

Voir ou ne pas voir ?

La visualisation des données hier et
aujourd'hui

Christophe Bontemps
Toulouse School of Economics (INRA)



... À UN GRAPHIQUE

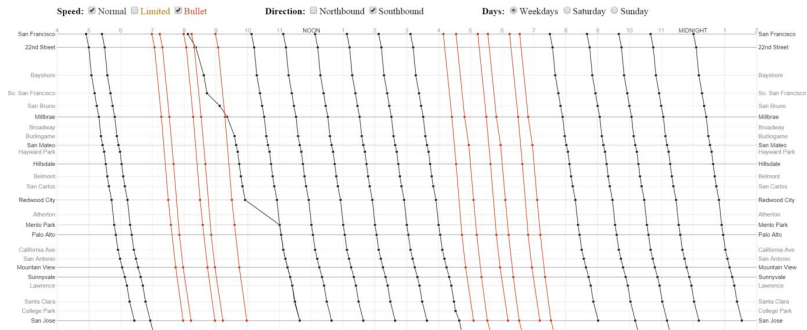


FIGURE: Horaires des trains entre San Francisco et San José (Caltrain)

... et c'est encore mieux en dynamique !

REPRÉSENTATION VISUELLE DES HORAIRES DE TRAIN... CRÉE EN 1885 !

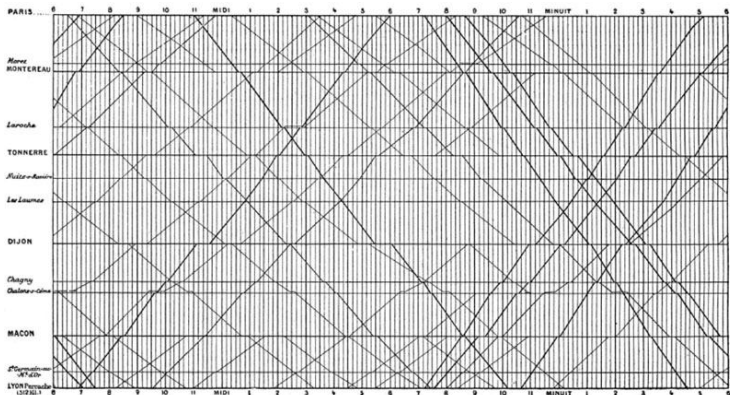


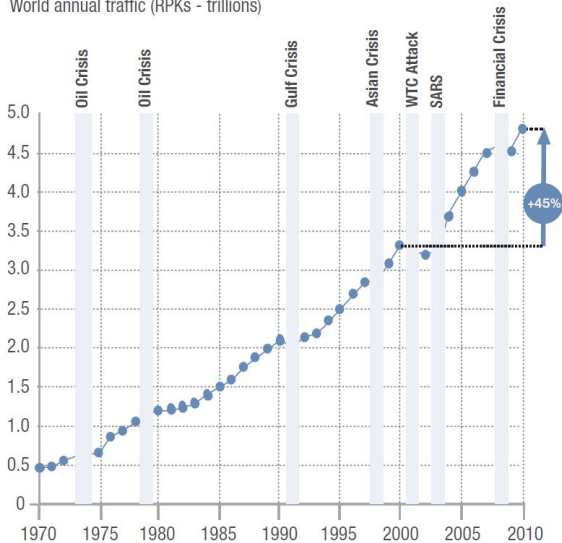
FIGURE: Horaires des trains entre Paris et Lyon- E.J. Marey (1885). Méthode due à l'ingénieur Français Ibry, ou au Russe Lt. Sergeev selon les sources (Source E. R. Tufte)

A QUOI SERT UN GRAPHIQUE ?

“On ne lit pas un graphique, on lui pose des questions”
Jacques Bertin, 1981

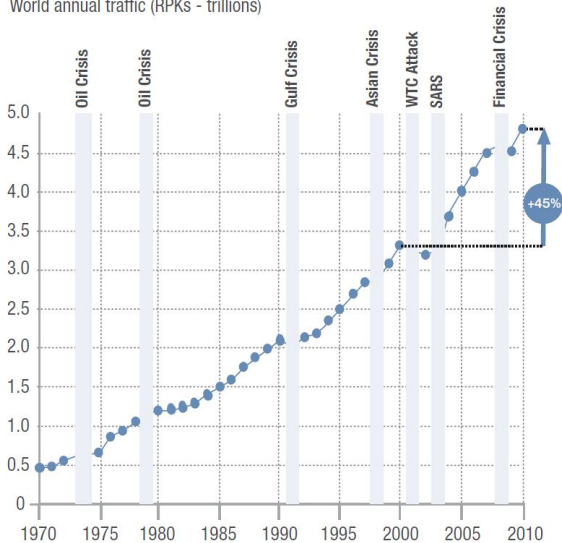
2 TYPES DE GRAPHIQUES : CEUX QUE L'ON COMPREND IMMÉDIATEMENT...

World annual traffic (RPKs - trillions)



2 TYPES DE GRAPHIQUES : CEUX QUE L'ON COMPREND IMMÉDIATEMENT...

World annual traffic (RPKs - trillions)



...ET CEUX QUE L'ON NE COMPREND PAS IMMÉDIATEMENT :



FIGURE: Visualizations showing relationships between citations in scholarly journals that are used to evaluate the importance of each journal.

QU'EST CE QU'UNE "BONNE" VISUALISATION ?

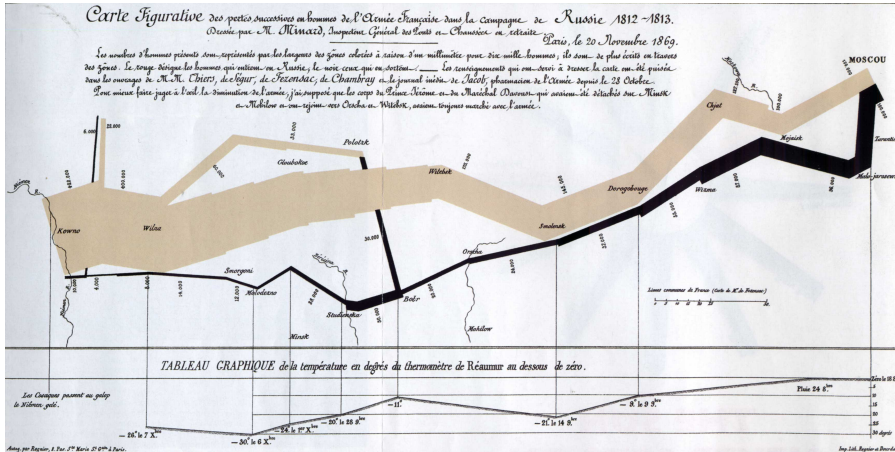


FIGURE: Carte figurative des pertes successives en hommes de l'armée Française dans la campagne de Russie en 1812-1813, Charles Minard (1869)

QU'EST CE QU'UNE "BONNE" VISUALISATION ?

