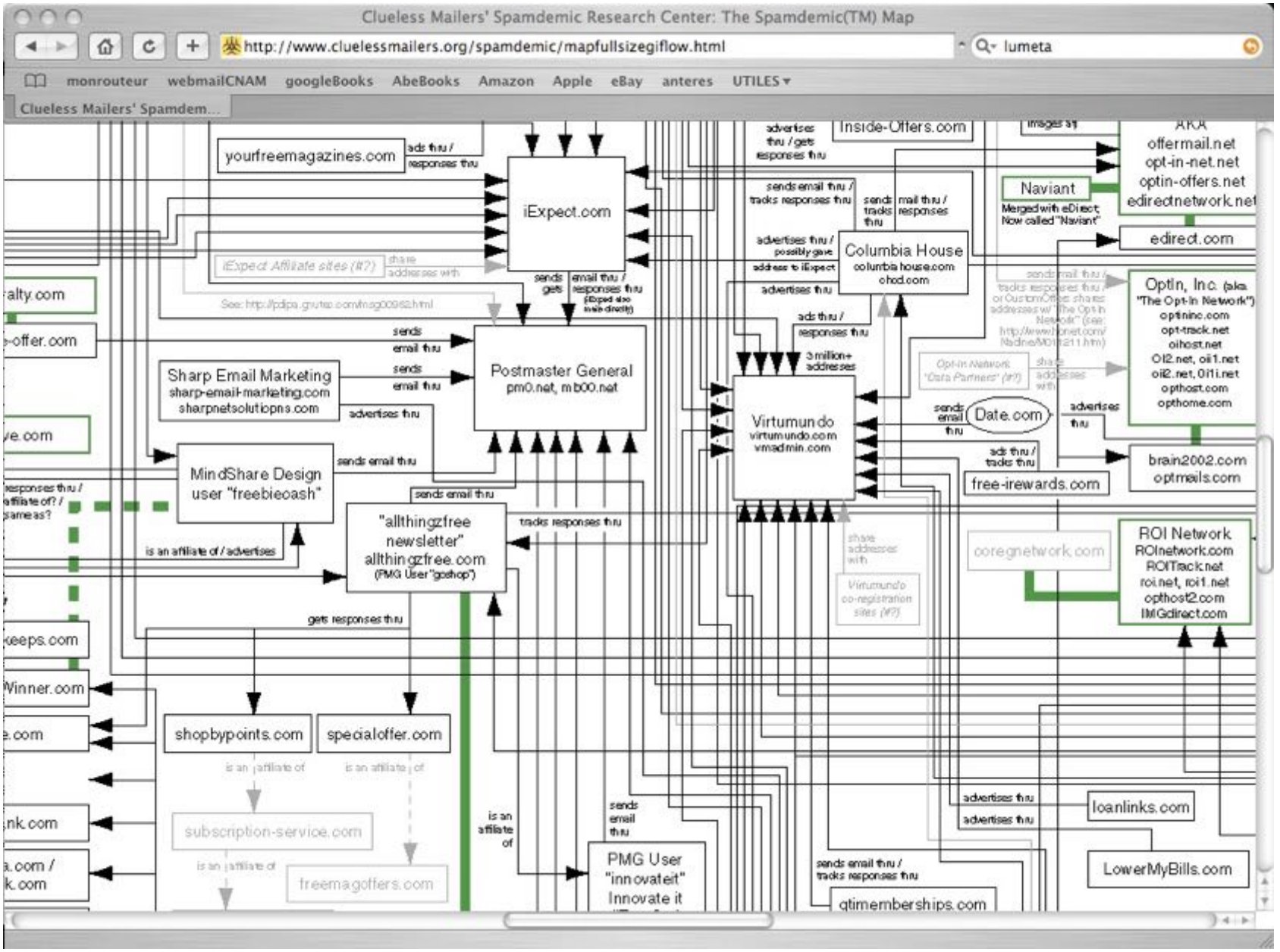
http://www.e-codices.unifr.ch/de/list/one/bbb/0347

exemple donné dans [Tufte] sans la couleur...

Aujourd'hui : cybergéographie

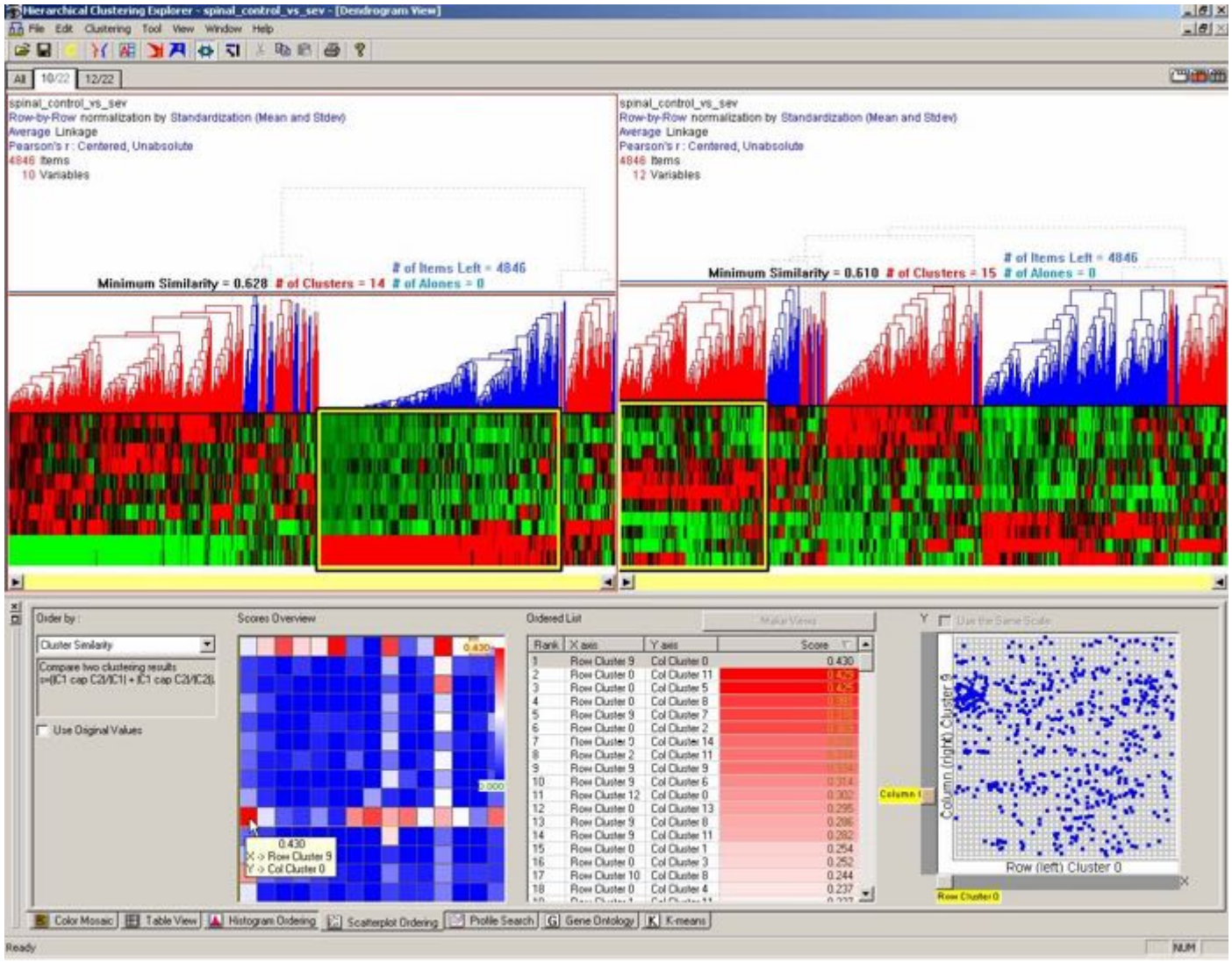


Lumeta (revue Nature)

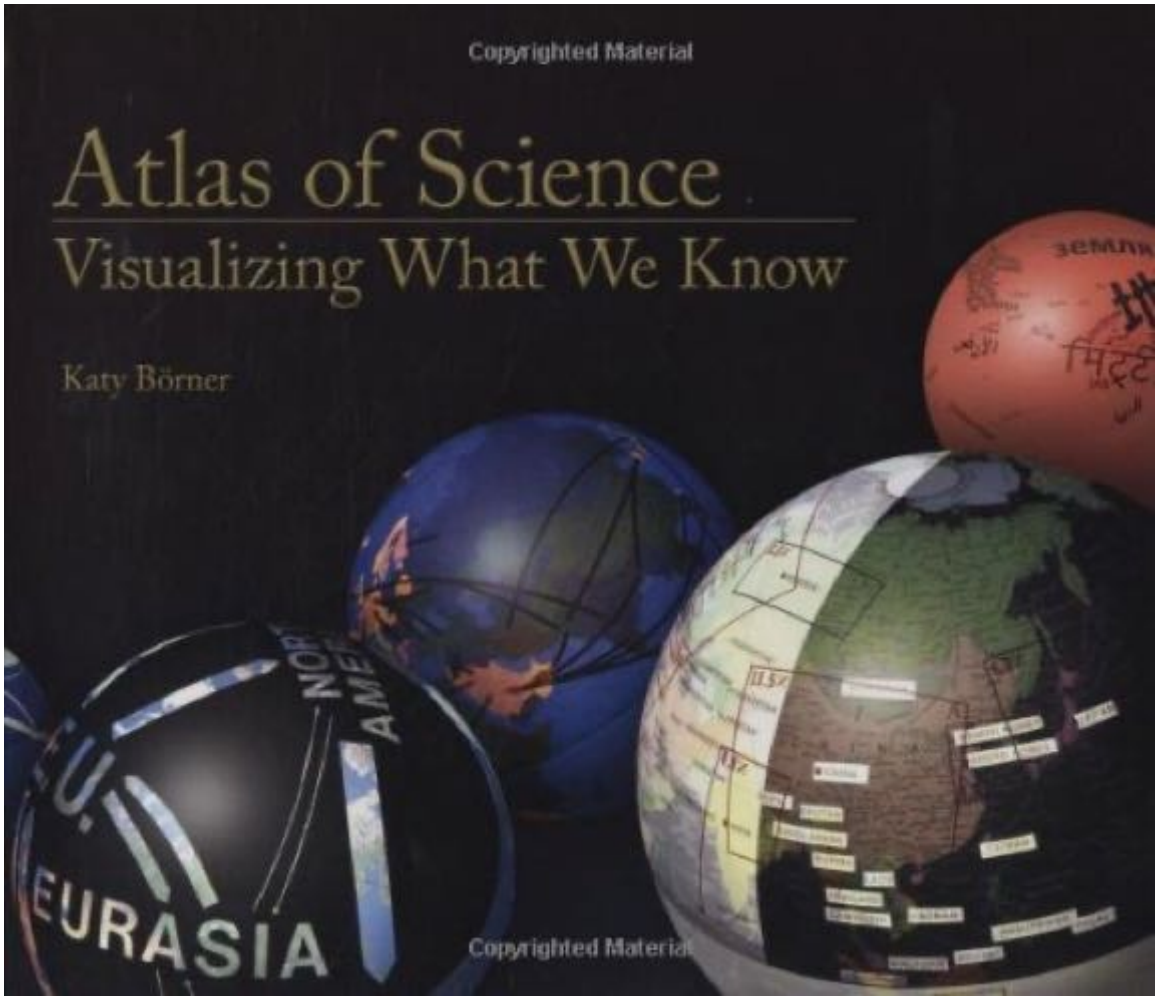


propagation d'un virus informatique

Génétique :


Hierarchical Clustering Explorer - spinal_control_vs_sev - [Dendrogram View]
File Edit Clustering Tool View Window Help
All 10/22 12/22
spinal_control_vs_sev
Row-by-Row normalization by Standardization (Mean and Stdev)
Average Linkage
Pearson's r : Centered, Unabsolute
4846 Items
10 Variables
# of Items Left = 4846
Minimum Similarity = 0.628 # of Clusters = 14 # of Alones = 0
spinal_control_vs_sev
Row-by-Row normalization by Standardization (Mean and Stdev)
Average Linkage
Pearson's r : Centered, Unabsolute
4846 Items
12 Variables
# of Items Left = 4846
Minimum Similarity = 0.610 # of Clusters = 15 # of Alones = 0
Order by :
Cluster Similarity
Compare two clustering results
Use Original Values
Scores Overview
Ordered List
Rank X axis Y axis Score
1 Row Cluster 9 Col Cluster 0 0.430
2 Row Cluster 0 Col Cluster 11
3 Row Cluster 0 Col Cluster 5
4 Row Cluster 0 Col Cluster 8
5 Row Cluster 9 Col Cluster 7
6 Row Cluster 0 Col Cluster 2
7 Row Cluster 3 Col Cluster 14
8 Row Cluster 2 Col Cluster 11
9 Row Cluster 9 Col Cluster 9
10 Row Cluster 9 Col Cluster 6 0.314
11 Row Cluster 12 Col Cluster 0 0.302
12 Row Cluster 0 Col Cluster 13 0.295
13 Row Cluster 9 Col Cluster 8 0.286
14 Row Cluster 9 Col Cluster 11 0.282
15 Row Cluster 0 Col Cluster 1 0.254
16 Row Cluster 0 Col Cluster 3 0.252
17 Row Cluster 10 Col Cluster 8 0.244
18 Row Cluster 0 Col Cluster 4 0.237
0.430
X -> Row Cluster 9
Y -> Col Cluster 0
Column (right) Cluster 9
Row (left) Cluster 0
Color Mosaic
Table View
Histogram Ordering
Scatterplot Ordering
Profile Search
Gene Ontology
K-means
Ready
NUM

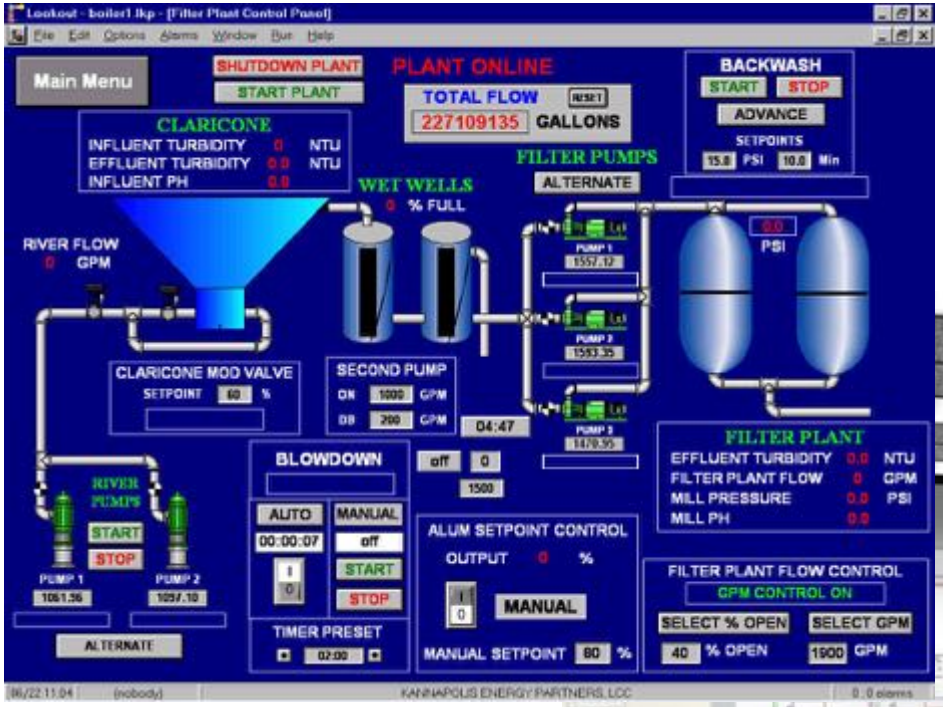
cartographie de la connaissance



livre magnifique (gd format) et pas trop cher (30€)

http://geography.sdsu.edu/People/Pages/skupin/

Autre exemple : piloter un processus complexe



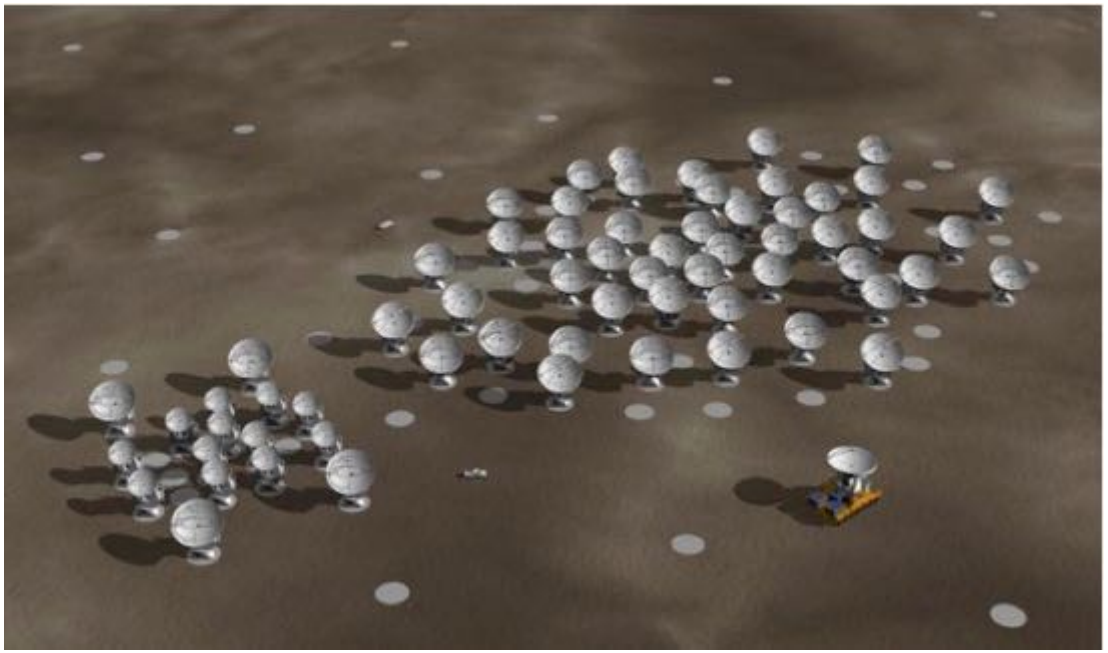
cockpit airbus A380 (wikipedia)

