

DATA VISUALISATION

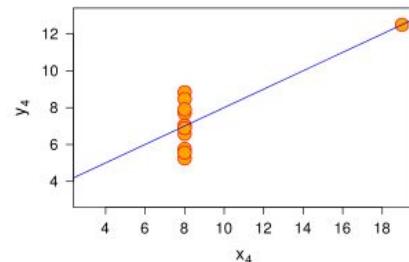
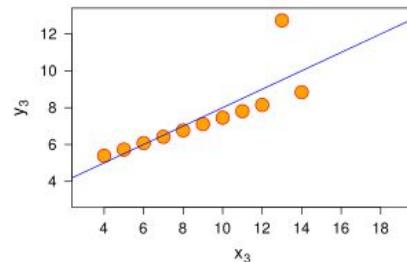
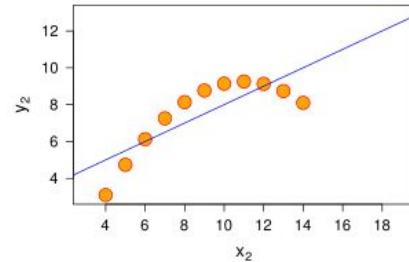
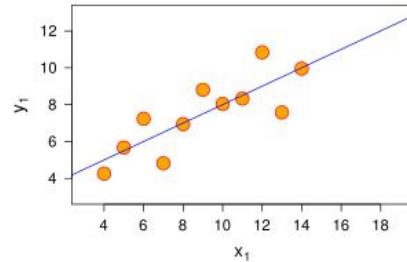
Quelques éléments de réflexion...

Pourquoi visualiser les données...

Informations supplémentaires

Quartet d'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



Moyenne, variance, corrélation, droite de régression identiques

Un volume toujours croissant...



Volume

10^{24}	yotta
10^{21}	zetta
10^{18}	exa
10^{15}	péta
10^{12}	téra
10^9	giga
10^6	méga
10^3	kilo

quadrillion

trillion

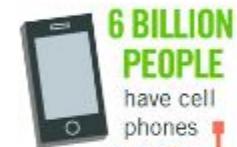
billion

million

40 ZETTABYTES

[43 TRILLION GIGABYTES]

of data will be created by 2020, an increase of 300 times from 2005



WORLD POPULATION: 7 BILLION



2005

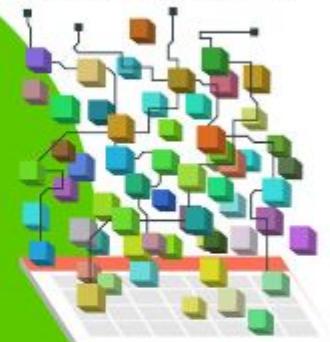
Volume
SCALE OF DATA

It's estimated that

2.5 QUINTILLION BYTES

[2.3 TRILLION GIGABYTES]

of data are created each day



Most companies in the
U.S. have at least

100 TERABYTES

[100,000 GIGABYTES]
of data stored



Variété

As of 2011, the global size of data in healthcare was estimated to be

150 EXABYTES

[161 BILLION GIGABYTES]



**30 BILLION
PIECES OF CONTENT**

are shared on Facebook every month



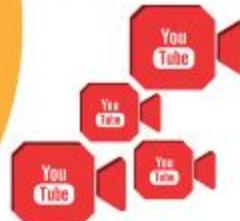
Variety DIFFERENT FORMS OF DATA

By 2014, it's anticipated there will be

**420 MILLION
WEARABLE, WIRELESS
HEALTH MONITORS**



**4 BILLION+
HOURS OF VIDEO**
are watched on YouTube each month



400 MILLION TWEETS
are sent per day by about 200 million monthly active users

Vitesse

The New York Stock Exchange captures
captures

1 TB OF TRADE INFORMATION

during each trading session



Modern cars have close to
100 SENSORS
that monitor items such as
fuel level and tire pressure