DÉTAIL

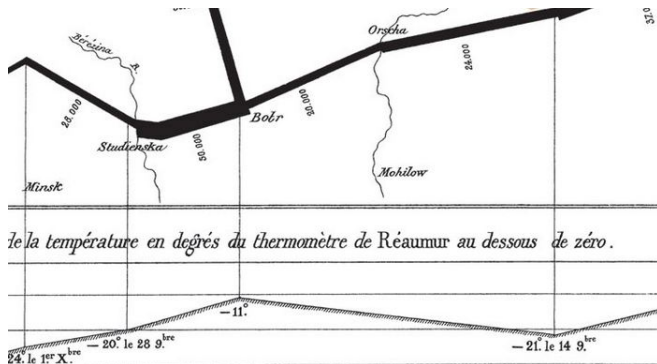


FIGURE: Carte figurative des pertes successives en hommes de l'armée Française dans la campagne de Russie en 1812-1813, Charles Minard (1869)

QU'EST CE QU'UNE "BONNE" VISUALISATION ?

DÉTAIL

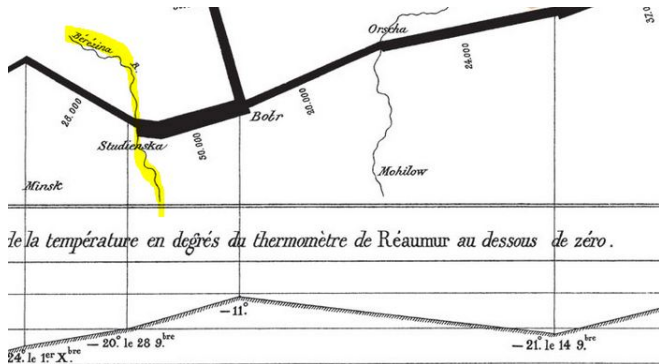


FIGURE: Carte figurative des pertes successives en hommes de l'armée Française dans la campagne de Russie en 1812-1813, Charles Minard (1869)

QU'EST CE QU'UNE "MAUVAISE" VISUALISATION ?

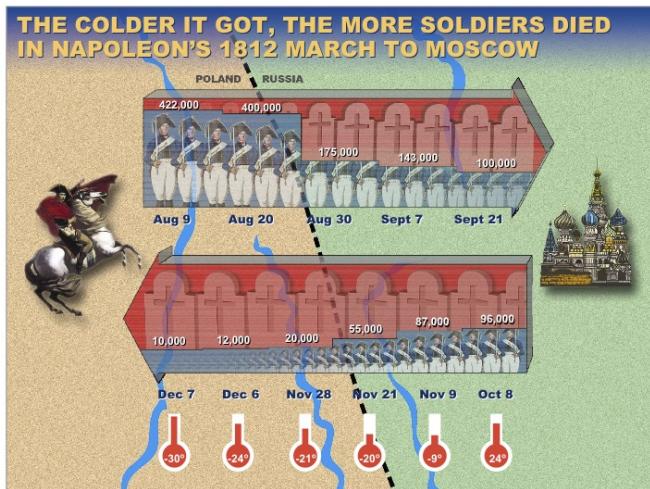
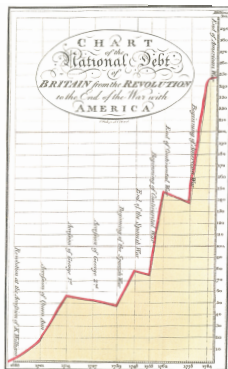


FIGURE: Carte figurative des pertes successives en hommes de l'armée Française dans la campagne de Russie en 1812-1813, Adrew Abela (2010)

EXEMPLES DE “MAUVAISES” VISUALISATIONS...



The Division of the Bottom are Years, & those on the Right hand Money.

FIGURE: Government spending "Skyrocketing". Source : E. R. Tufte from Playfair(1786). Un des rares graphiques plus haut que long ()

EXEMPLES DE "MAUVAISES" VISUALISATIONS...

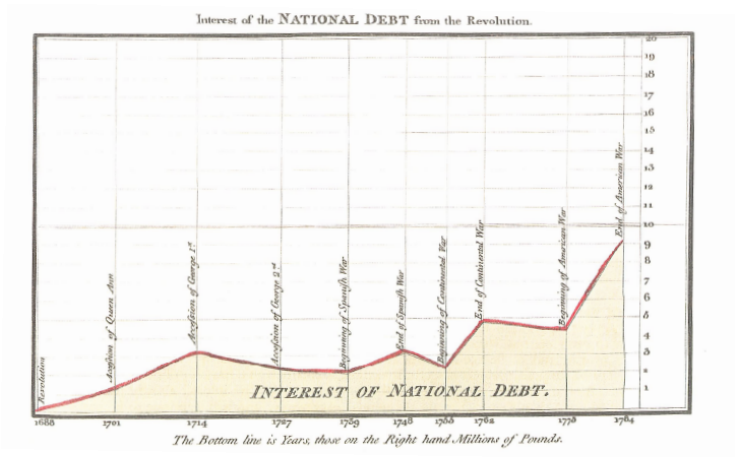


FIGURE: Government spending "Skyrocketing" (revisited)- E. R. Tuft from Playfair(1786).

EXEMPLES DE "MAUVAISES" VISUALISATIONS...