Traders

http://www.stocktradingtogo.com/2007/09/25/are-20-monitors-enough-for-one-trader-you-decide/

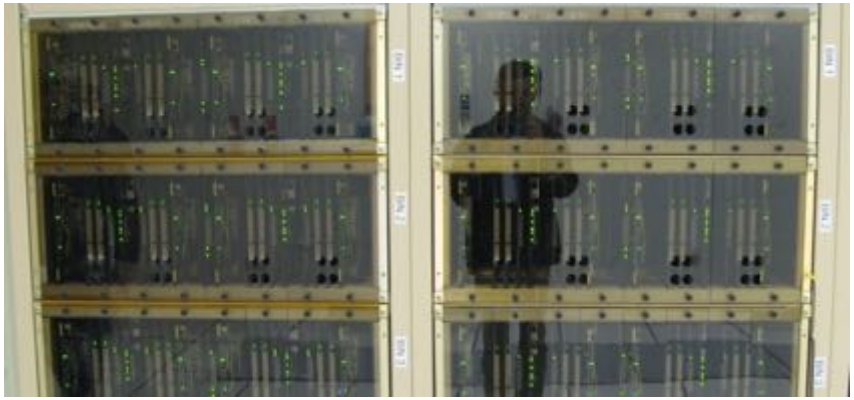
Grands instruments scientifiques : exemple d'ALMA

Chili (Atacama, 5000m)



interféromètre : 66 antennes (12m + 7m) et 2145 lignes de bases (<16 km)

ex de freq : 714 GHz !



corrélateur : 170 TFlops, 8200 FPGA, 180 KW

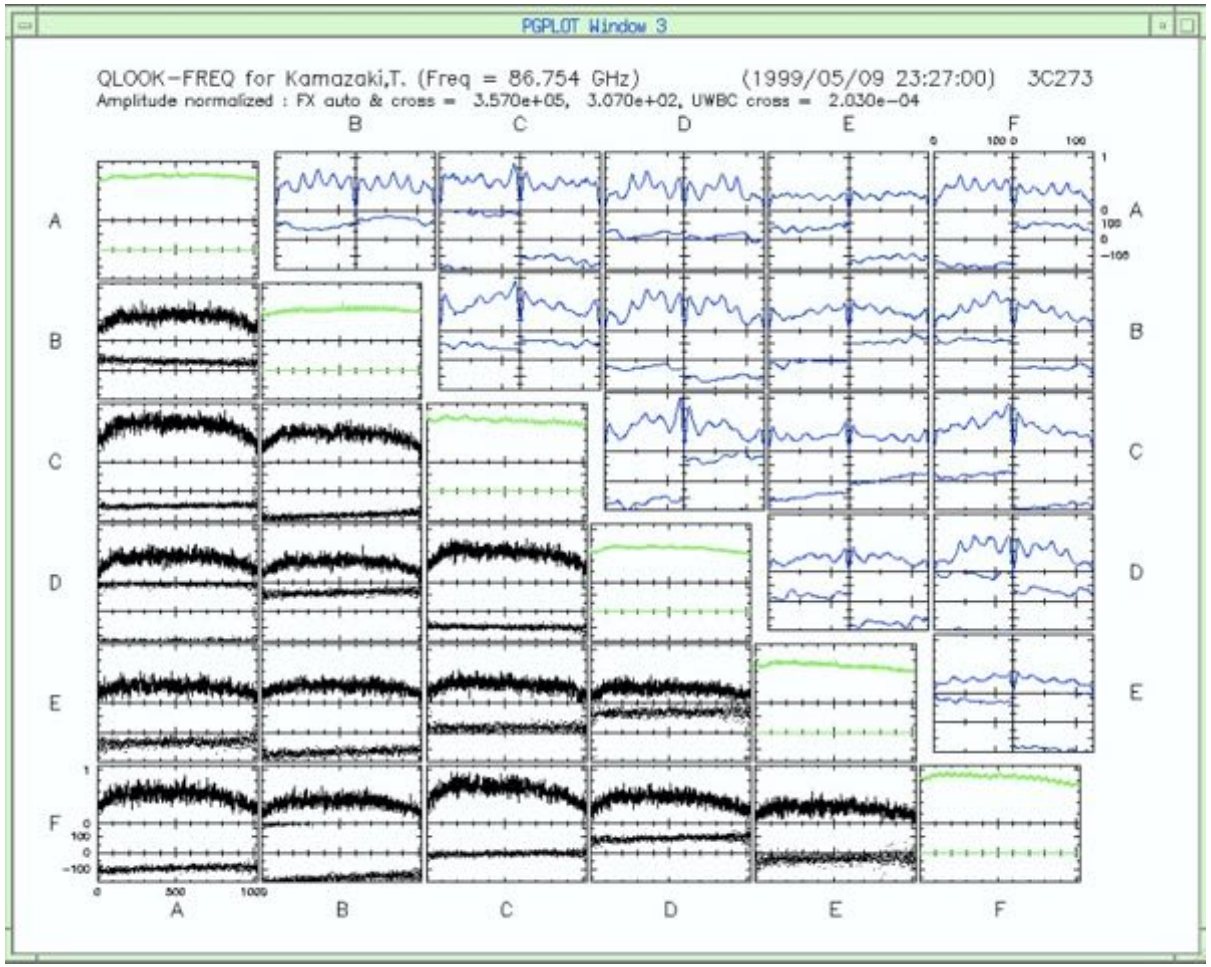
Salle de contrôle
en 2010-11

Java+Corba (avant 2000 !)

débit de capture : 1 Go/s
débit d'archivage : 180 To/an

http://www.eso.org/public/images/max_a129560s/

Problème du passage à l'échelle des interfaces



Exemple d'écran du VLA (Arizona) : 6 antennes => 6*5/2 lignes de base

Comment faire avec 2145 ?? A VOIR PLUS TARD

2. Définition du sujet

Compact graphical presentation and user interface for
- manipulating large numbers of items
- possibly extracted from far larger datasets

Enables users to make
- discoveries,
- decisions, or
- explanations

about
- patterns (trend, cluster, gap, outlier...),
- groups of items, or
- individual items.

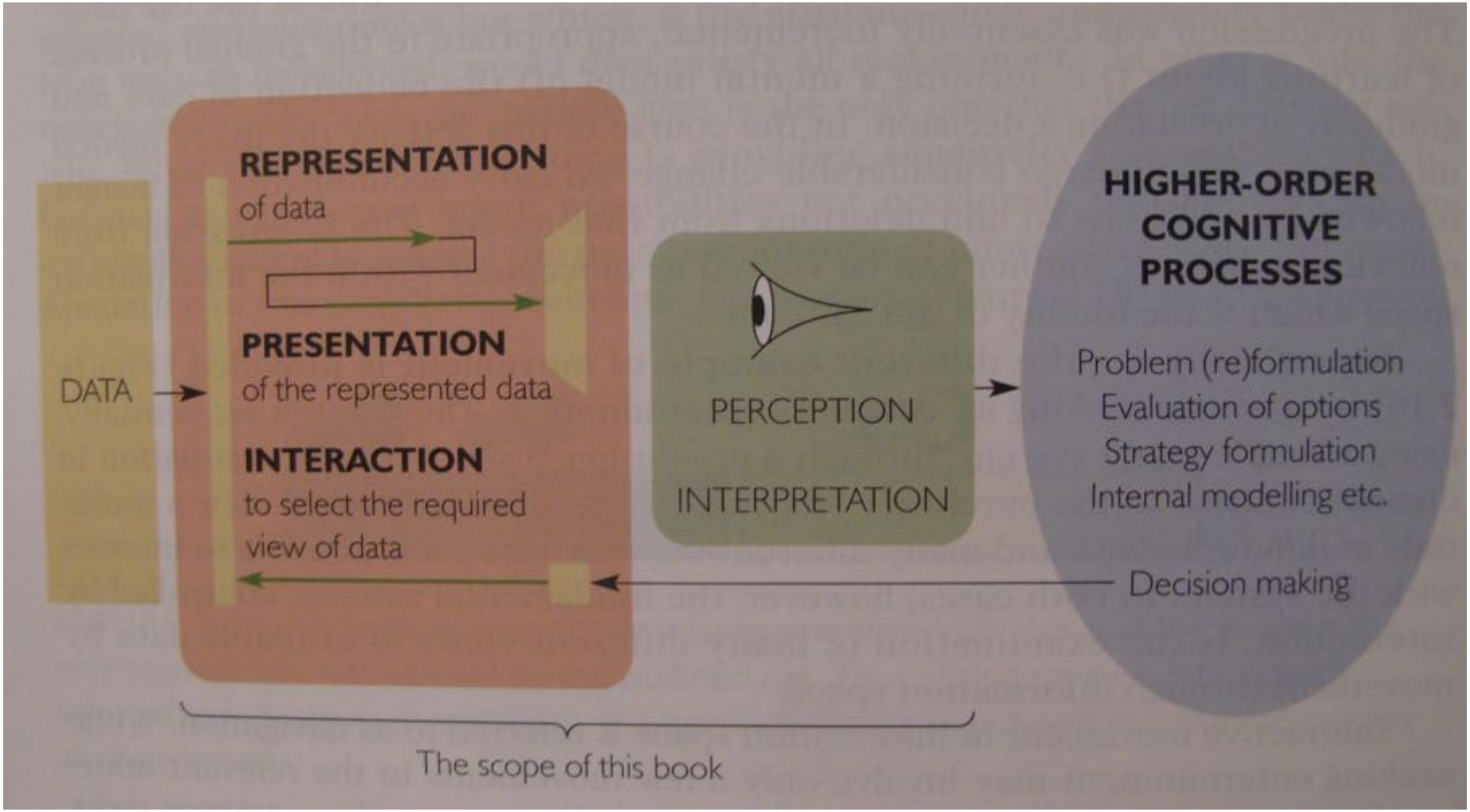
[Plaisant, 2001]

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.

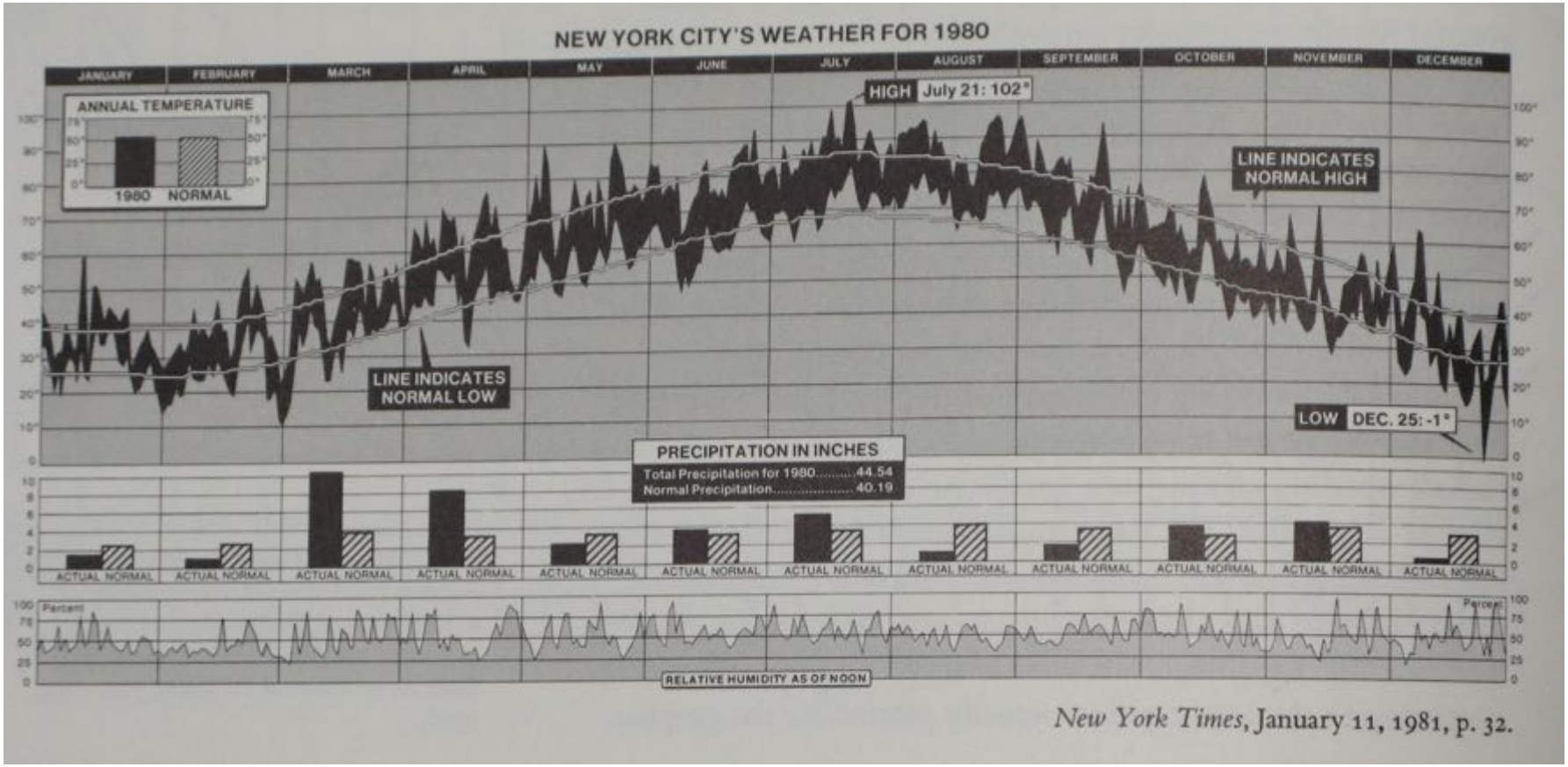
http://www.infovis-wiki.net/index.php/Information_Visualization

"using vision to think" (Card et al.)