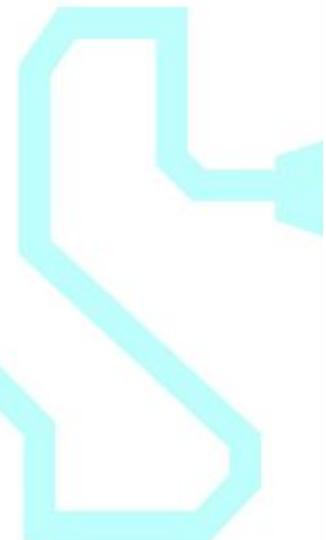
Velocity

ANALYSIS OF STREAMING DATA

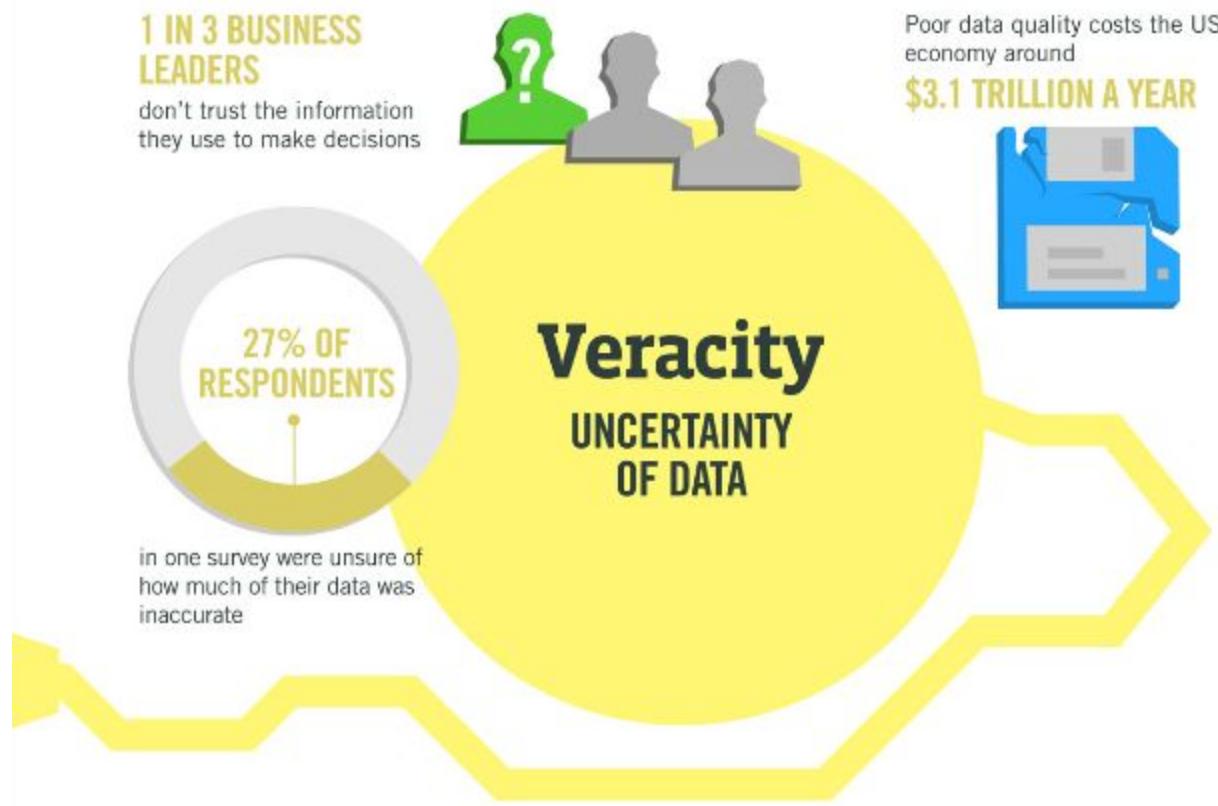
By 2016, it is projected
there will be

**18.9 BILLION
NETWORK CONNECTIONS**

– almost 2.5 connections
per person on earth



Véracité



Types de données

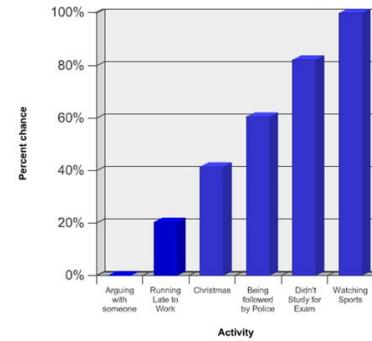
Données catégorielles

Valeurs ou observations qui peuvent être regroupées par catégories. Aucun ordonnancement