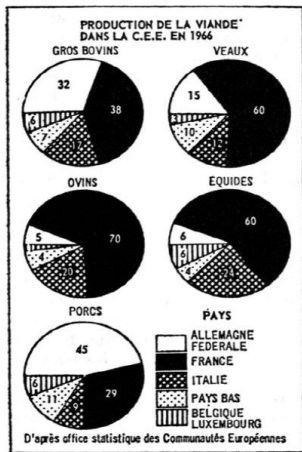


Fig. 4. Exemple de construction inefficace et inutile.

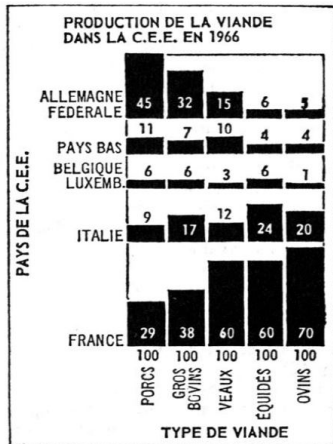
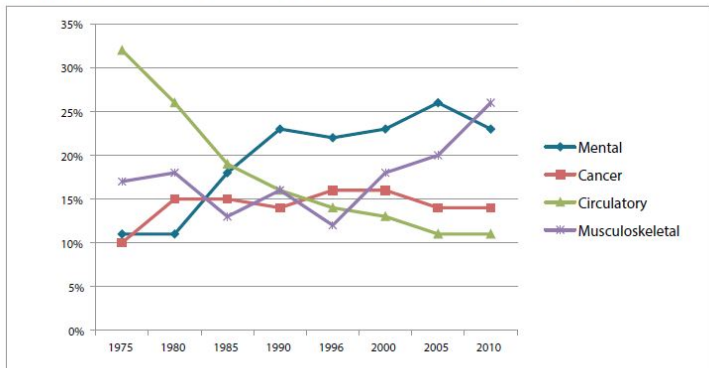


Fig. 5. Exemple de construction efficace et utile. Même information que fig. 4. Mais construction en x , y , z . En x , les types de viande; en y , les pays de la C.E.E.; en z , les quantités.

EXEMPLES DE “MAUVAISES” VISUALISATIONS : SPAGHETTIS

27. Initial DI Worker Awards by Major Cause of Disability—Calendar Years 1975-2010



Source: Social Security Advisory Board (2012).

FIGURE: Major Cause of Disability - 1975-2010 (J. Schwabish, 2014).

EXEMPLES DE “MAUVAISES” VISUALISATIONS : SPAGHETTIS

Initial DI Worker Awards by Major Cause of Disability—
Calendar Years 1975–2010
(Percent)

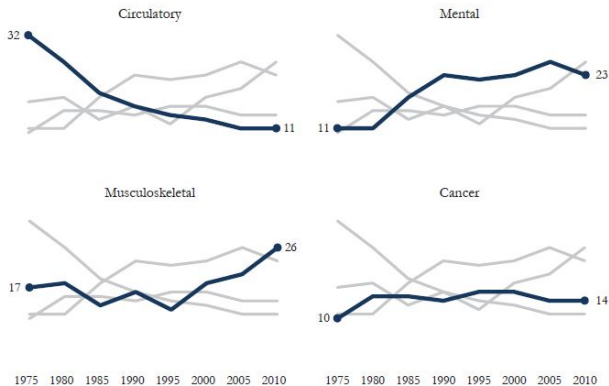


FIGURE: Major Cause of Disability- 1975-2010 (J. Schwabish).

“BONS” OU “MAUVAISES” GRAPHIQUES ?

“Il n’y a pas de bons ou de mauvais diagrammes (...), il y a des constructions qui répondent et d’autres qui ne répondent pas aux questions qu’on est en droit de se poser ”

Jacques Bertin, 1981

PETIT TEST...TROUVER LES BONNES PROPORTIONS

Sur l'ensemble des prisonniers américains :

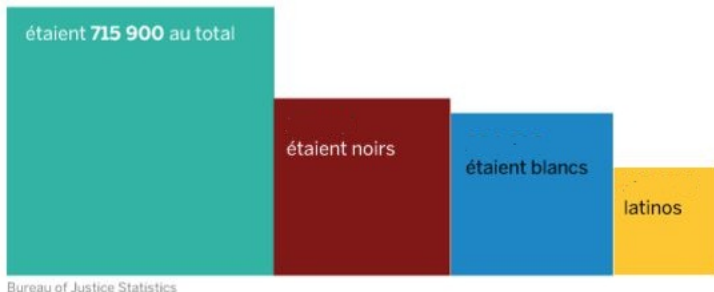


FIGURE: Composition ethnique des prisonniers purgeant une peine en 2008 aux USA. (Le Monde 5/12/2014)

PAS SI SIMPLE...

Sur l'ensemble des prisonniers américains :

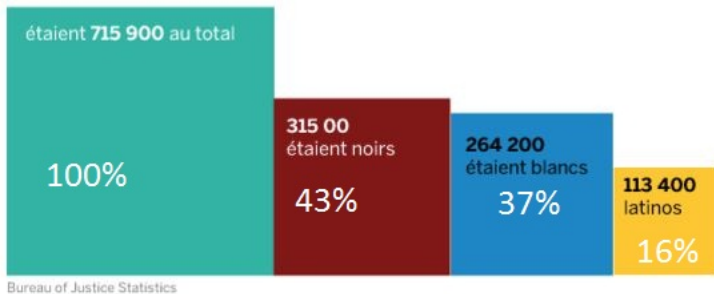
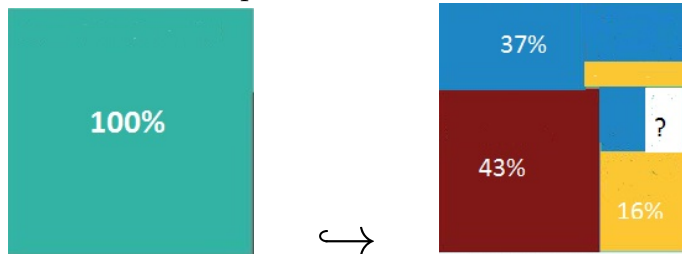


FIGURE: Composition ethnique des prisonniers purgeant une peine en 2008 aux USA. (Le Monde 5/12/2014)

VERIFICATION

Sur l'ensemble des prisonniers américains :