REPRESENTATION
of data
PRESENTATION
of the represented data
INTERACTION
to select the required
view of data
DATA
PERCEPTION
INTERPRETATION
HIGHER-ORDER
COGNITIVE
PROCESSES
Problem (re)formulation
Evaluation of options
Strategy formulation
Internal modelling etc.
Decision making
The scope of this book

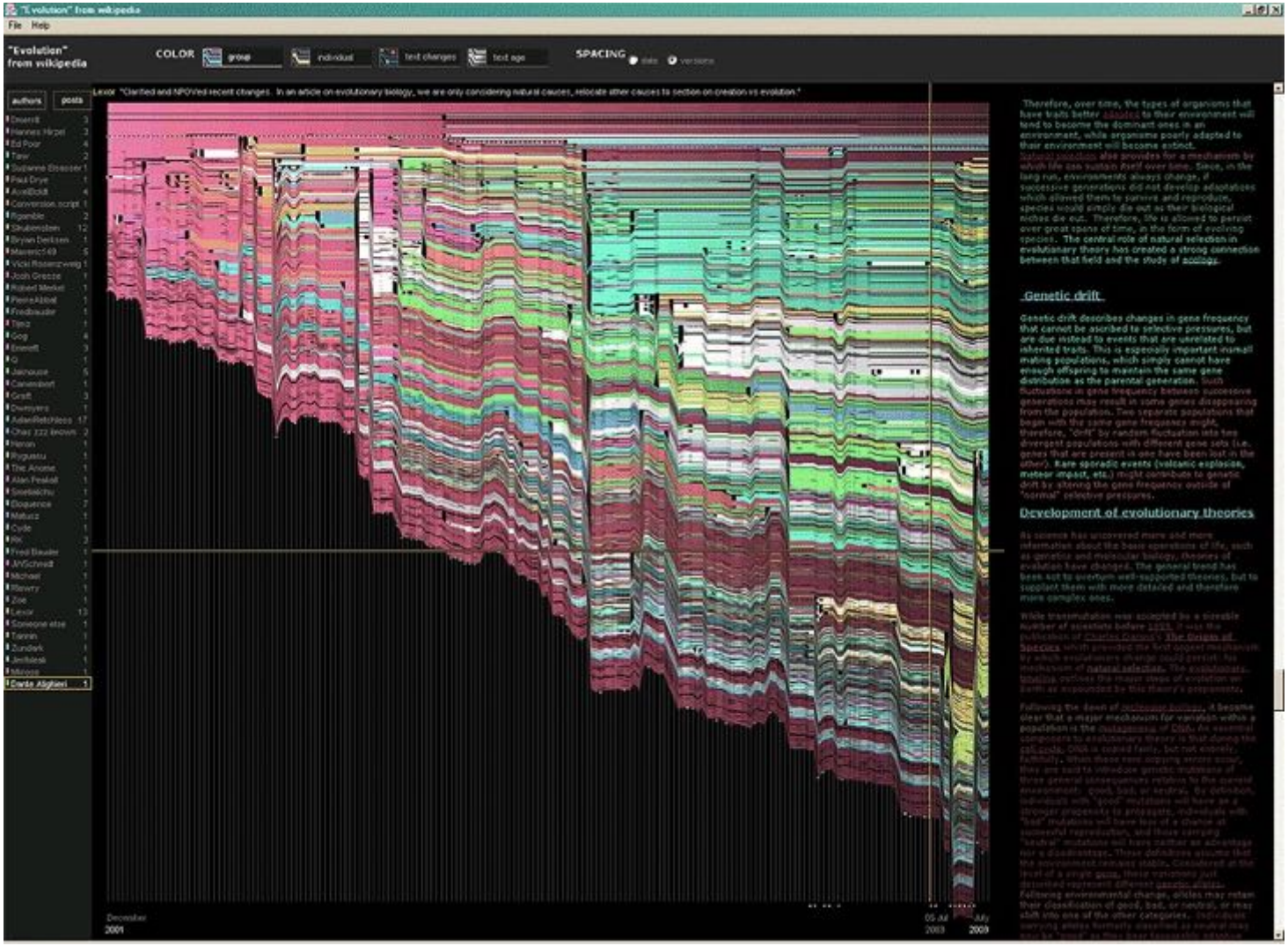
[Spence] p.26

Un schéma pour 1888 nombres



[Tufte, Visual display … p.30]

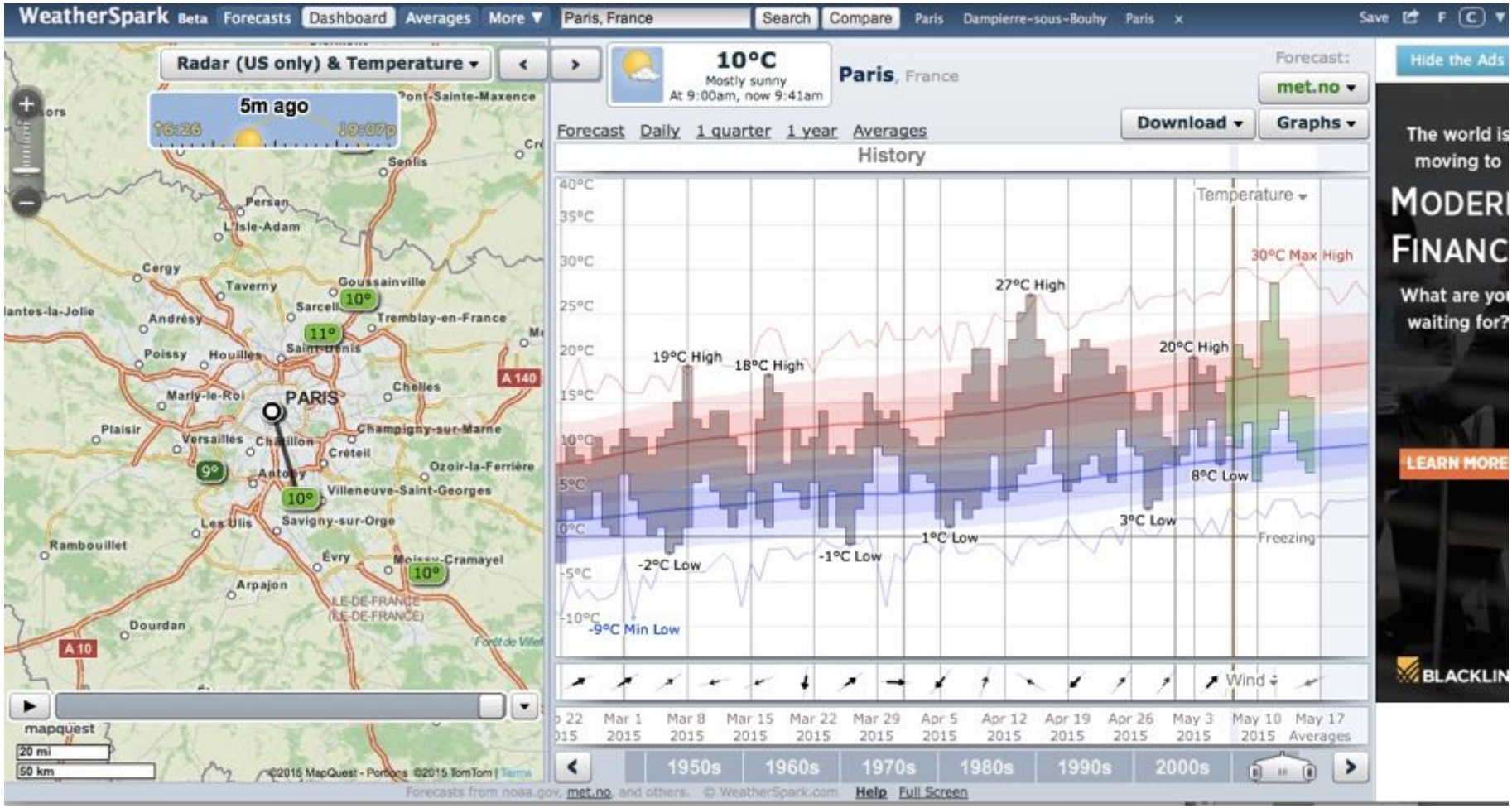
Si + de données, une interaction indispensable (zoom, etc.)



versions de l'article "evolution" sur Wikipedia

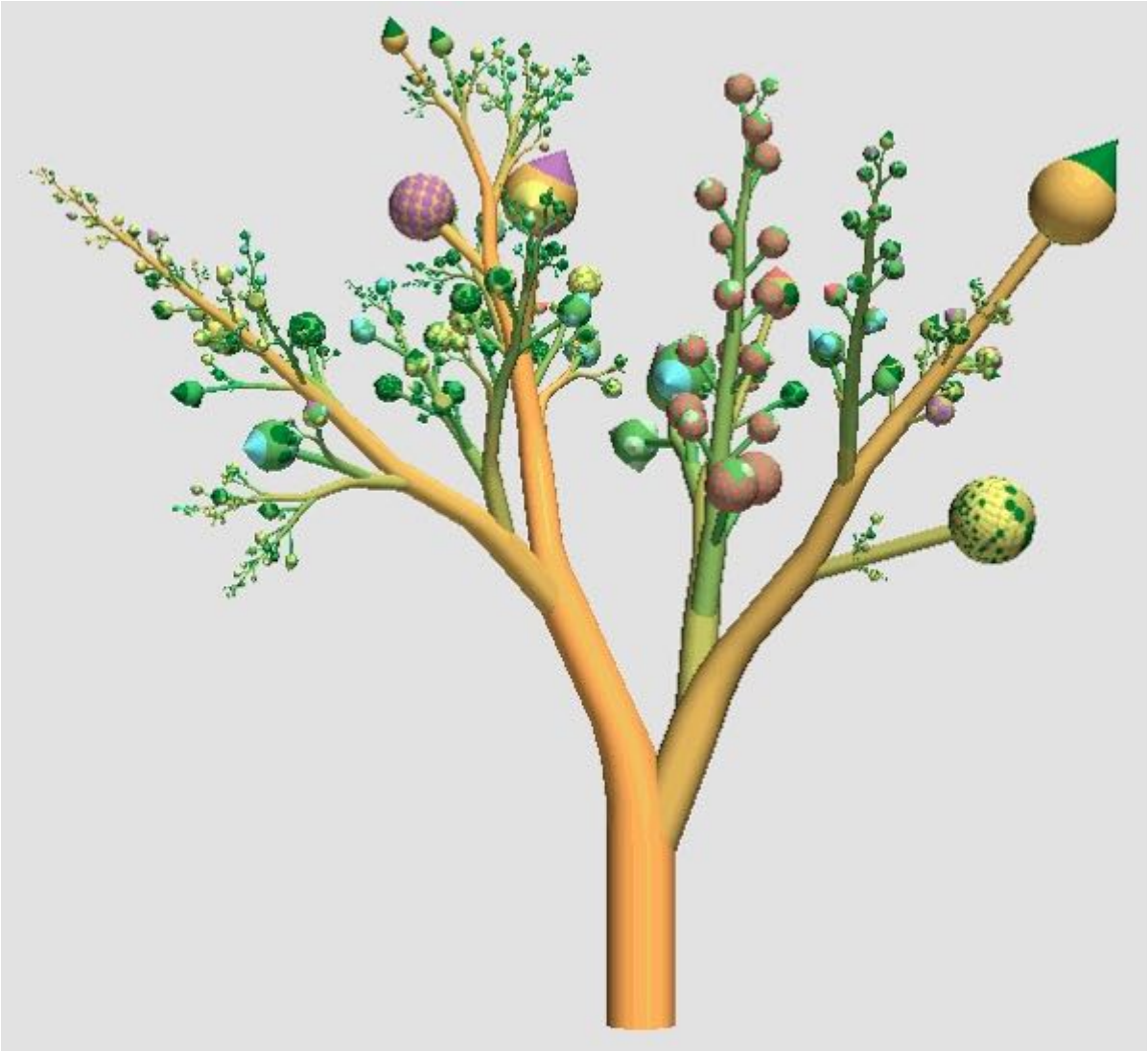
http://www.research.ibm.com/visual/projects/history_flow/

Autre exemple avec la météo



http://weatherspark.com/

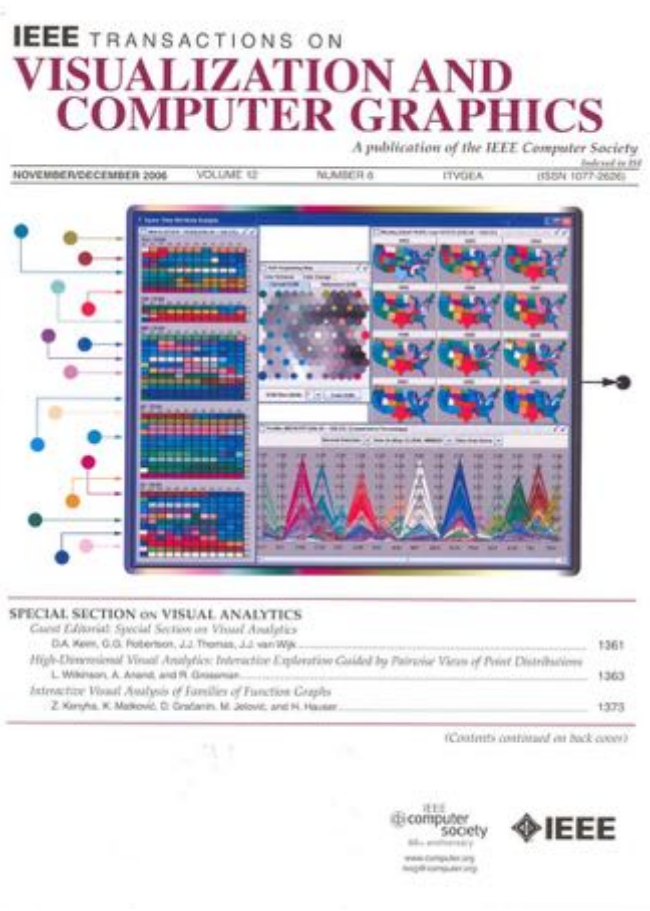
Représenter l'abstrait ? une usabilité à évaluer…



http://www.infovis-wiki.net/index.php?title=Image:Boom.gif

Infovis : une discipline scientifique

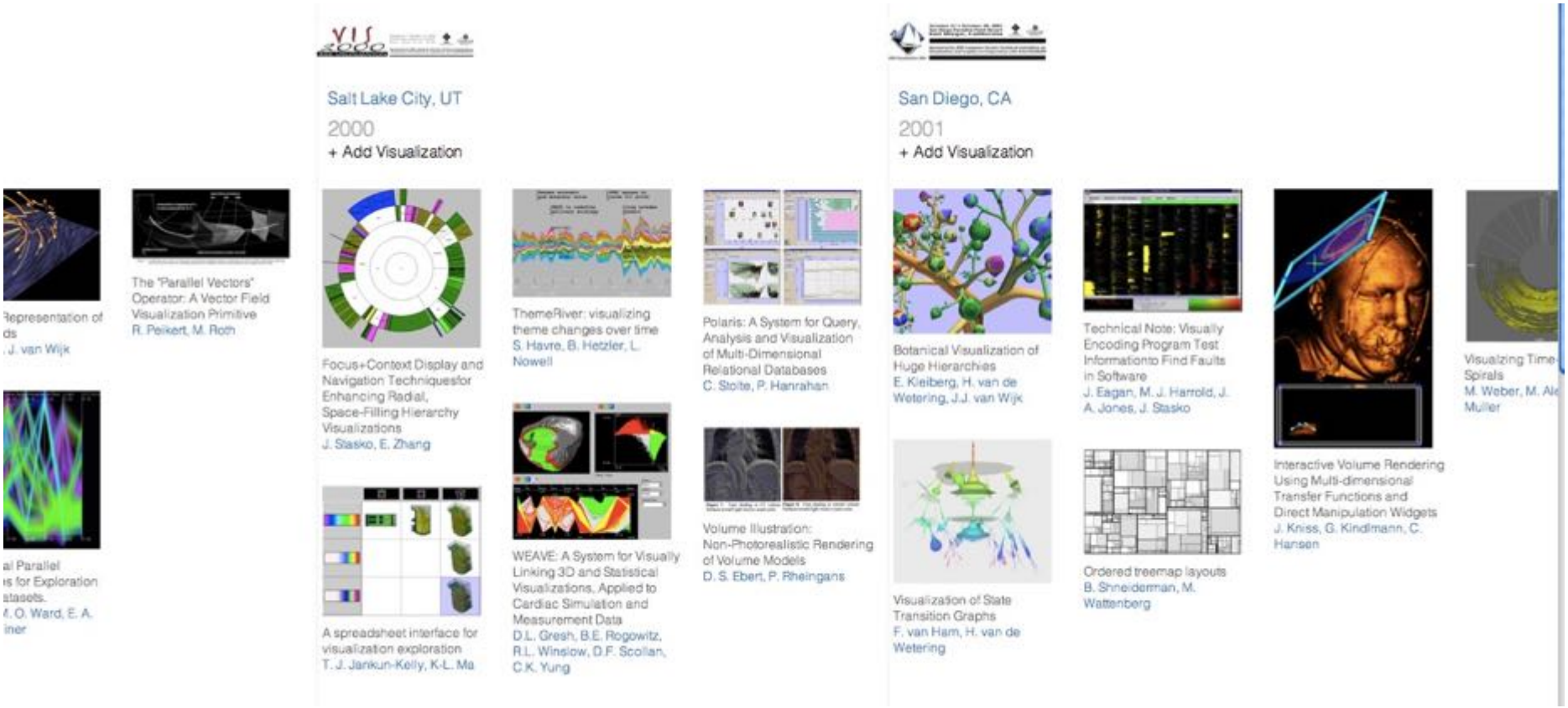
conférence VIS



revue TVCG

à la croisée de plusieurs autres (info, psycho, ergo, design, ...)