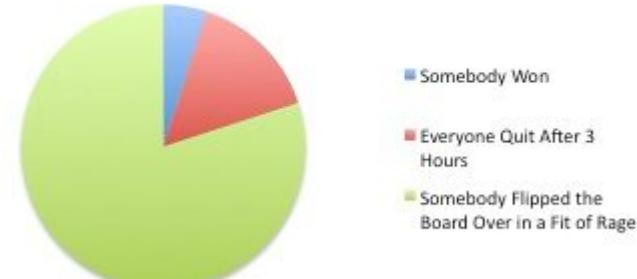
Opérations : = , ≠

Exemple : genre (masculin, féminin), couleur des yeux

Chances of an Athiest believing in God

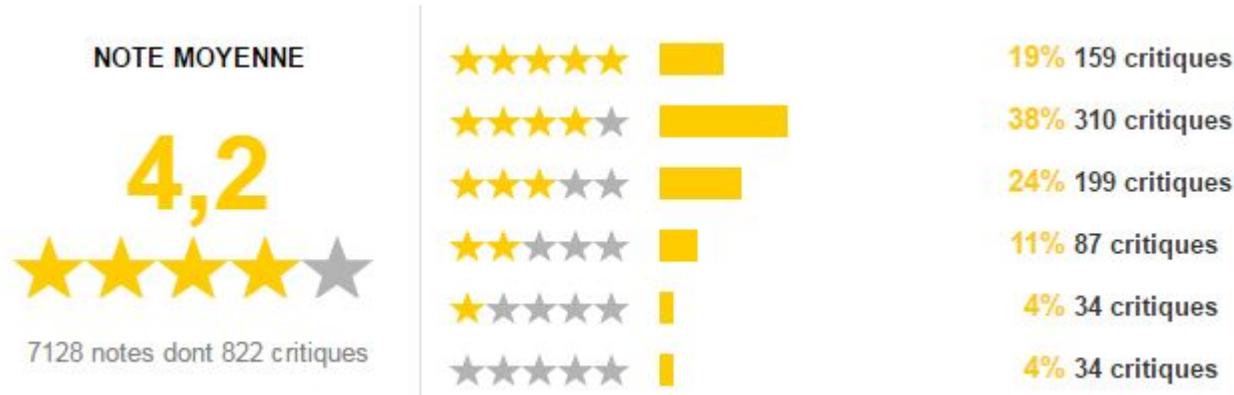


Reasons for Ending a Game of Monopoly



Données ordinales

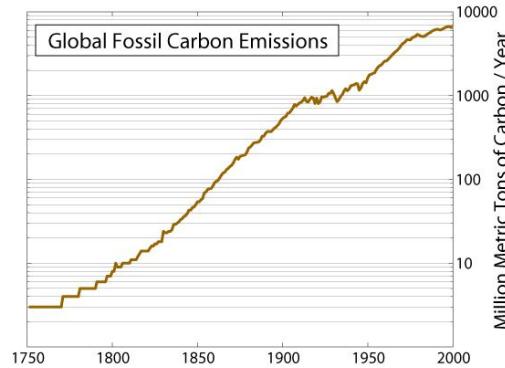
variable catégorielle numérique qui possède une propriété d'ordonnancement
opérations : =, ≠, <, >



Données numériques

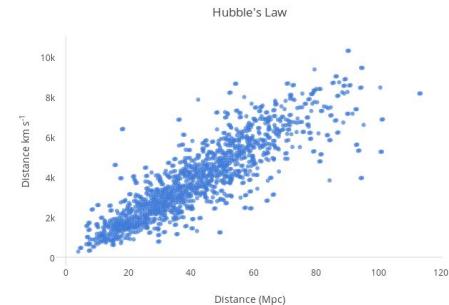
Données discrètes

- peuvent être représentées par un nombre entier
- généralement obtenues par comptage



Données continues

- peuvent être représentées par un nombre réel
- généralement obtenues par la mesure



Intervalle

Données numériques sans référence absolue

Les intervalles sont constants et la différence a du sens

Opérations : $=$, \neq , $<$, $>$, $-$

Exemple : degrés Celsius/Farenheit, dates, heures, QI, etc...

Ratio

Données numériques avec une référence absolue

Opérations : =, ≠, +, -, ×, /

Exemple : degrés Kelvin, taille, poids, vitesse, temps de réaction, etc...