LES “règles” D’EDWARD R. TUFTE

1. *Avant tout, montrer les données*
2. *Maximiser le ratio encre/données*
3. *Effacer ce qui n’est pas les données*
4. *Effacer ce qui est redondant*
5. *Reviser and refaire*

EXEMPLE D'EPURATION À LA TUFTE

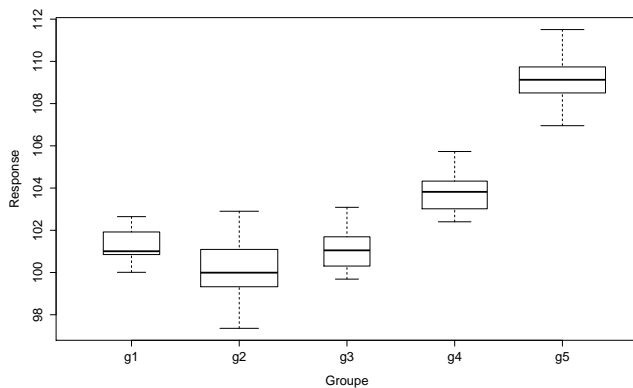


FIGURE: Distribution d'une variable continue sur 4 groupes d'individus

EXEMPLE D'EPURATION À LA TUFTE

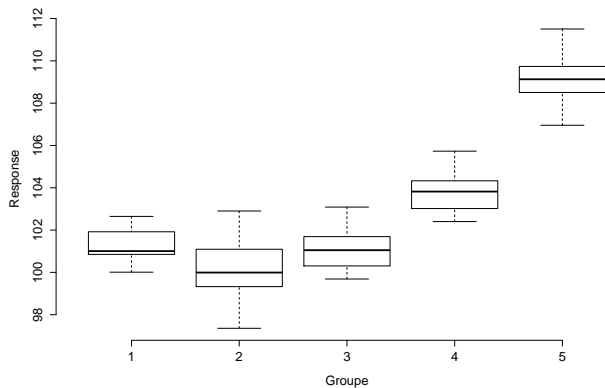


FIGURE: Distribution d'une variable continue sur 4 groupes d'individus

EXEMPLE D'EPURATION À LA TUFTE

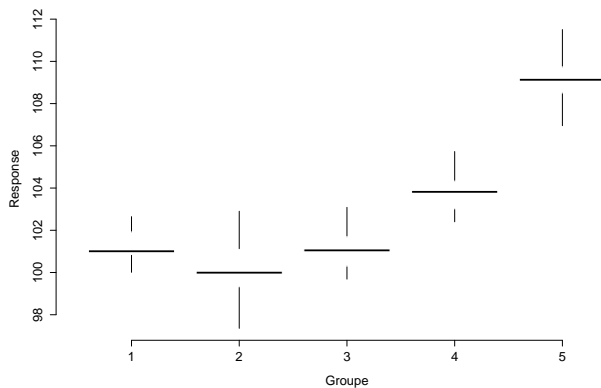


FIGURE: Distribution d'une variable continue sur 4 groupes d'individus

EXEMPLE D'EPURATION À LA TUFTE

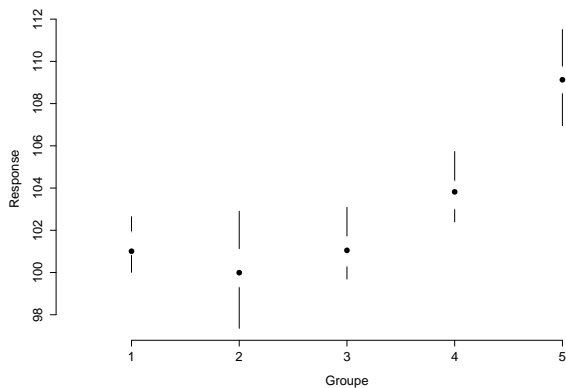


FIGURE: Distribution d'une variable continue sur 4 groupes d'individus

ET, PUISQU'IL FAUT MONTRER LES DONNÉES...

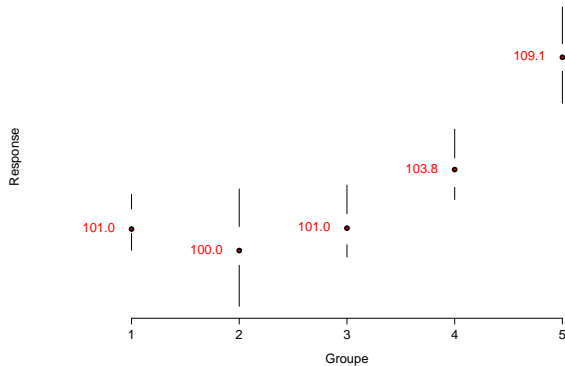
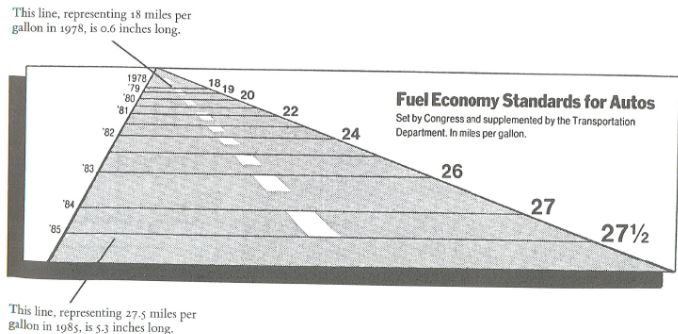


FIGURE: Distribution d'une variable continue sur 4 groupes d'individus

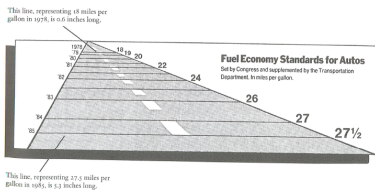
LE FACTEUR DE MENSONGE (E. TUFTE)

$$\text{LieFactor} = \frac{\text{Taille de l'effet visuel}}{\text{Taille de l'effet dans les donnees}} \quad (1)$$



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

FIGURE: Fuel economy standards. (E. Tufte - from NY Times 1978)



New York Times, August 9, 1978, D-1.