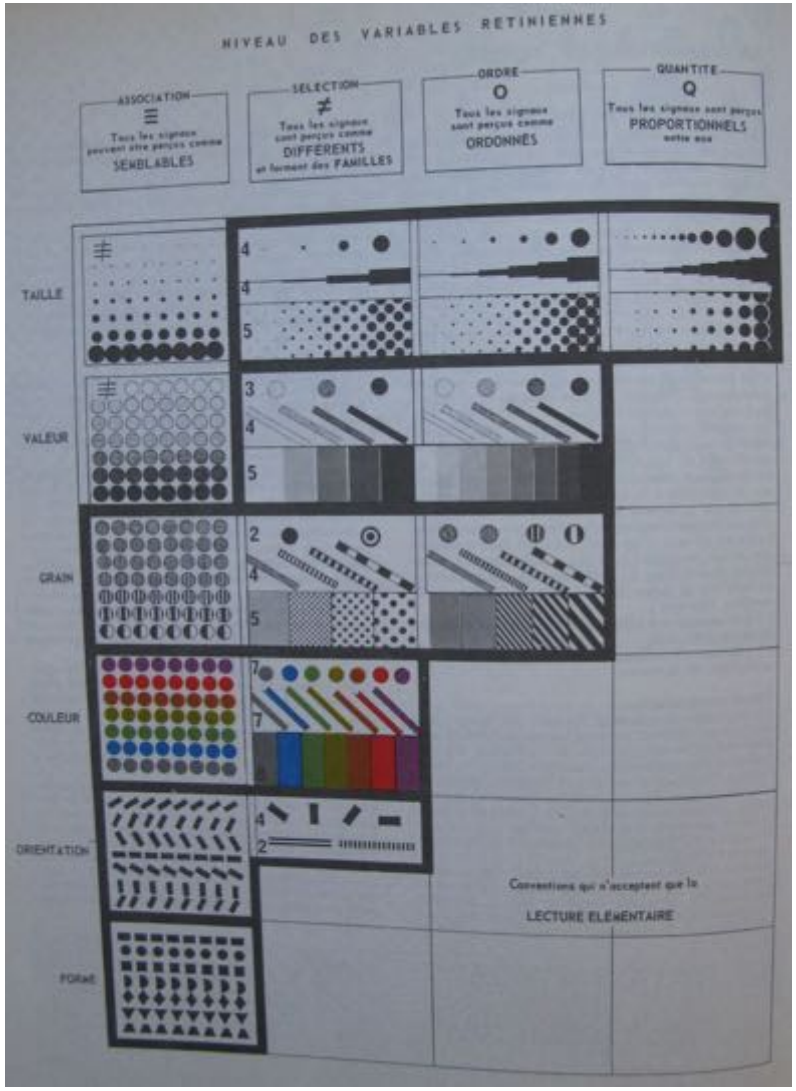
A voir : une chronologie des 25 ans de la conference sur la visualisation



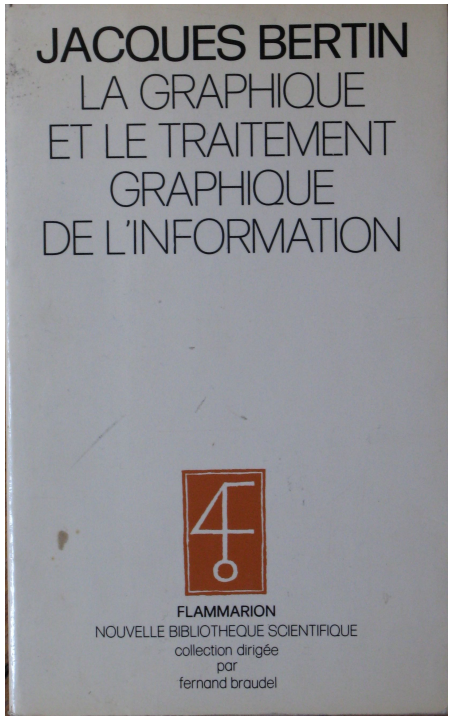
http://www.aviz.fr/~bbach/vis25timeline/

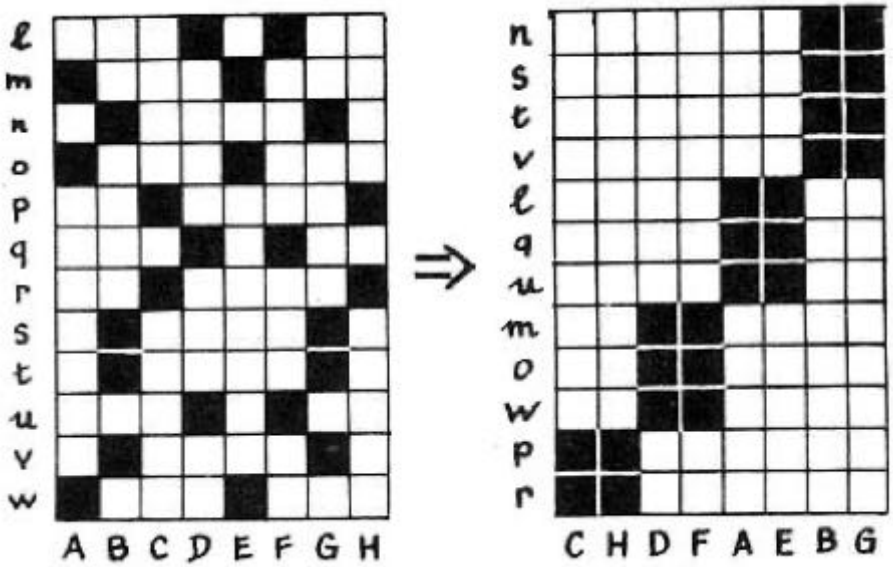
3. Travaux de Bertin et Tufte

J. Bertin
Sémiologie graphique
Mouton & Gauthier-Villars, 1967



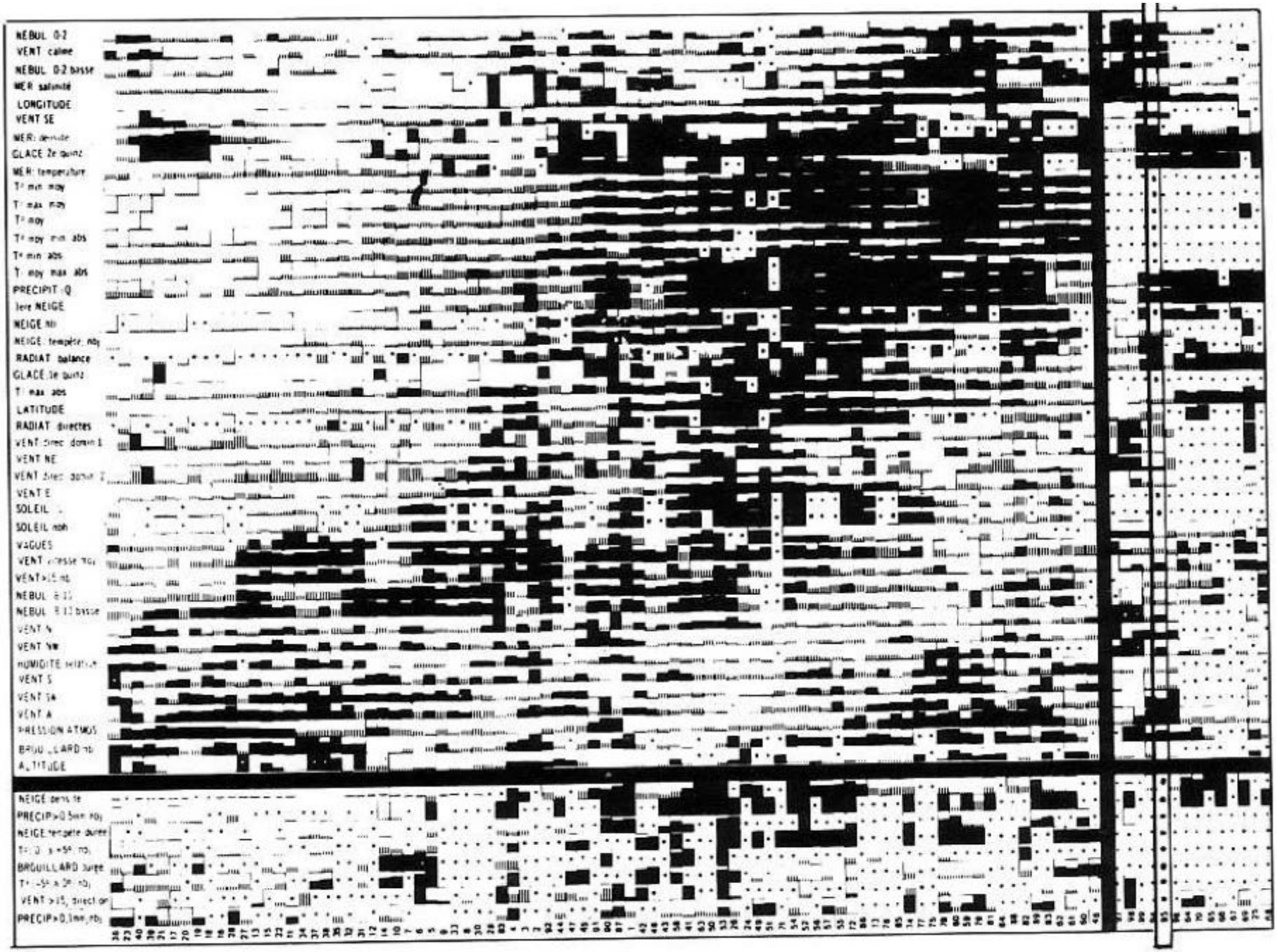
propriétés des "variables rétiennes" => cours #2





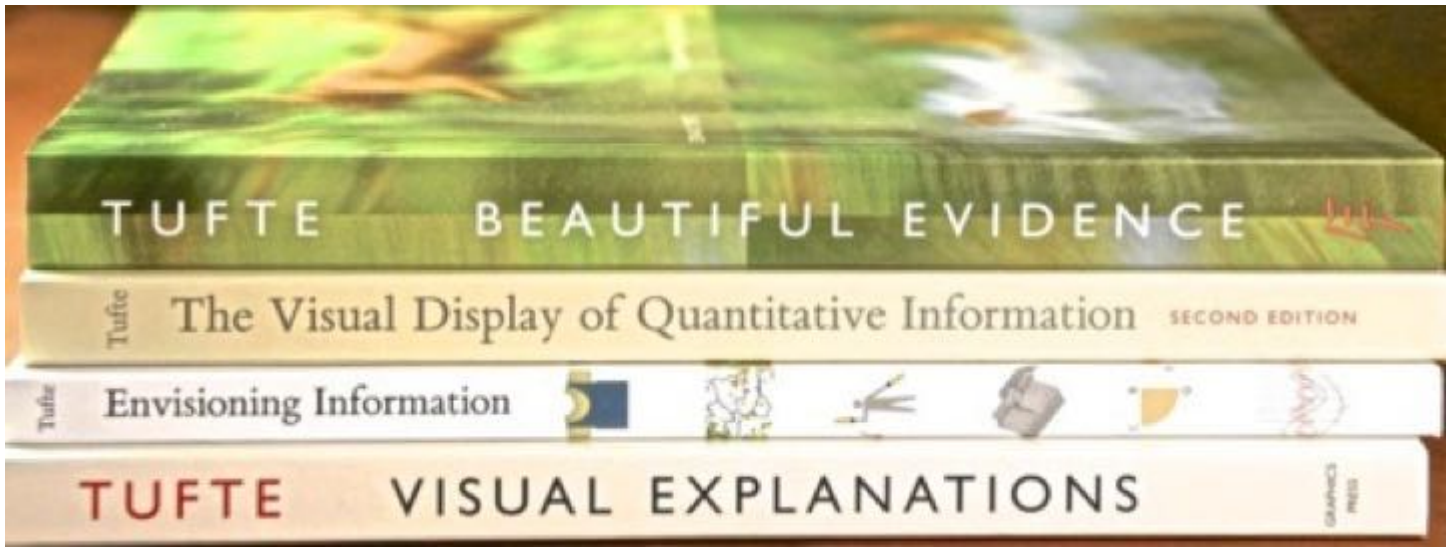
2ème livre : 1977

Matrices de permutation
et système "Domino"


NEBUL 0-2
VENT calme
NEBUL 0-2 basse
MER salinité
LONGITUDE
VENT SE
MER: densité
GLACE 2e quint
MER: température
T° min moy
T° max moy
T° moy
T° moy min abs
T° min abs
T° moy max abs
PRECIPIT >0
1ere NEIGE
NEIGE nb
NEIGE: tempête, nbj
RADIAT balance
GLACE 1e quint
T° max abs
LATITUDE
RADIAT directes
VENT direc dominante
VENT NE
VENT direc dominante 2
VENT E
SOLEIL
SOLEIL nbh
VAGUES
VENT vitesse moy
VENT>15 nb
NEBUL 8-10
NEBUL 8-10 basse
VENT N
VENT NW
HUMIDITE relative
VENT S
VENT SW
VENT W
PRESSION ATMOS
BROUILLARD nb
ALTITUDE
NEIGE densité
PRECIP>0.5mm nbj
NEIGE tempête durée
T° 0° à 5° nbj
BROUILLARD durée
T° -5° à 0° nbj
VENT >15, direction
PRECIP>0.1mm nbj

Edward Tufte

4 livres :



http://www.edwardtufte.com/tufte/

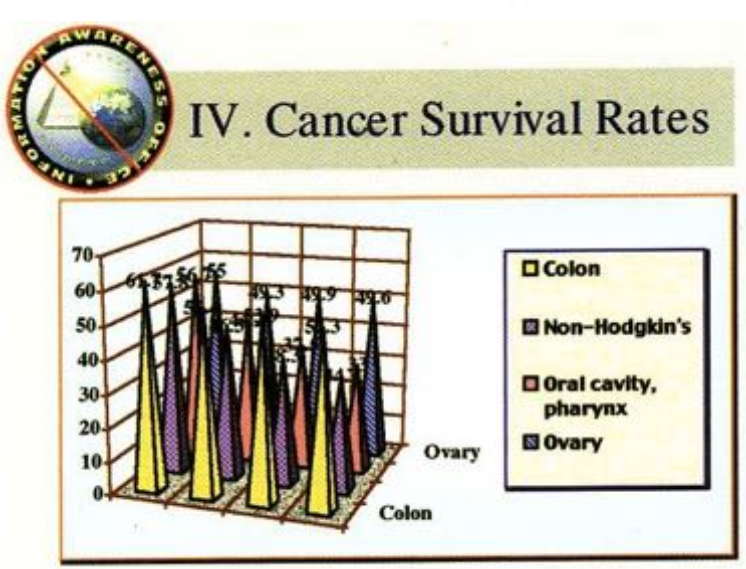
- concept du data/ink ratio
- "maximize data density (within reason)"

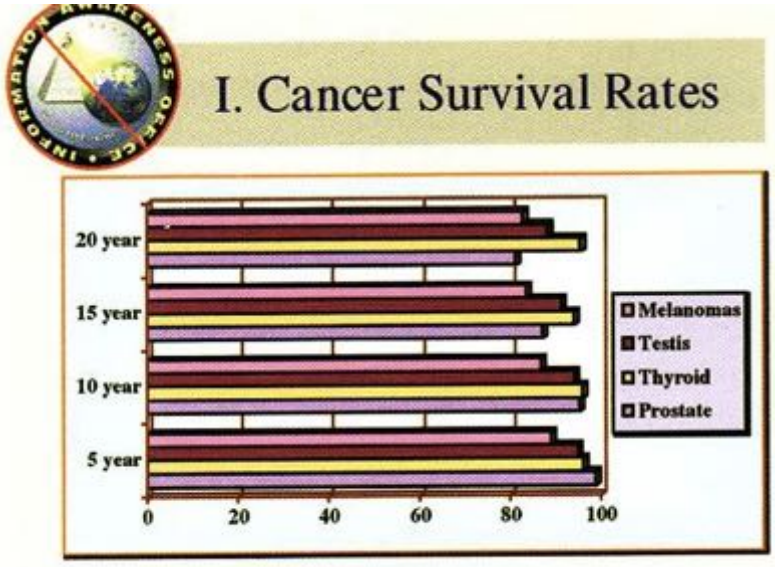
  densité= dim matrice données / surface du graphique

- "The graphics (…) tells a story"