Synthèse

	Catégorielle	Ordinal	Interval	Ratio
Ordonnancement		✓	✓	✓
Fréquence de distribution	✓	✓	✓	✓
Mode	✓	✓	✓	✓
Médiane		✓	✓	✓
Moyenne			✓	✓
+ , -			✓	✓
x , /				✓
Référence absolue				✓

Example : 100 m Londres 2012

C : catégorielle
O : ordinaire
I : Intervalle
R : ratio

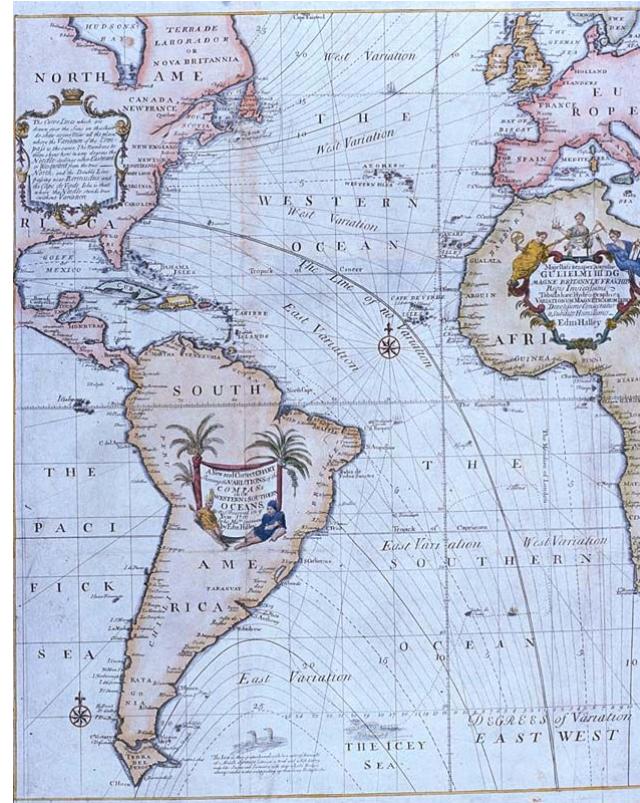
Rang	Ligne	Nom	Pays	Réaction	Temps
1	7	Usain Bolt	Jamaica	0.165	9.63
2	5	Yohan Blake	Jamaica	0.179	9.75
3	6	Justin Gatlin	United States	0.178	9.79
4	8	Ryan Bailey	United States	0.176	9.88
5	9	Churandy Martina	Netherlands	0.139	9.94
6	2	Richard Thompson	Trinidad and Tobago	0.160	9.98
7	3	Asafa Powell	Jamaica	0.155	11.99
?	?	?	?	?	?

Historique

Champ magnétique terrestre

Halley 1701

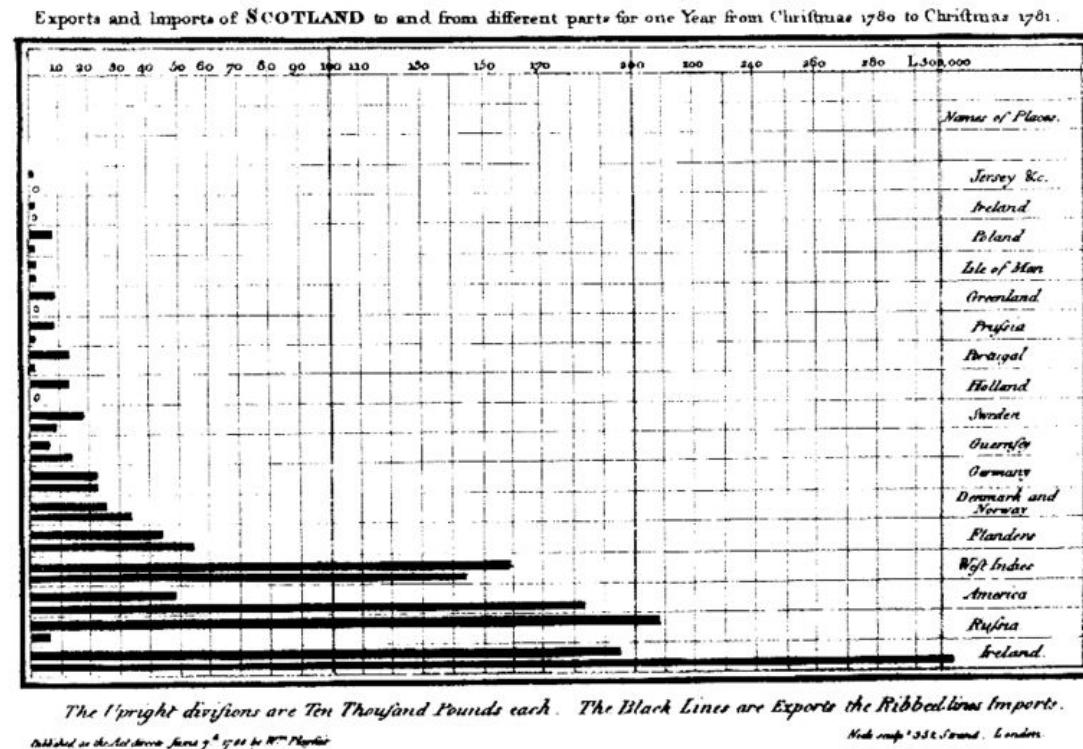
Première carte isogone (relie des points de même valeur)