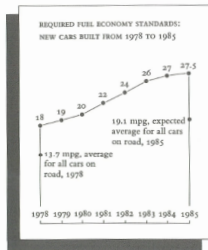
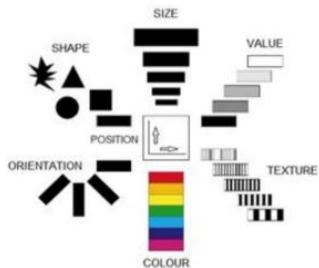


FIGURE: Fuel economy standards (revisited)

La ligne de "18 mpg" mesure 1.5 cm (1978) ; celle de "27,5 mpg" mesure 13 cm (1985)

→ Lie factor = 14.5% !!!

LES VARIABLES VISUELLES DE JACQUES BERTIN



- ▶ Position dans le plan
- ▶ Taille
- ▶ Valeur
- ▶ Grain
- ▶ Couleur
- ▶ Orientation
- ▶ Forme

LES VARIABLES VISUELLES DE JACQUES BERTIN

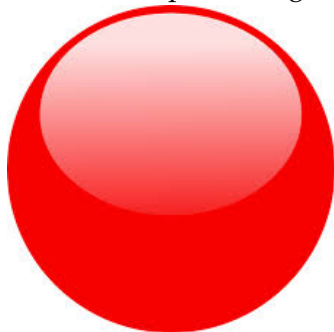
	Niveau d'organisation des variables visuelles			
Position dans le plan	≡	≠	O	∝
Taille	≡	≠	O	∝
Valeur	≡	≠	O	∝
Grain	≡	≠	O	
Couleur	≡	≠		
Orientation	≡	≠		
Forme	≡			

≡ : Equivalence ≠ : différences ; O : relation d'Ordre ;

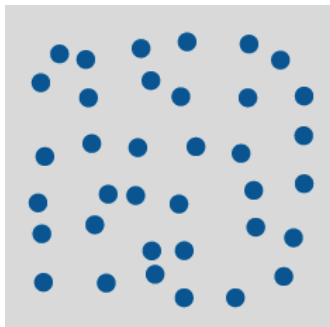
∝ : Proportionnalité

ET MAINTENANT...

Cherchez le point rouge !

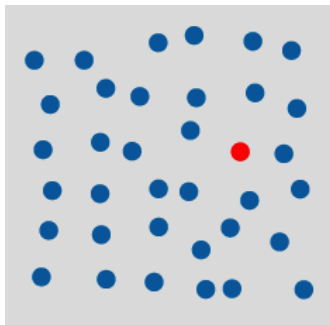


TESTS : TROUVER LE POINT ROUGE



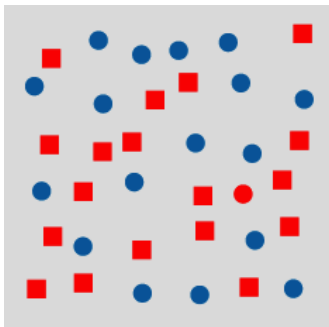
TESTS : TROUVER LE POINT ROUGE

TESTS : TROUVER LE POINT ROUGE



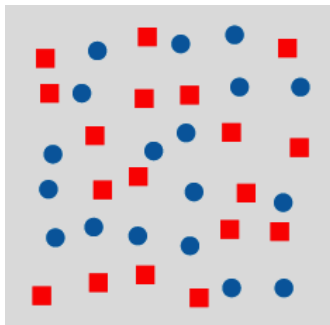
TESTS : TROUVER LE POINT ROUGE

TESTS : TROUVER LE POINT ROUGE



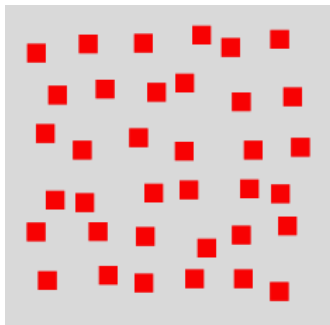
TESTS : TROUVER LE POINT ROUGE

TESTS : TROUVER LE POINT ROUGE



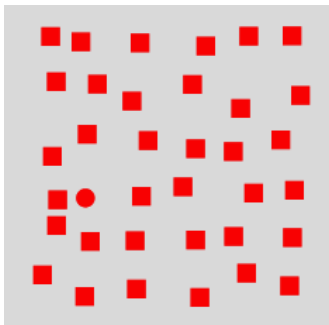
TESTS : TROUVER LE POINT ROUGE

TESTS : TROUVER LE POINT ROUGE



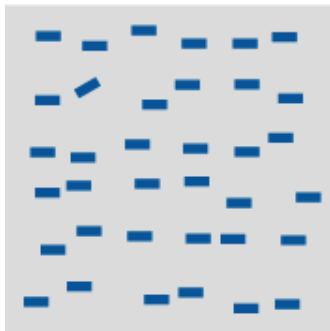
TESTS : TROUVER LE POINT ROUGE

TESTS : TROUVER LE POINT ROUGE



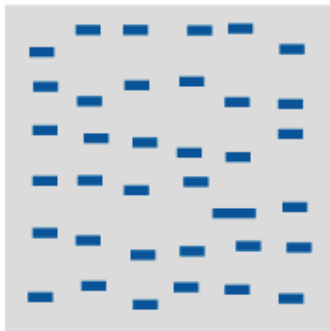
PLUS DUR : Y A T'IL UN ÉLÉMENT DIFFÉRENT ?

PLUS DUR : Y A T'IL UN ÉLÉMENT DIFFÉRENT ?



PLUS DUR : Y A T'IL UN ÉLÉMENT DIFFÉRENT ?

PLUS DUR : Y A T'IL UN ÉLÉMENT DIFFÉRENT ?



PLUS DUR : Y A T'IL UN ÉLÉMENT DIFFÉRENT ?

PLUS DUR : Y A T'IL UN ÉLÉMENT DIFFÉRENT ?



C'ÉTAIT PAS FACILE !