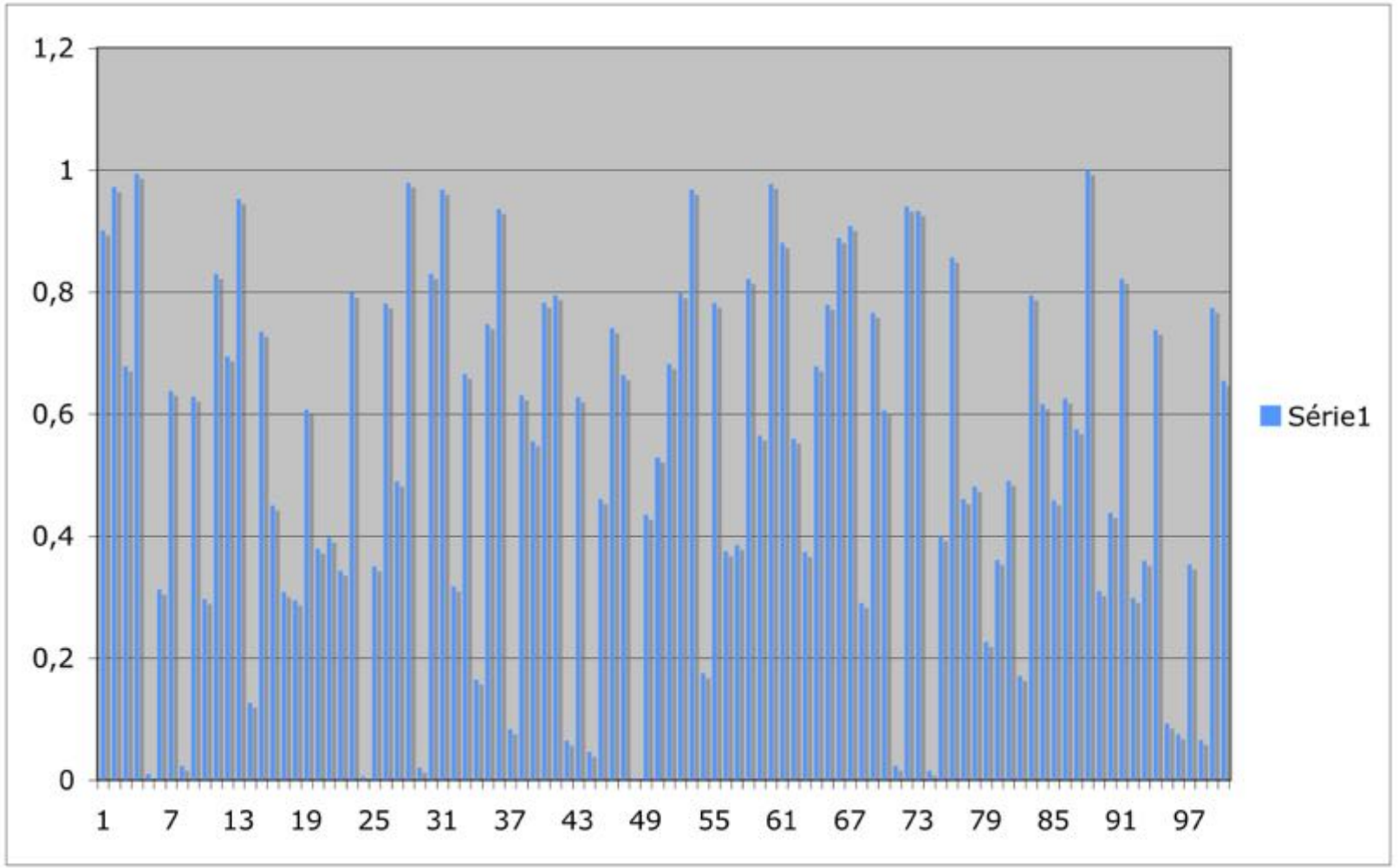
Critique du style PowerPoint (PP)

Applying the PowerPoint templates for statistical graphics to this nice straightforward table yields the analytical disasters on the facing page. These PP default-designs cause the data to explode into 6 separate chaotic slides, consuming 2.9 times the area of the table. *Everything* is wrong with these smarmy, incoherent graphs: uncomparative, thin data-density, chartjunk, encoded legends, meaningless color, logotype branding, indifference to content and evidence. Chartjunk is a clear sign of statistical stupidity; use these designs in your presentation, and your audience will quickly and correctly conclude that you don't know much about data and evidence.[20] Poking a finger into the eye of thought, these data graphics would turn into a nasty travesty if used for



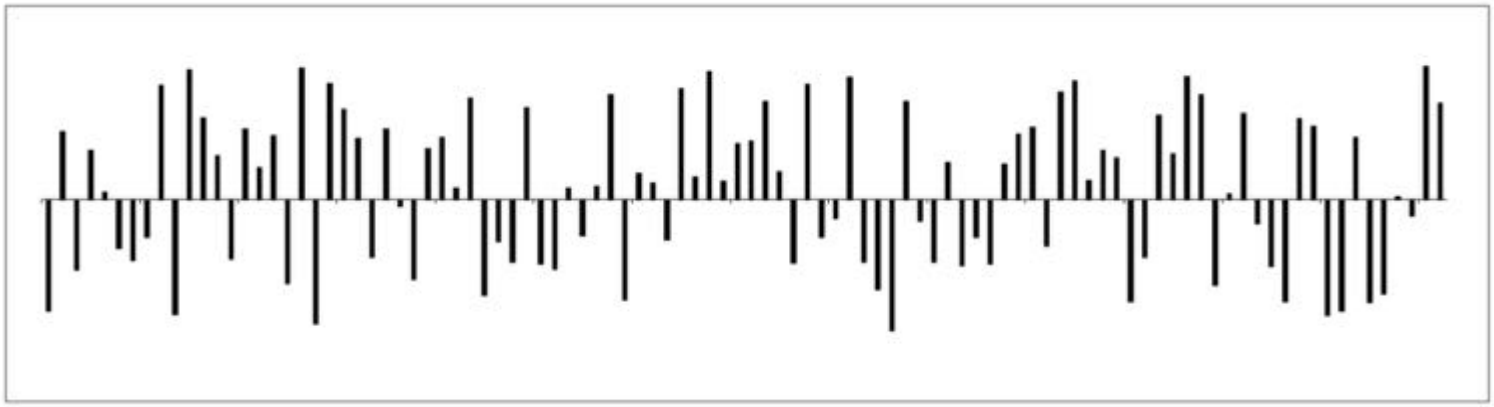


Exemple du grapheur d'Excel :



100 valeurs aléatoire entre 0 et 1, graphe par défaut

Après application des principes de Tufte (et Bertin)



Passage à l'échelle ?

4. Le Xerox PARC : S. Card, J. Mackinlay, R. Robertson

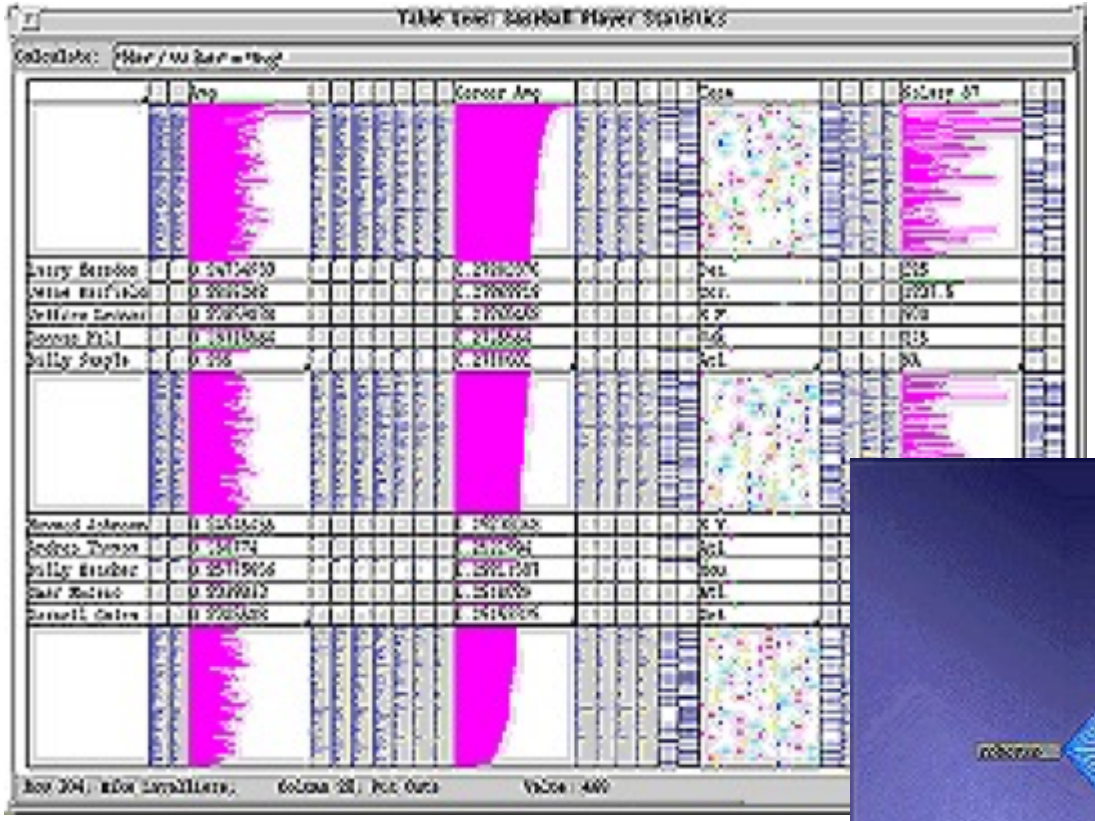
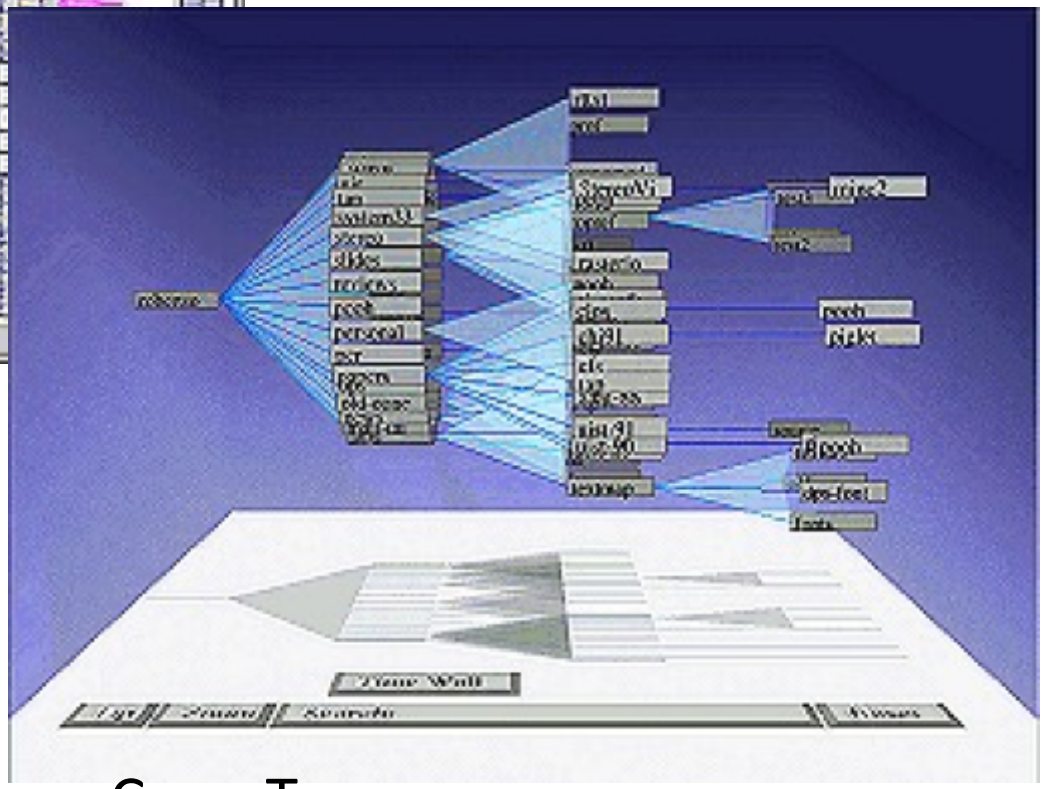
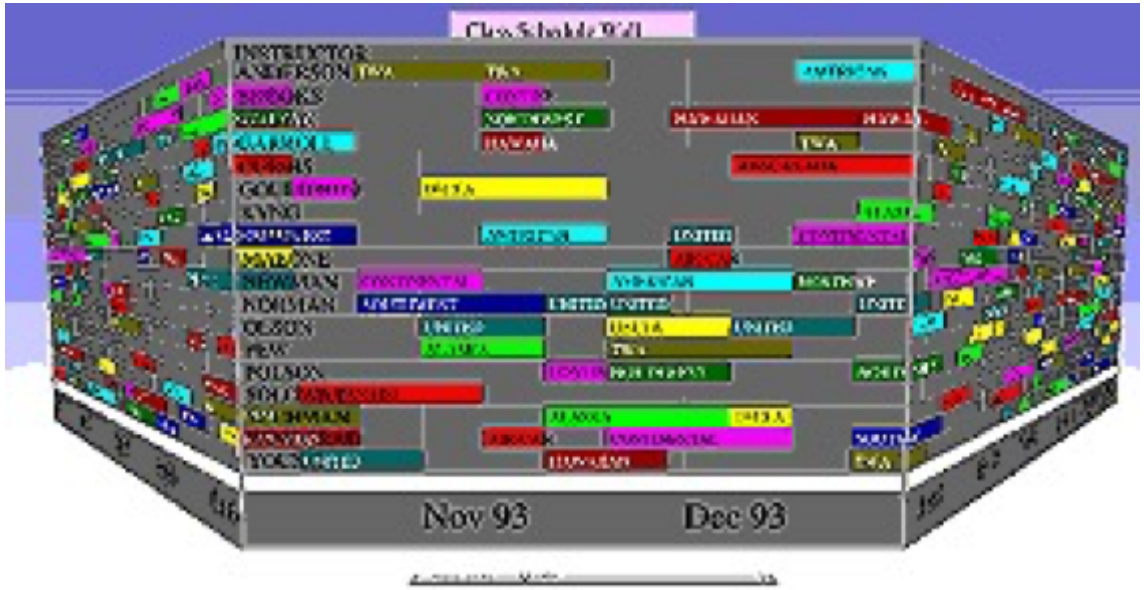


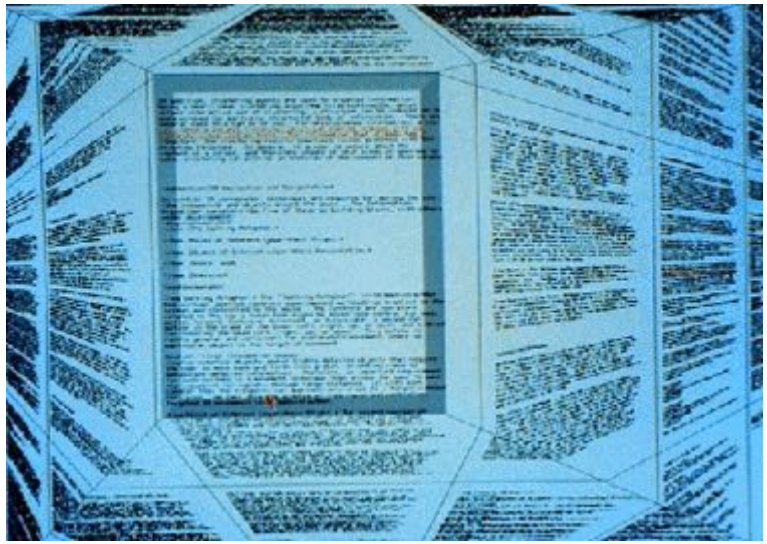
Table lens



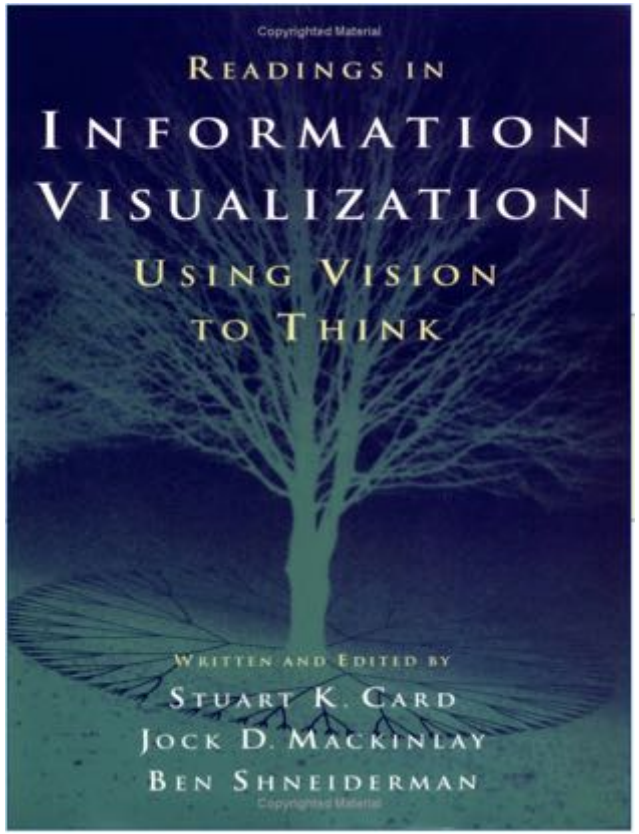
Cone Tree



Perspective Wall



Data lens



(niveau recherche)