Exportations de l'Ecosse

Playfair 1786

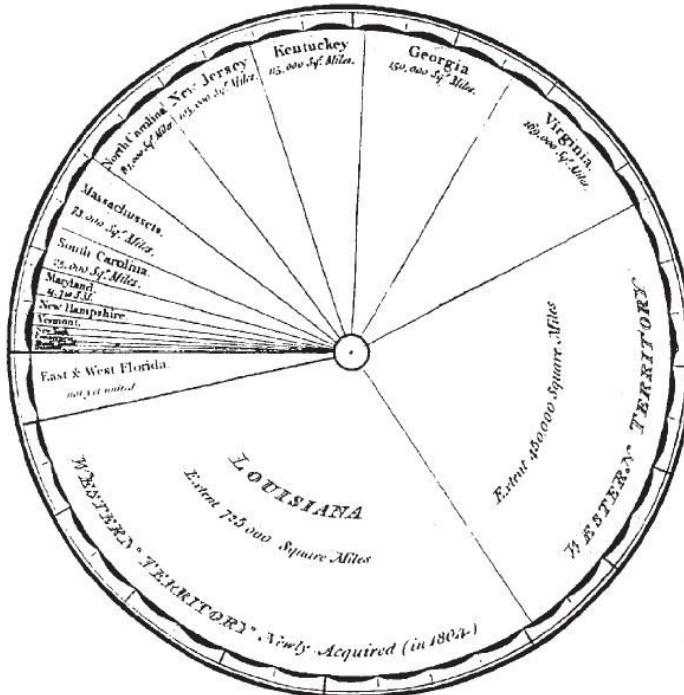
Premier bar graph



Commerce aux USA

Playfair 1805

Première “pie chart”



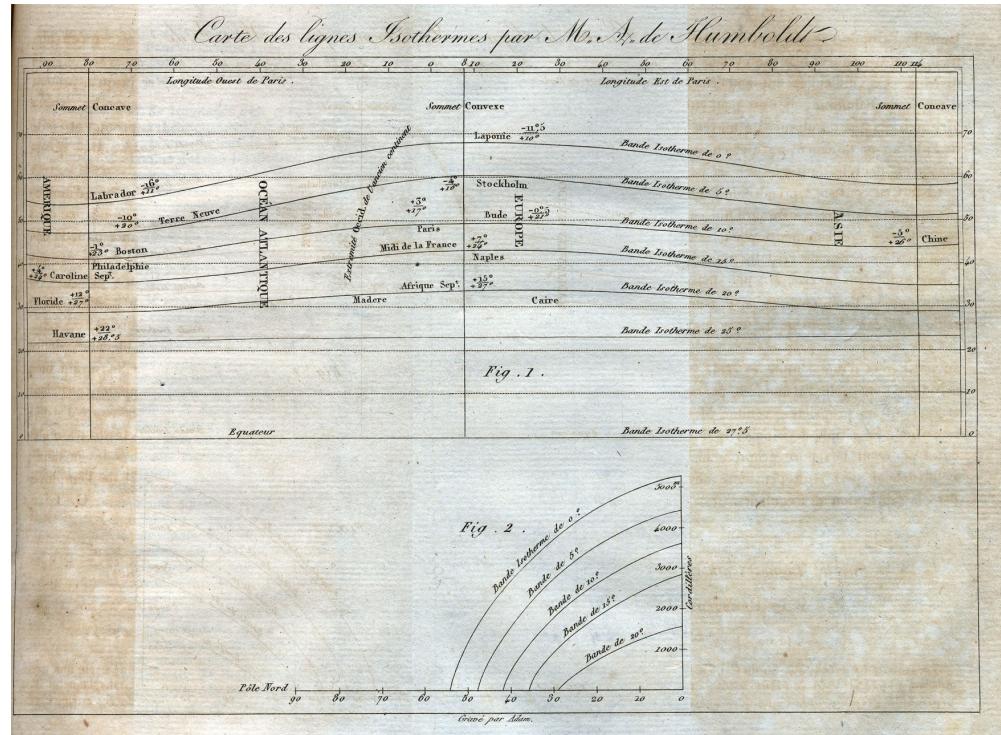
STATISTICAL REPRESENTATION of the UNITED STATES of AMERICA,
in 1805.