- ▶ Concept de perception préattentive (Anne Treisman, 1985)
- ▶ Certaines propriétés visuelles sont détectées très vite
- ▶ Mais interférences entre elles (exemple couleur et forme)
- ▶ Il existe des tonnes d'exemples et de conseils sur cela

ILLUSTRATION : FORME ET PROPORTIONALITÉ

Le prix du terrain dans la France de l'Est (J. Bertin 1970)

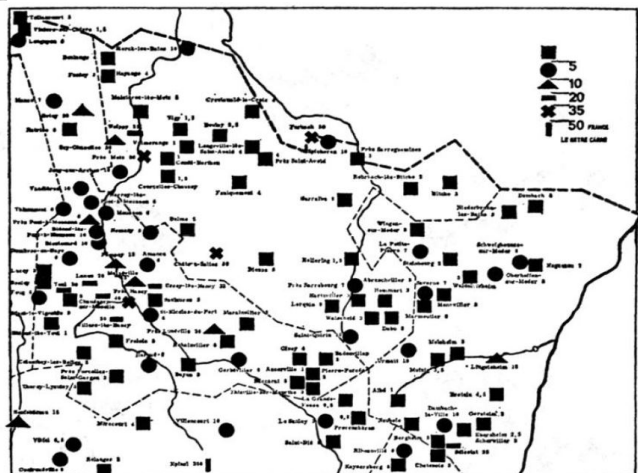
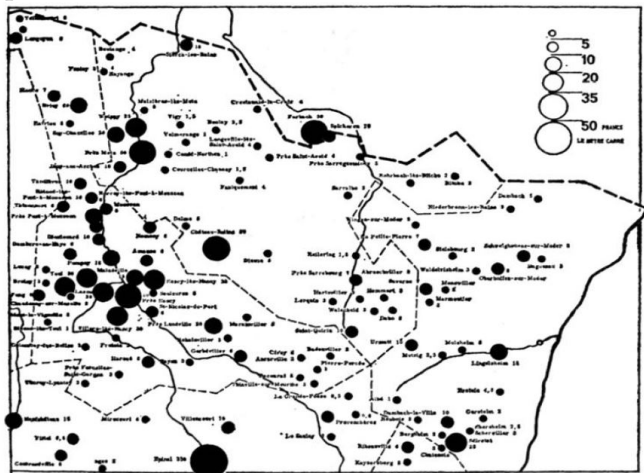


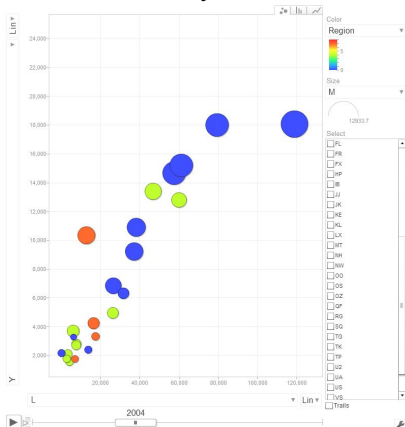
ILLUSTRATION : TAILLE ET PROPORTIONALITÉ

Le prix du terrain dans la France de l'Est (J. Bertin 1970)



COULEUR, TAILLE ET PROPORTIONALITÉ

Productivity of Airlines



(voir la démo dans 1 minute)

POURQUOI LA VISUALISATION ?

“The greatest value of a picture is when it forces us to notice what we never expected to see”

John Tukey, 1977

Impact de la visualisation pour comprendre/monttrer un pb :

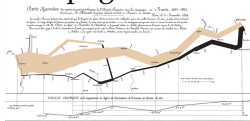
- ▶ London Cholera Map - John Snow (1854)
- ▶ Campagne de Russie- Minard (1861)
- ▶ War Mortality - Florence Nightingale (1855)
- ▶ Philipps curve - A. W. Phillips (1958)

POURQUOI LA VISUALISATION ? : IMPACTS !

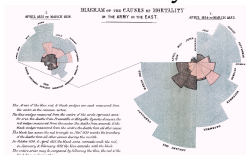
- ▶ London Cholera Map - John Snow (1854)



- ▶ Campagne de Russie- Minard (1861)



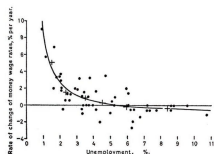
- ▶ War Mortality - Florence Nightingale (1855)



IMPACTS AUJOURD'HUI !

- ▶ Philipps curve - A. W. Phillips (1958)

1861-1913



- ▶ Gapminder (Hans Rosling)



(Demo !)

- ▶ Google map et la concurrence...



A QUOI SERT LA VISUALISATION DES DONNÉES ?

Ce qu'en dit Jacques Bertin.

1. *A traiter des données pour comprendre et en tirer de l'information*
↔ Idée de la *fouille* de données
2. *A communiquer cette information*
↔ Il faut quelques heures pour construire un graphique ; quelques secondes pour le lire.

*“Communiquer implique la simplification.
Traiter des données implique l'exhaustivité”.*

A QUOI SERT LA VISUALISATION DES DONNÉES EN STATISTIQUE ÉCONOMÉTRIE

1. *A traiter des données pour comprendre et en tirer de l'information*
 - ↔ résumés graphiques, intuitions, réduction de la complexité/dimension
2. *A communiquer cette information*
 - ↔ Présentation de résultats synthétiques
 - ↔ Comparaison de résultats, de modèles, d'hypothèses.

EX. 1 : DONNÉES D'ANSCOMBE (1973)

4 couples de variables (X_1, Y_1) , (X_2, Y_2) , (X_3, Y_3) and (X_4, Y_4) ,
aparamment avec les mêmes distributions et les mêmes
corrélations 2 à 2.