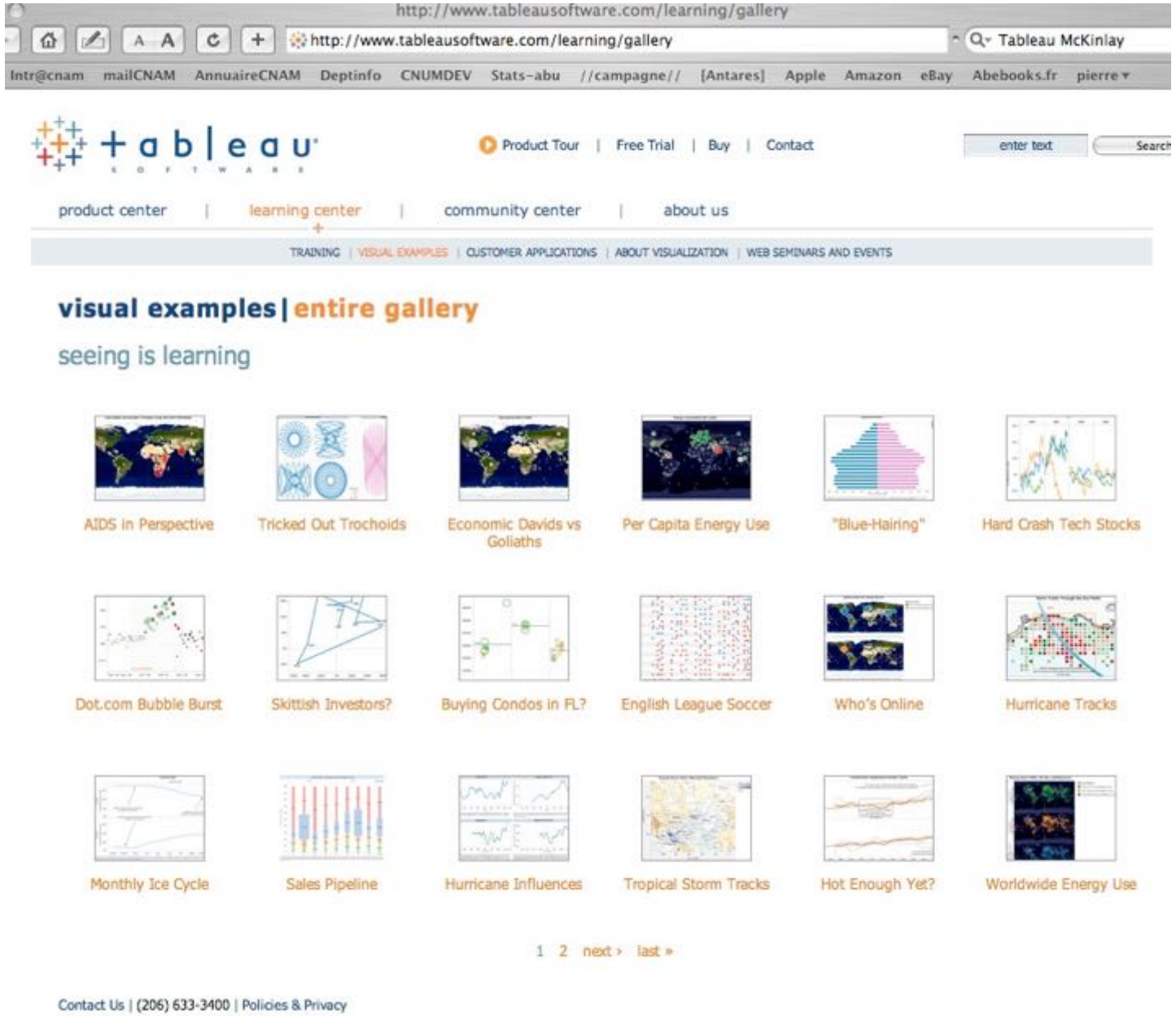
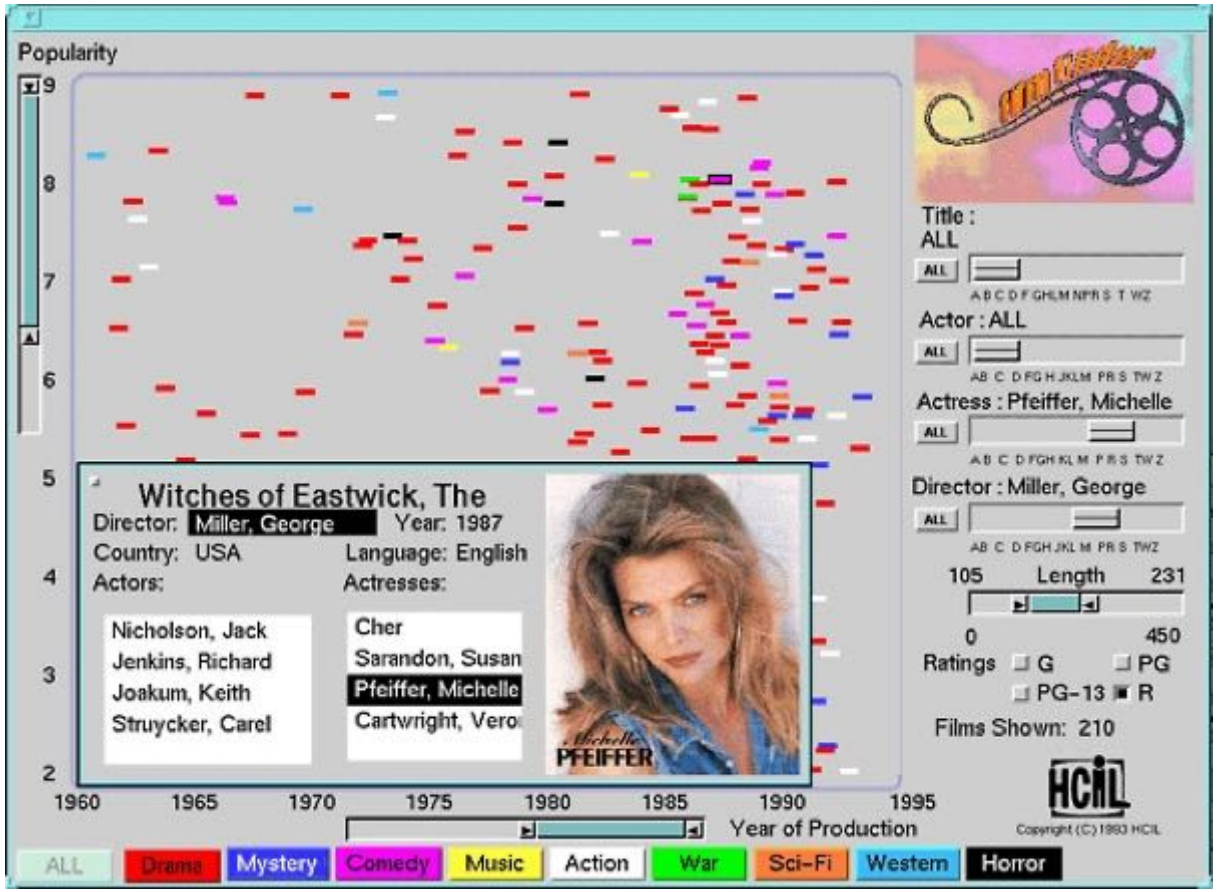


Tableau (Jock Mackinlay)

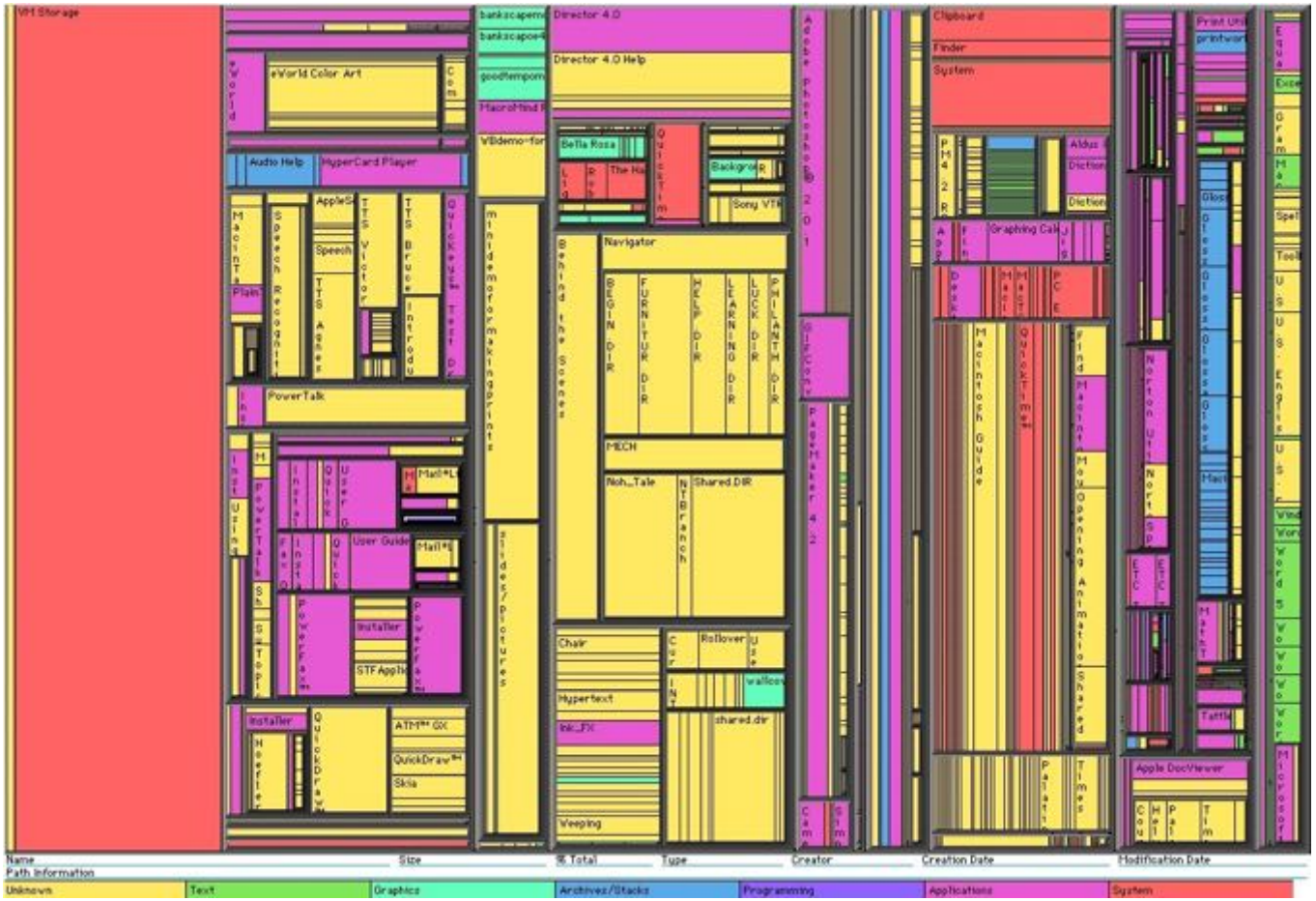
Le HCI lab, univ. Maryland

Dynamic queries (1993)



the FilmFinder comprised 1,838 film titles, 5,468 actors and 1,463 directors.
Shneiderman and Ahlberg (CHI, 1994)

Treemaps (1991)



Infovis mantra (1996)

Overall, the bandwidth of information presentation is potentially higher in the visual domain than for media reaching any of the other senses. Humans have remarkable perceptual abilities that are greatly under-utilized in current designs. Users can scan, recognize, and recall images rapidly, and can detect changes in size, color, shape, movement, or texture. They can point to a single pixel, even in a megapixel display, and can drag one object to another to perform an action. User interfaces have been largely text-oriented, so as visual approaches are explored, appealing new opportunities are emerging.

There are many visual design guidelines but the basic principle might be summarized as the Visual Information Seeking Mantra:

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

Each line represents one project in which I found myself rediscovering this principle and therefore wrote it down it as a reminder. It proved to be only a starting point

actions that users wish to perform.

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:

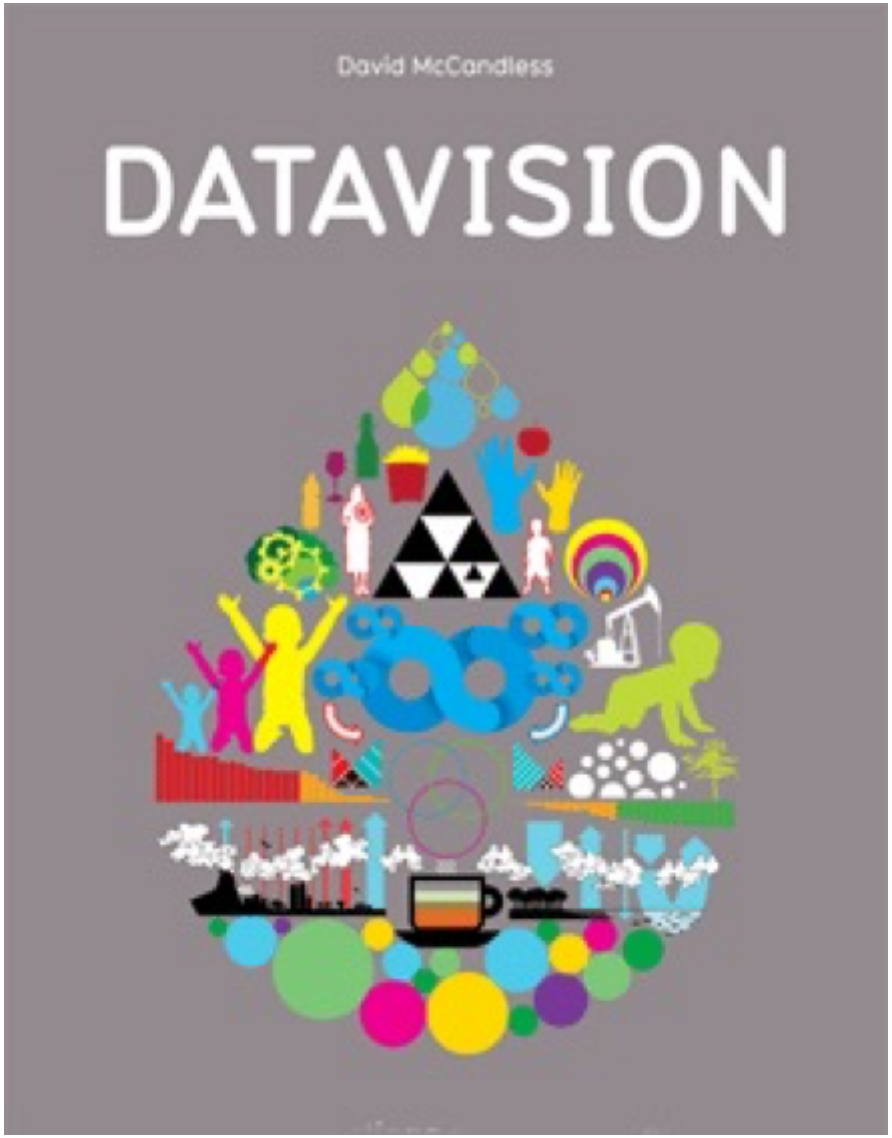
Overview: Gain an overview of the entire collection.
Zoom : Zoom in on items of interest
Filter: filter out uninteresting items.
Details-on-demand: Select an item or group and get details when needed.
Relate: View relationships among items.
History: Keep a history of actions to support undo, replay, and progressive refinement.
Extract: Allow extraction of sub-collections and of the query parameters.

Further discussion of the tasks follows the descriptions of the seven data types:

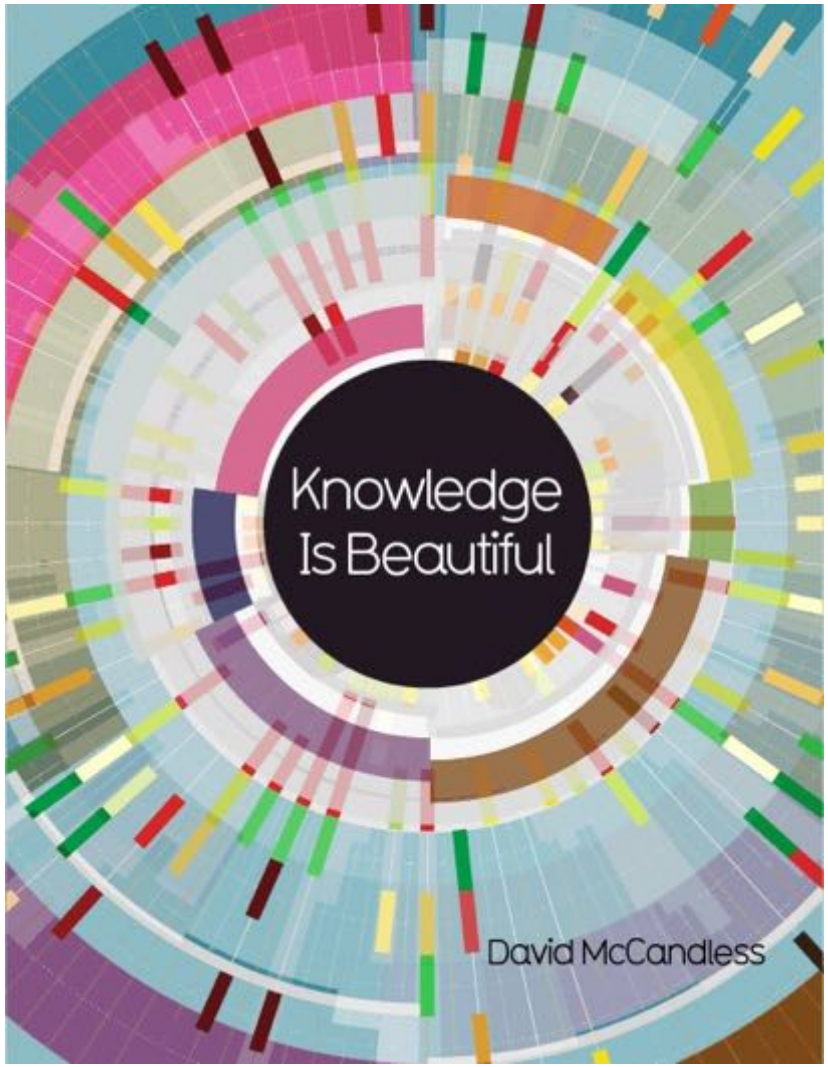
1-dimensional: linear data types include textual documents, program source code, and alphabetical lists of names which are all organized in a sequential manner. Each item in the collection is a line of text containing a string of characters. Additional line attributes might be the date of last update or author name. Interface design issues include what fonts, color, size to use and what overview, scrolling, or selection methods can be used. User problems might be to find the number of items, see items having