This Novel Invented Method is intended to show the Proportions between the dimensions in a striking manner:
Total Extent 450,000 Square Miles or 112,5 Millions of acres.

Isothermes

Humboldt 1817

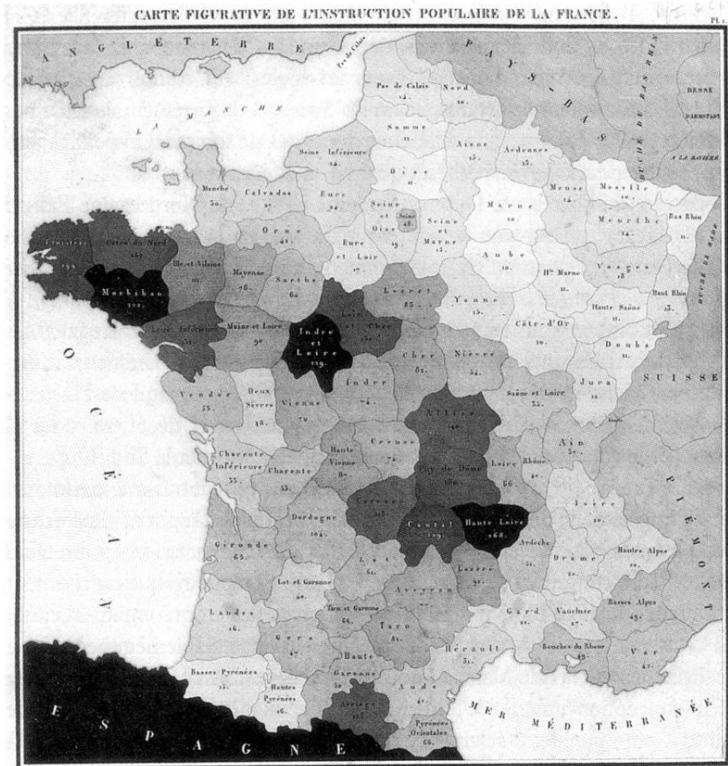
Premier graphe isotherme



L'instruction en France

Dupin 1826

Première carte thématique choroplète



Contours de température

Lalanne 1843

Représentation de 3 variables numériques ($y =$ heure, $x =$ mois, $z =$ température)

Heures

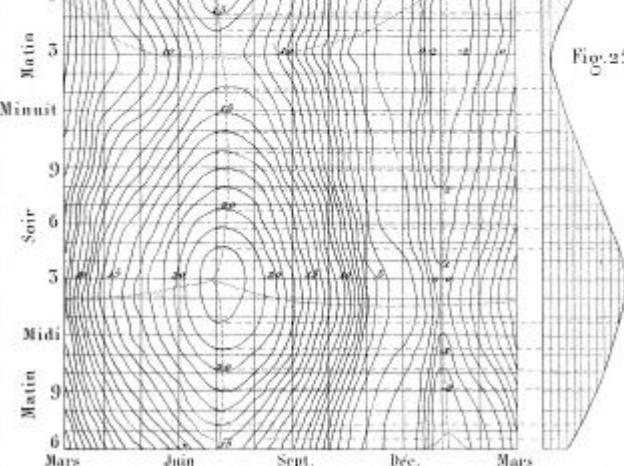


Fig. 2^{me}: Plan topographique exprimant les variations de la température à Metz.