Xs	Mean	Std. Dev.	Ys	Mean	Std. Dev.	<i>corr</i> (X_i, Y_i)	N
X_1	9	3.32	Y_1	7.5	2.03	0.8164	11
X_2	9	3.32	Y_2	7.5	2.03	0.8162	11
X_3	9	3.32	Y_3	7.5	2.03	0.8163	11
X_4	9	3.32	Y_4	7.5	2.03	0.8165	11

LE JEU DE DONNÉES D'ANSCOMBE (1973)

4 petites régressions ($Y_i = \alpha + \beta X_i + \epsilon_i$) confirment que l'on a les mêmes relations entre ces variables 2 à 2

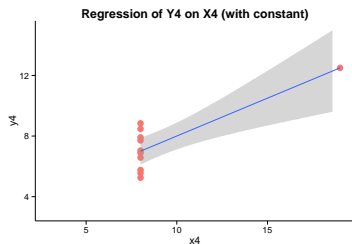
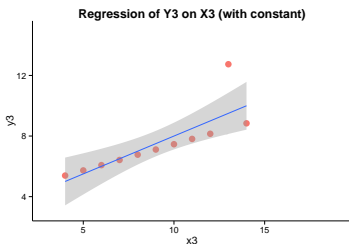
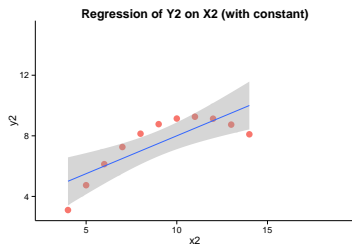
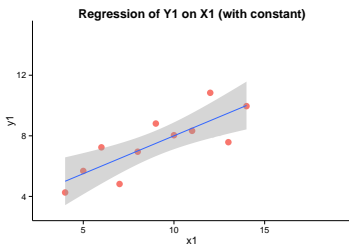
	<i>Dependent variable :</i>			
	Y_1	Y_2	Y_3	Y_4
Regressed on :	X_1	X_2	X_3	X_4
	0.500 ***	0.500***	0.500***	0.500***
Constant	3.000**	3.001**	3.002**	3.002**
R^2	0.667	0.666	0.666	0.667
Resid Std. Error	1.237	1.237	1.236	1.236
F Statistic	17.990***	17.966***	17.972***	18.003***

Note : Data from Anscombe (1973).

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

LE JEU DE DONNÉES D'ANSCOMBE (1973)

Une visualisation simple donne une toute autre information



EX. 2 : TABLEAUX ET MATRICES (J. BERTIN)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P			
							■			■							1	COLLÈGE
	■	■	■				■	■			■				■		2	COOPÉRATIVE AGRIC.
								■									3	GARE
■				■	■			■	■		■	■	■	■	■		4	ÉCOLE CLASSE UNIQUE
	■						■				■				■		5	VÉTÉRINAIRE
■		■	■	■	■			■	■			■	■	■	■		6	PAS DE MÉDECIN
									■					■			7	PAS D'ADDUCTION D'EAU
		■						■			■						8	GENDARMERIE
	■	■	■				■				■				■		9	REMBREMENT

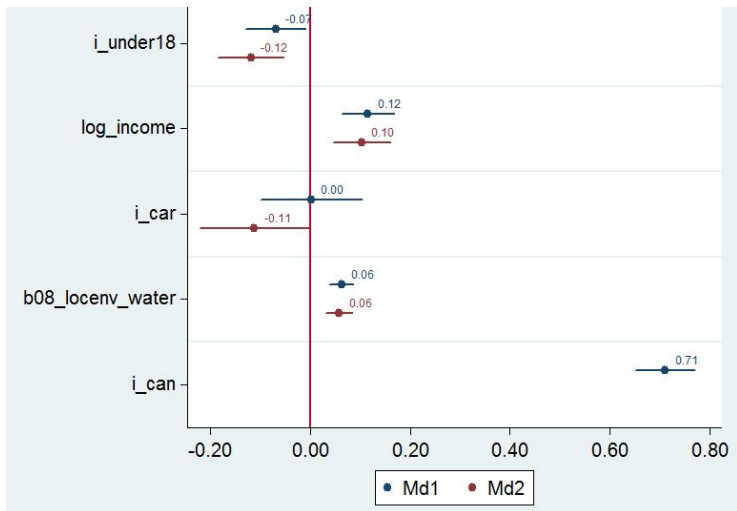
TABLEAUX ET MATRICES (J. BERTIN)

VILLAGES													BOURGS				VILLES			
N	J	P	M	I	F	E	A	B	O	L	G	D	C	H	K					
																1	COLLÈGE	6	URBAIN	
																3	GARE			
																8	GENDARMERIE			
																2	COOPÉRATIVE AGRIC.	→		
																5	VÉTÉRINAIRE			
																9	REMBREMENT	→		
																4	ÉCOLE CLASSE UNIQUE	→		
																6	PAS DE MÉDECIN		RURAL	
																7	PAS D'ADDUCTION D'EAU			

EX. 3 : REGRESSION COMME UN TABLEAU

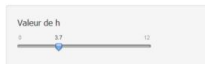
	(Mod. 1) Special	(Mod. 2) Special
i_under18	-0.0692* (-2.25)	-0.119*** (-3.57)
log_income	0.116*** (4.31)	0.102*** (3.51)
i_car	0.00131 (0.03)	-0.112* (-2.00)
b08_locenv_water	0.0624*** (4.99)	0.0583** (4.28)
i_can	0.710*** (23.27)	
Constant	-1.467*** (-5.38)	-0.961** (-3.24)

REGRESSION COMME UN GRAPHIQUE

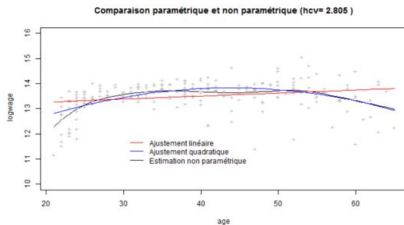


LE FUTUR (DE LA VISUALISATION) SERA T'IL DYNAMIQUE ?

Regression Dynamique (Café IMT)

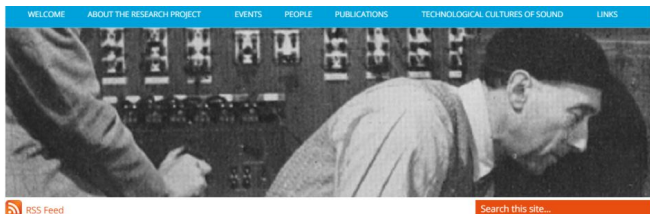


Linear NP-Reg Quadratic Comparaison



Demo

LE FUTUR (DE LA VISUALISATION) SERA T'IL ACOUSTIQUE ?