Shneiderman "The eyes have it : a task by data type taxonomy for information visualizations"

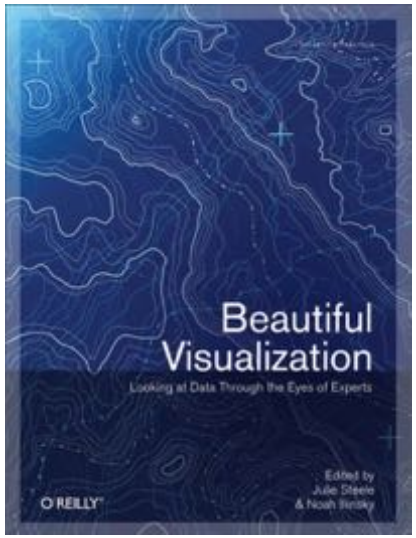
5. Aujourd'hui : "Information is beautifull"



D. McCandess

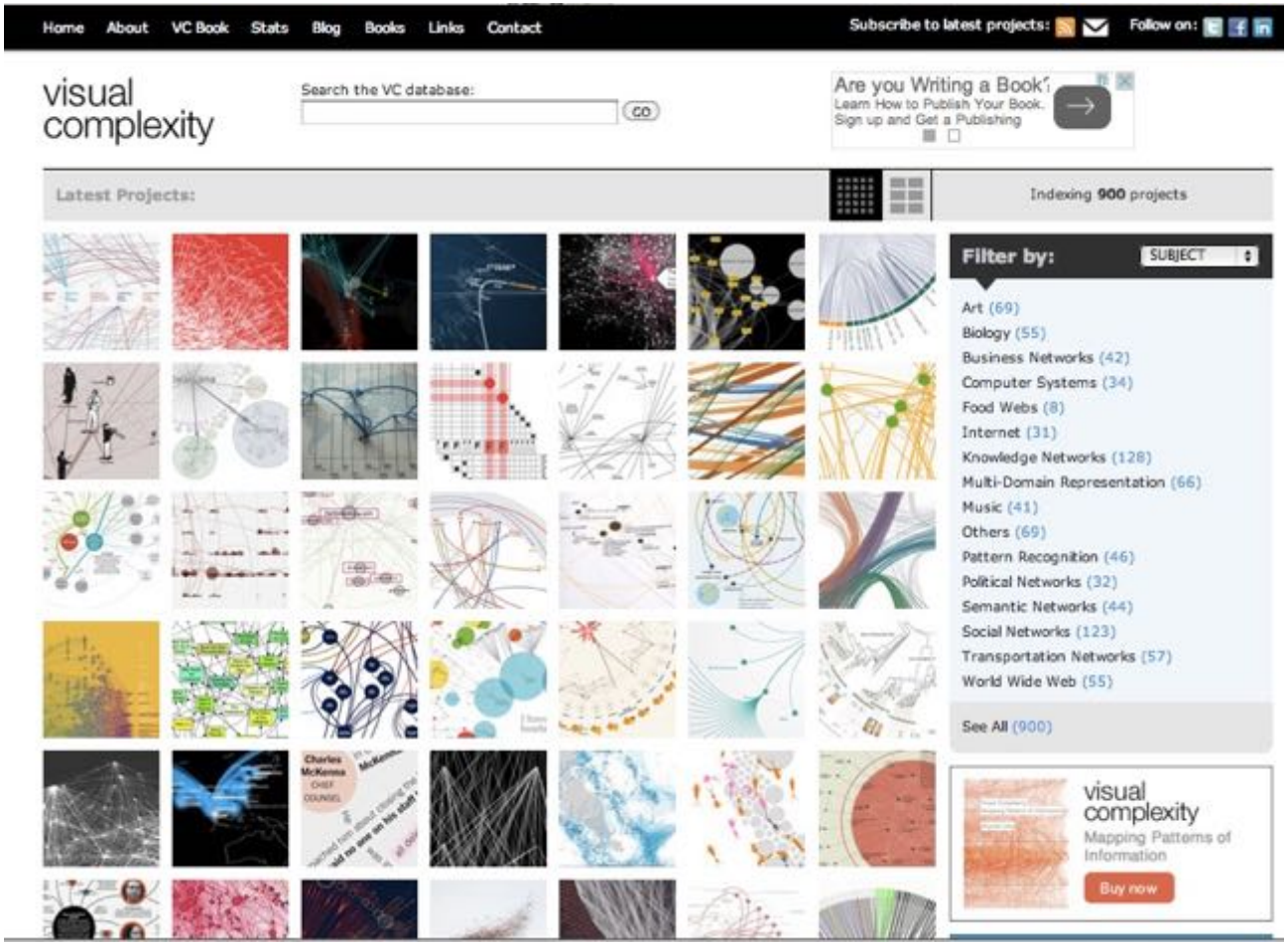


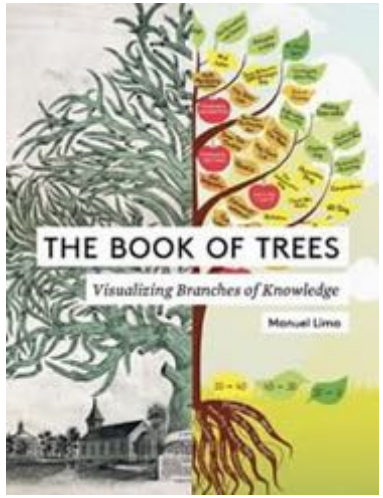
Wordle - Jonathan Feinberg



Aaron Koblin "Flight patterns" 2009

Manuel Lina

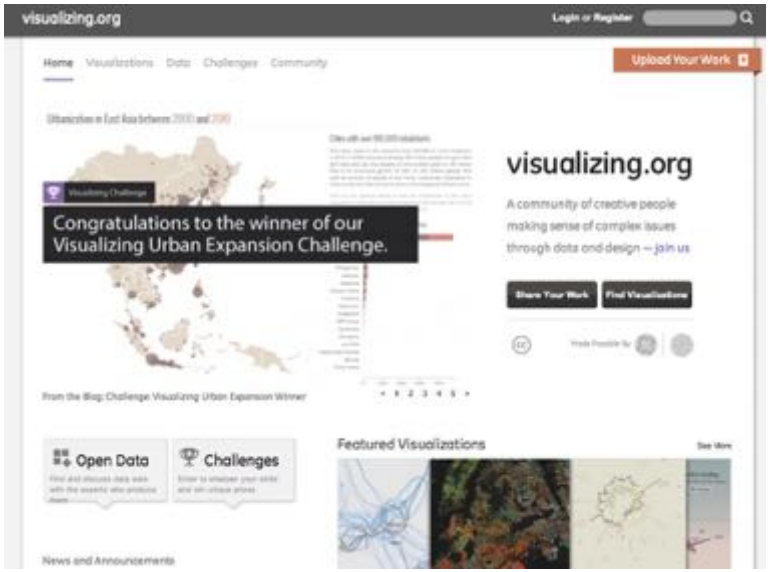




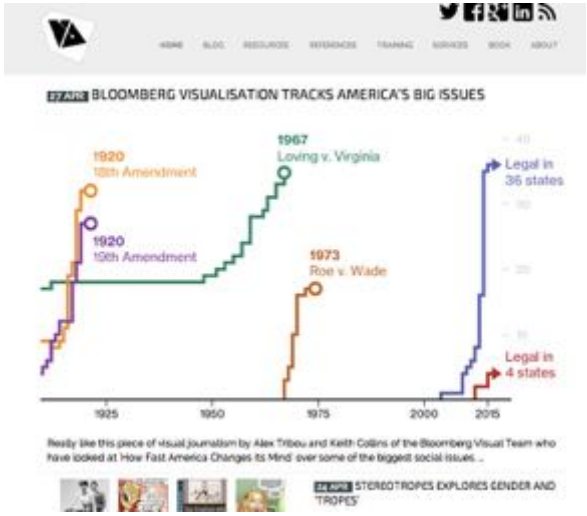
http://www.visualcomplexity.com/

6. Explosion de sites et de logiciels !

http://www.infovis-wiki.net/

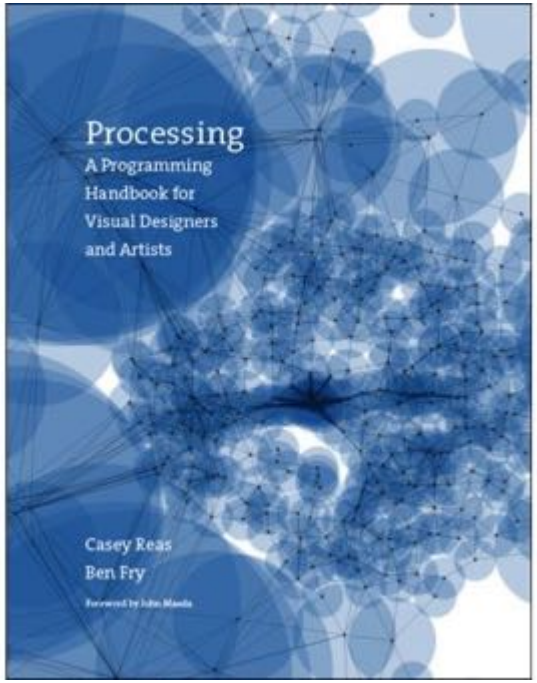


http://visualizing.org/

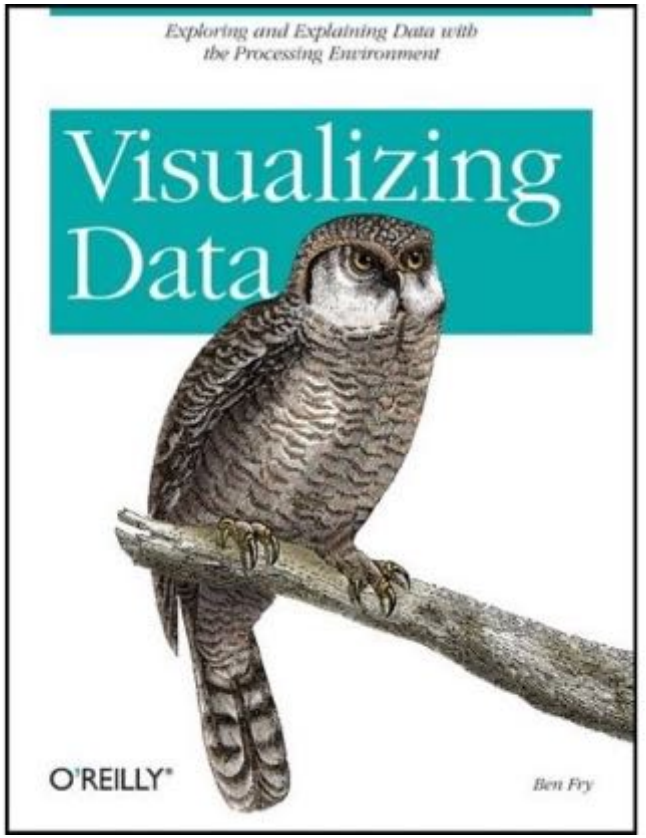


http://www.visualisingdata.com/

http://processing.org



C. REAS & B. FRY

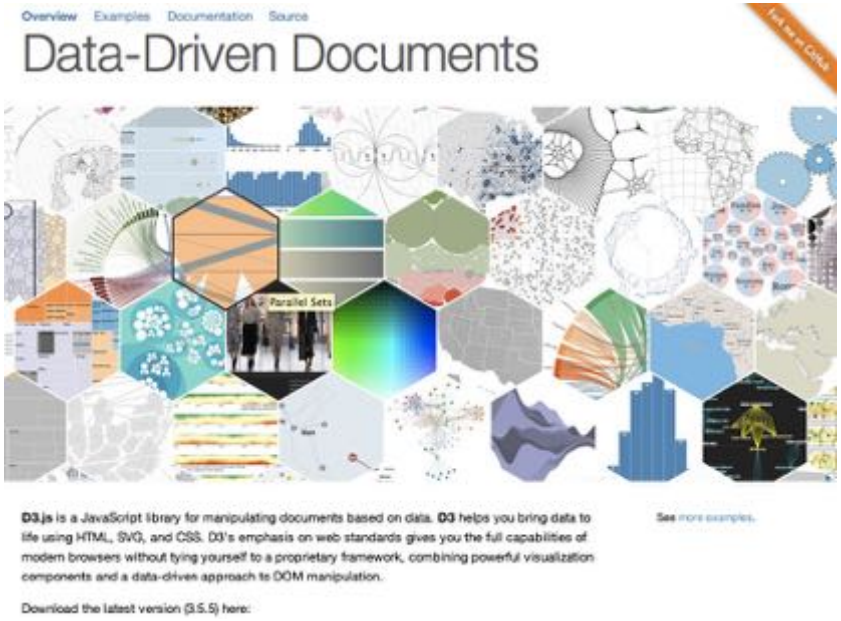


B. FRY

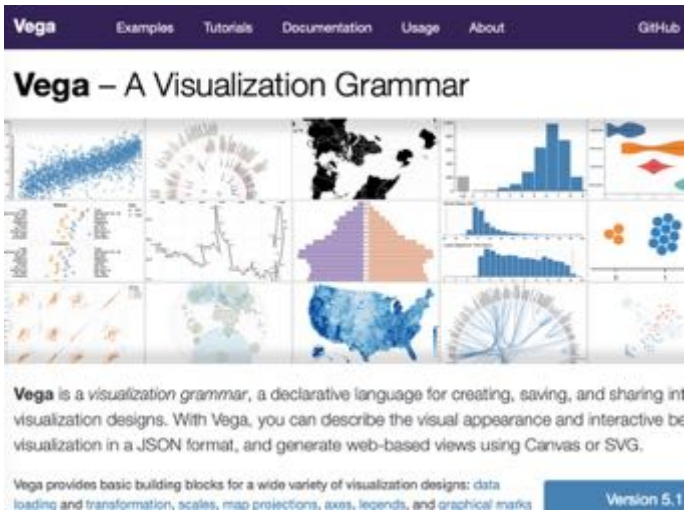
Tableau

=> pour les "dashboard"

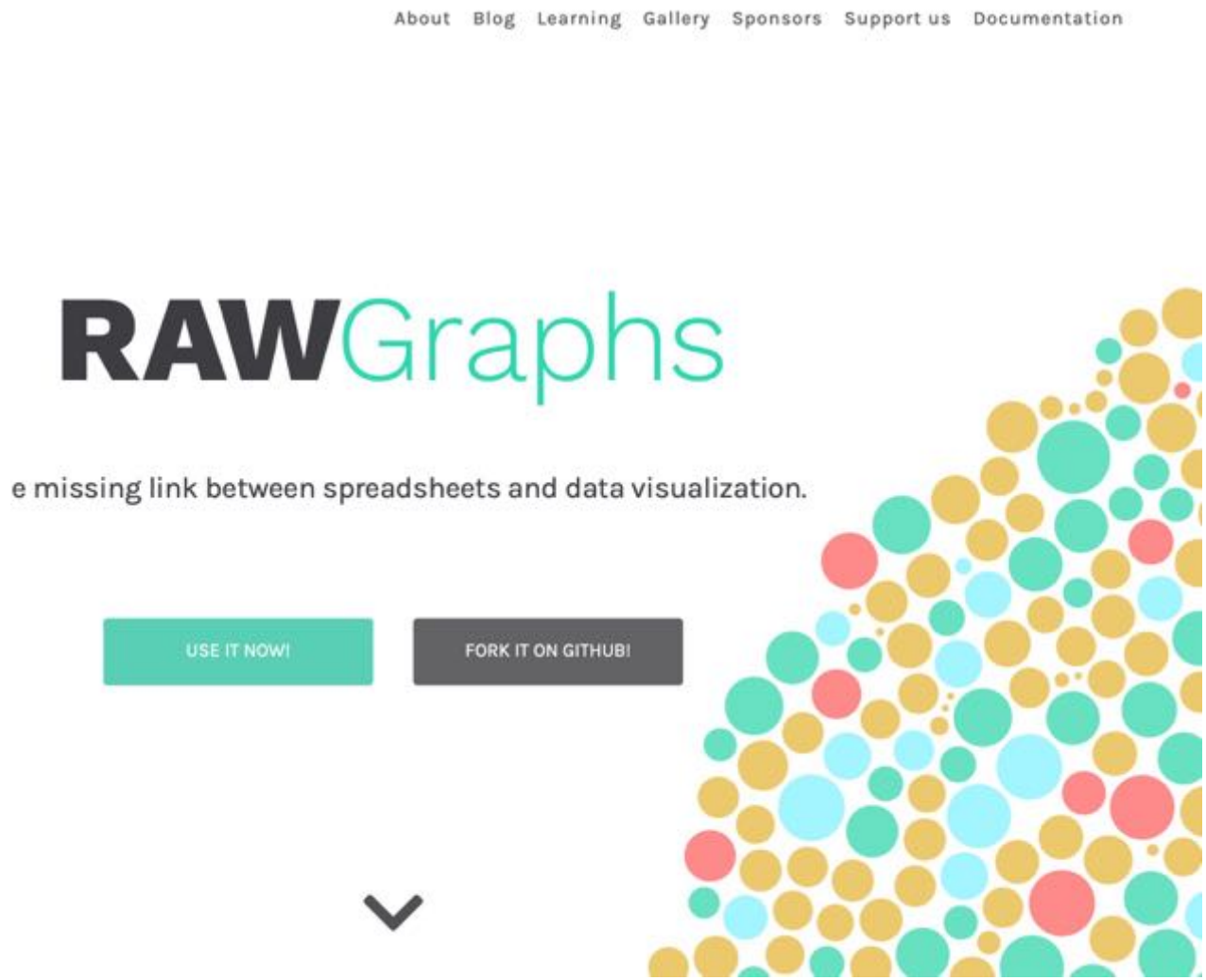
D3 http://d3js.org/



+ vega (sur github)