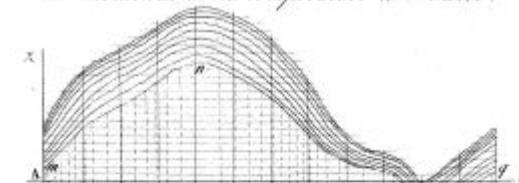
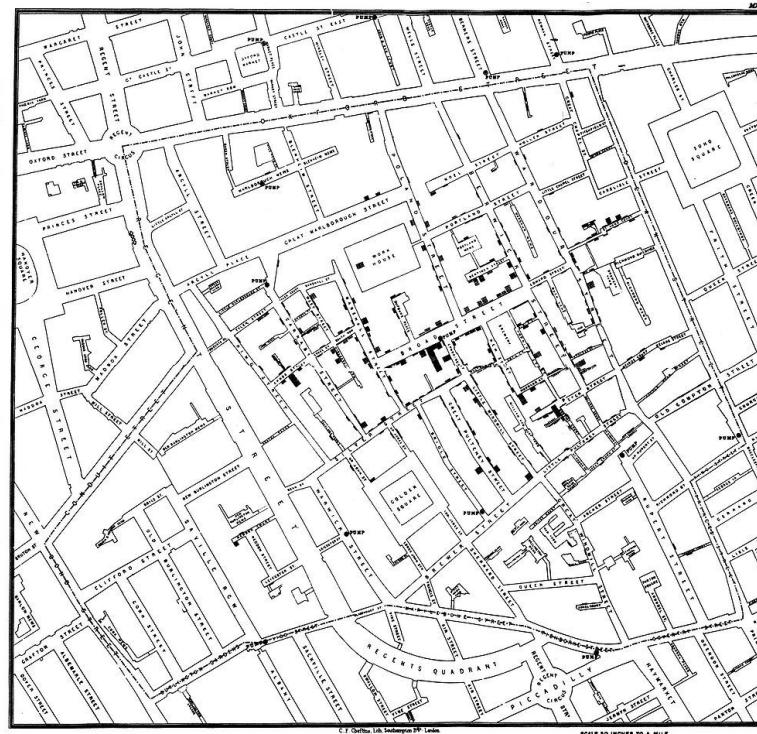


Fig. 2.

Fig. 2^{me}

Epidémie de Choléra à Londres

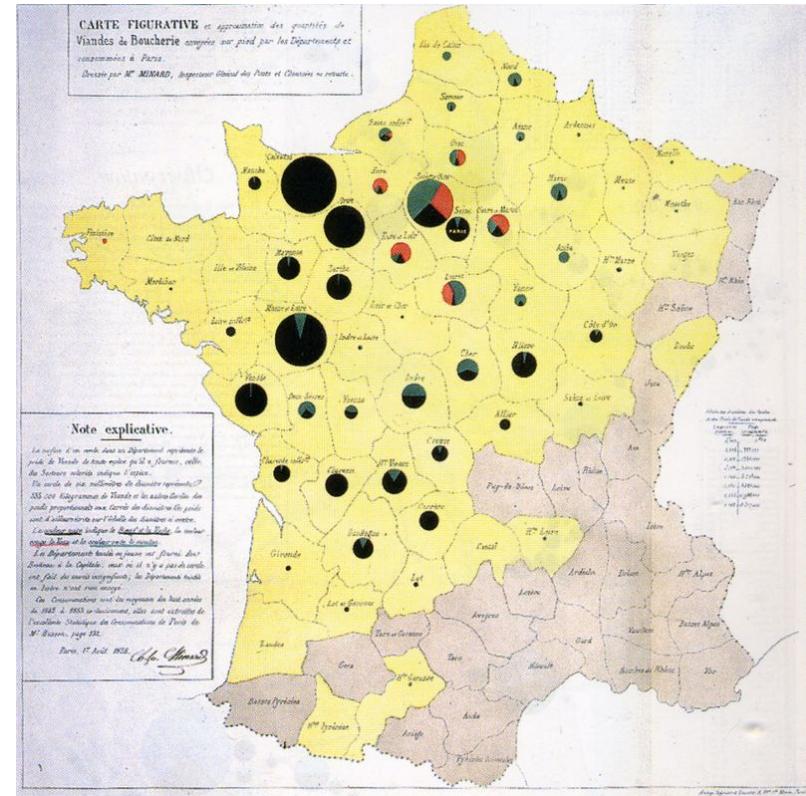
John Snow (1855)