SONIFICATION OF SCIENTIFIC DATA

Main researcher: Alexandra Supper

This project investigates the practice of the auditory display of scientific data, also known as data sonification. Since the early 1990s, a small interdisciplinary community of scientists and artists has been systematically exploring the possibility of using sound to represent scientific data – sometimes as an alternative, other times as a supplement to the more established practice of data visualization. The project has traced the emergence of the sonification community, and in particular, their efforts to have listening to scientific data accepted as a credible scientific technique. The fact that sonification has received increasing media attention in recent years has not always been particularly helpful in these efforts, as the media discourse tends to emphasize listening as a source of sublime emotional experiences – an idea which is quite at odds with the notion that listening can serve analytic insights and



NEWS

- **Website on Sonic Science Festival**
10 Oct 2014
In January 2015, the Sonic Science Festival will take place, [Read more](#) →
- **New PhD Project on traffic information and car radio**
22 May 2014
The PhD project "Talking You Through: Traffic Information and Car" [Read more](#) →
- **Sonic Skills on the Radio**

(source : Maastricht University)

LE FUTUR SERA T'IL ACOUSTIQUE ?

the audiolyzR

Select an audioPlot
Hemoglobin-BMI.json

Transpose!

Click Here to Loop

Hemoglobin

BMI

Volume: 1.9

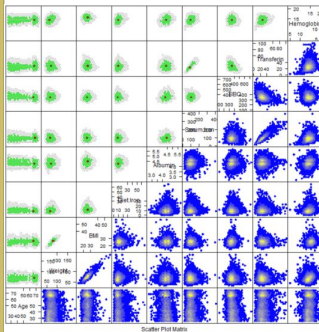
Tempo: 115 bpm

Pause between loops (in seconds) 0.2

Chord Qualities

- Major
- Minor
- Augmented
- Diminished
- V 7

Starting note 40 (MIDI Scale)



CONCLUSION ?

- ▶ On trouve de nombreuses visualisations (“dataviz”) dans de nombreux domaines
- ▶ Beaucoup d’outils (Google, D3.js, R, Tableau, ..)
- ▶ Difficile de s’y retrouver..
- ▶ Beaucoup sont **difficiles à interpréter**, peu apportent des éléments intéressants
- ▶ Savant mélange d’expression artistique et de statistique
- ▶ Conscience actuelle de l’importance des graphiques pour les sciences (cours, livres, papiers,)
- ▶ Visualisation dynamique ou interactive très à la mode

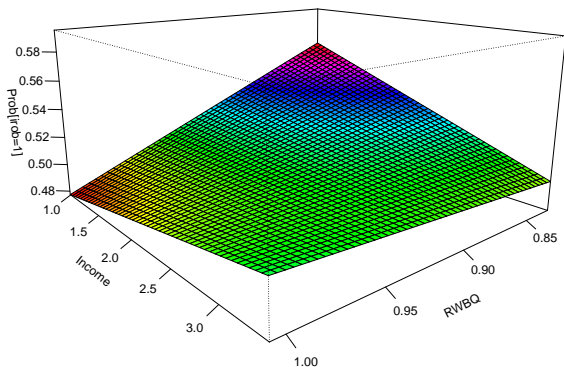
MERCI!



QUELQUES GRAPHIQUES AUXQUELS VOUS AVEZ ÉCHAPPÉS)

- Probabilité de boire de l'eau du robinet comme une fonction du **revenu** et d'un **indice de qualité** .

Estimated conditional prob of drinking tap water using Probit estimator

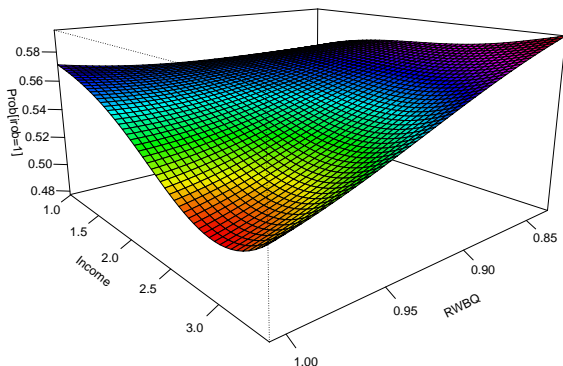


Region= 3 , Retired= 0 , Diploma= 1 , Rural= 0 .

QUELQUES GRAPHIQUES AUXQUELS VOUS AVEZ ÉCHAPPÉS)

- Probabilité de boire de l'eau du robinet comme une fonction du **revenu** et d'un **indice de qualité** .

Estimated conditional prob of drinking tap water using NP estimator

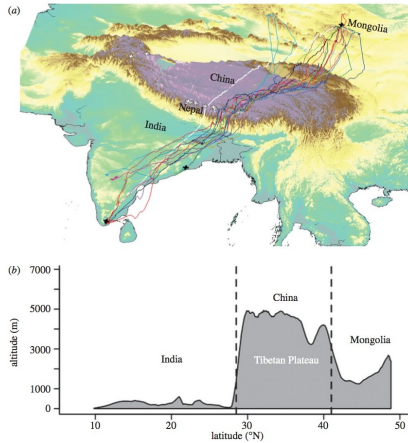


Region= 3 , Retired= 0 , Diploma= 1 , Rural= 0 .

AIR TRANSPORTATION



OIES SAUVAGES



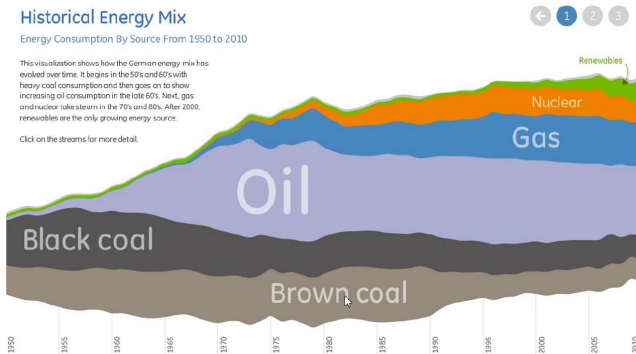
ENERGIE

Historical Energy Mix

Energy Consumption By Source From 1950 to 2010

This visualization shows how the German energy mix has evolved over time. It begins in the 50's and 60's with heavy coal consumption and then goes on to show increasing oil consumption in the late 60's. Next, gas and nuclear take steam in the 70's and 80's. After 2000, renewables are the only growing energy source.

Click on the streams for more detail.



Imagination at work

Sources: AG Energiebilanzen, Kolert, BPE

D3.js



LIVING STANDARDS COMPARISON