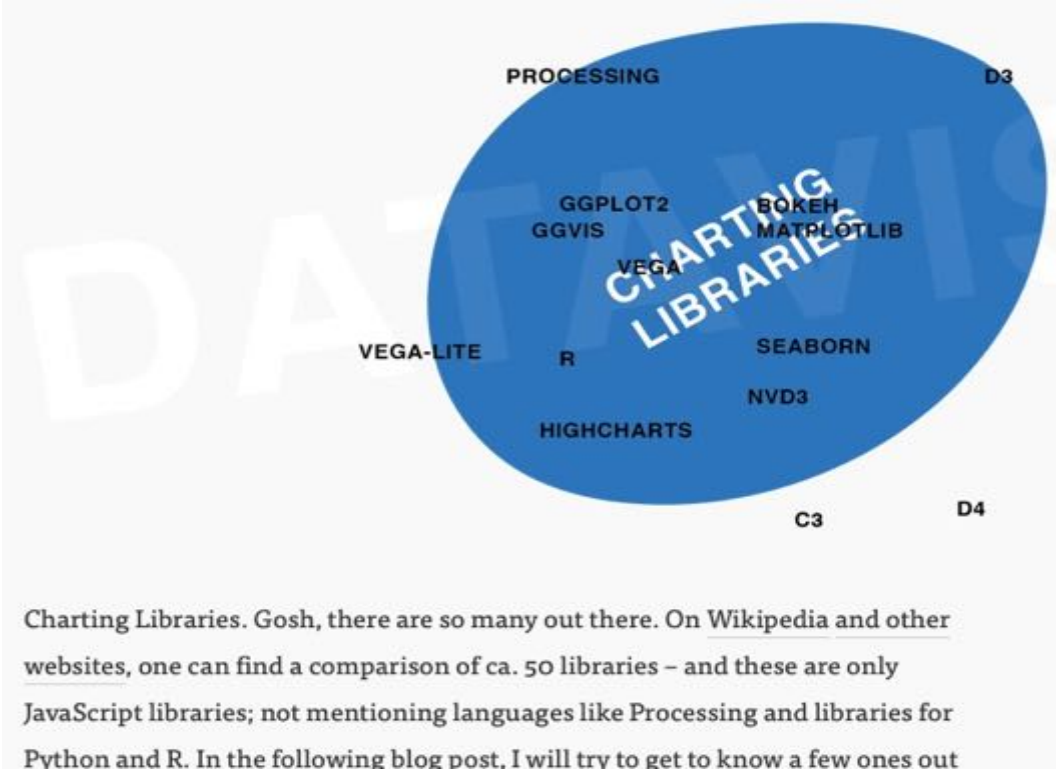


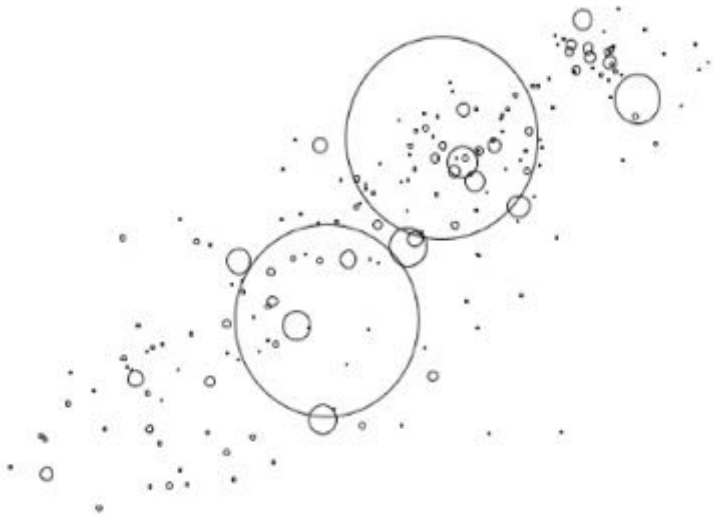
A ETUDIER EN TP ?



https://rawgraphs.io

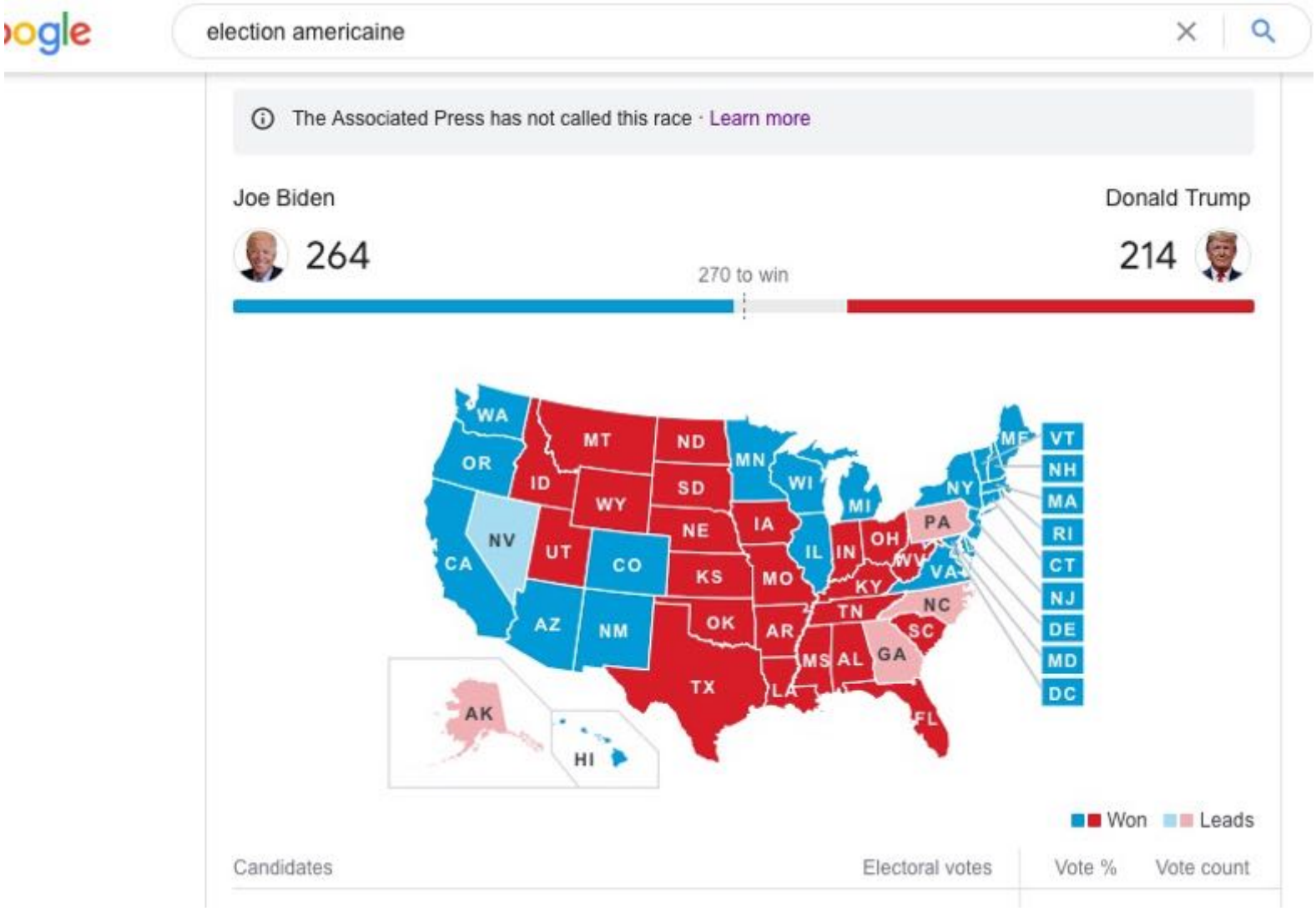
Un comparatif à étudier





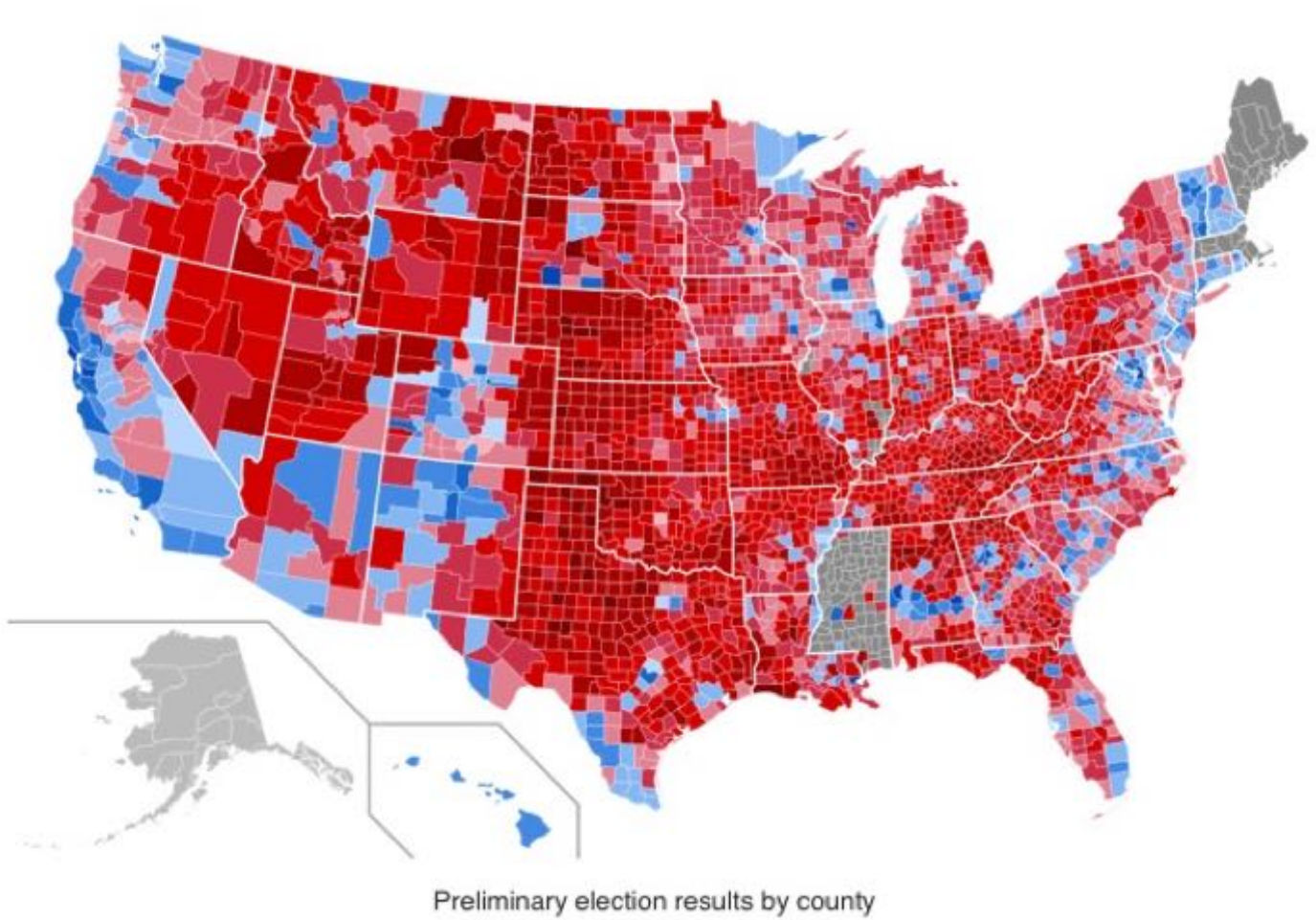
https://lisacharlottemuth.com/2016/05/17/one-chart-code/

Un exemple d'avant le COVID : élections USA



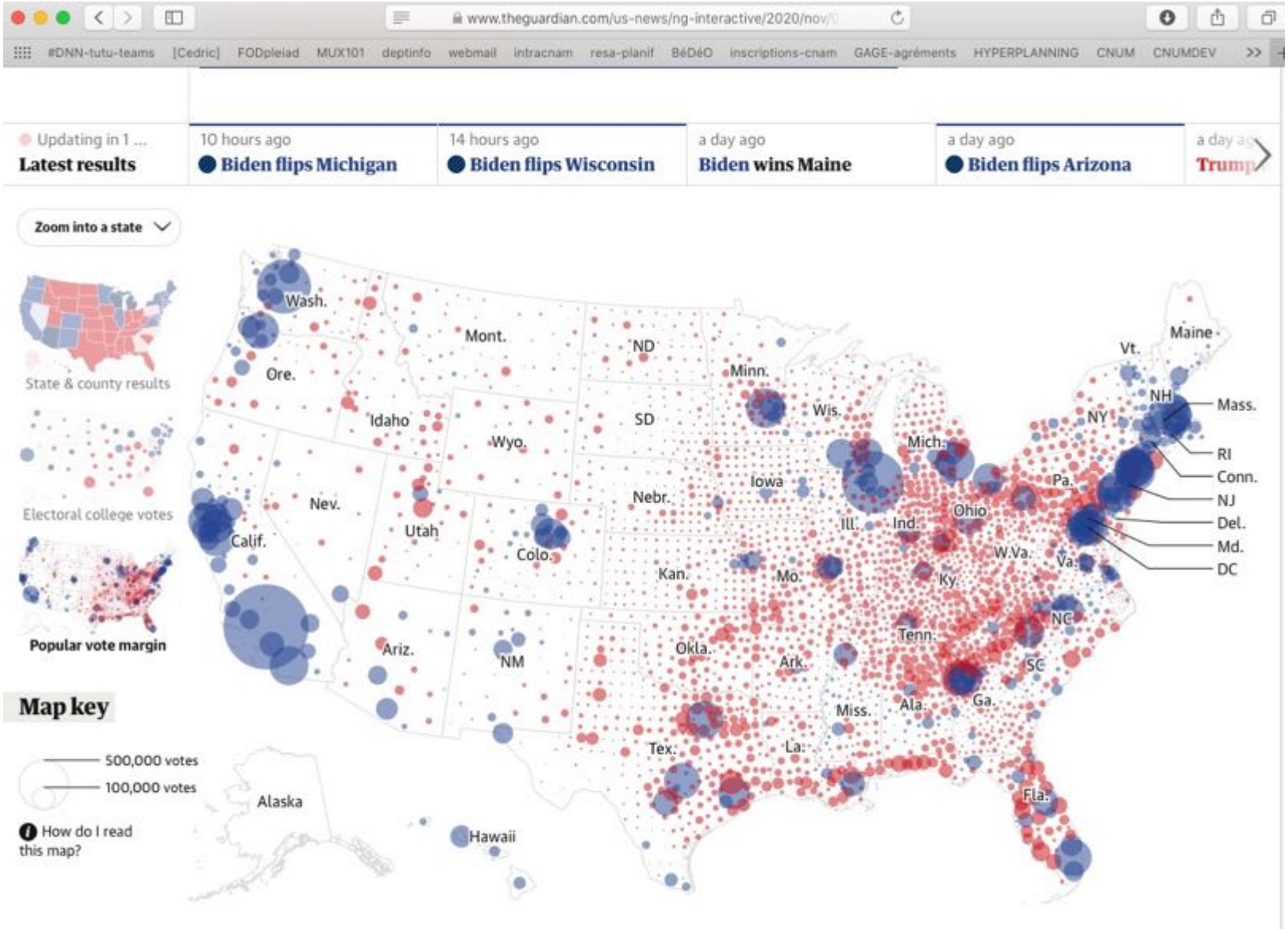
qui gagne ? rouge ou bleu ?

niveau de détail accru => permet une analyse plus fine



(wikipedia)

Prise en compte de la densité de population



idem par anamorphose => dernier cours