Quantité de viande consommée à Paris

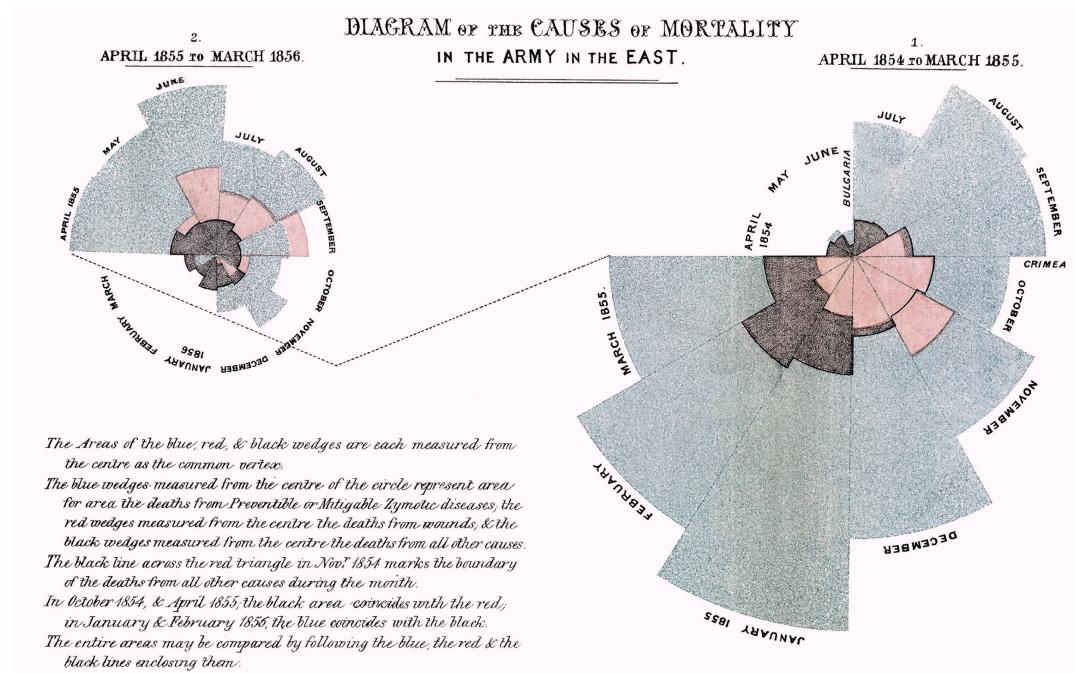
Minard 1858

Première carte incorporant des diagrammes



Mortalité dans l'armée anglaise

Nightingale 1858



Données météorologiques

Galton 1863

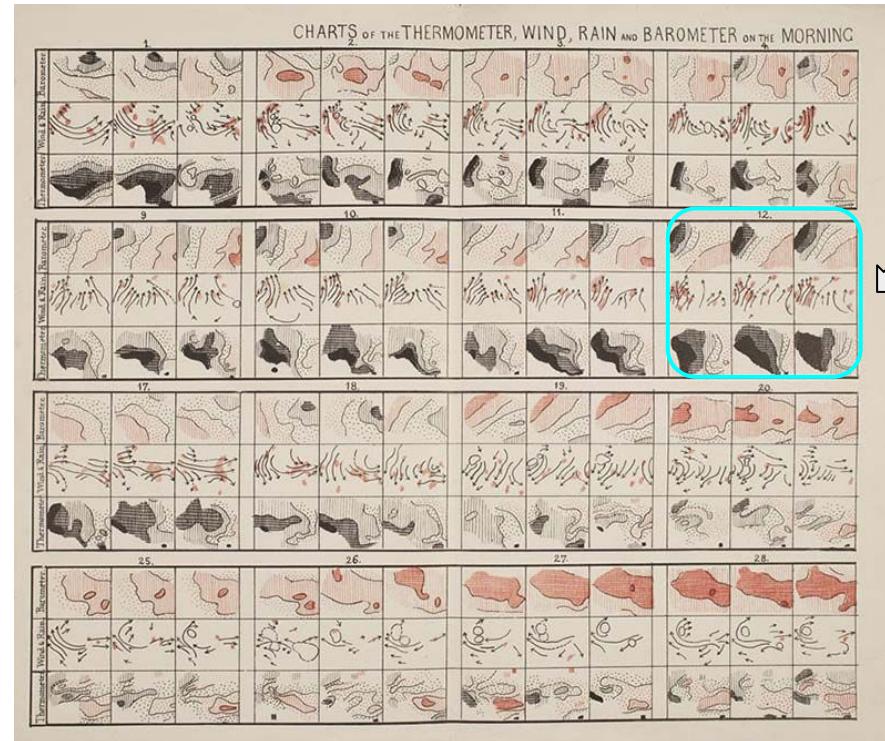
Pression atmosphérique
Vent et pluie
Température



Matin - Midi - Soir



Jour



La retraite de Russie (Minard 1869)

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812 ~ 1813.

Dessinée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie; le noir ceux qui en sortent. — Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. Chiers, de Léger, de Fezensac, de Chambray et le journal médical de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout, qui avaient été défaits sur Minsk et Malibolz et qui rejoignent vers Ossaka en Witelk, avaient toujours marché avec l'armée.

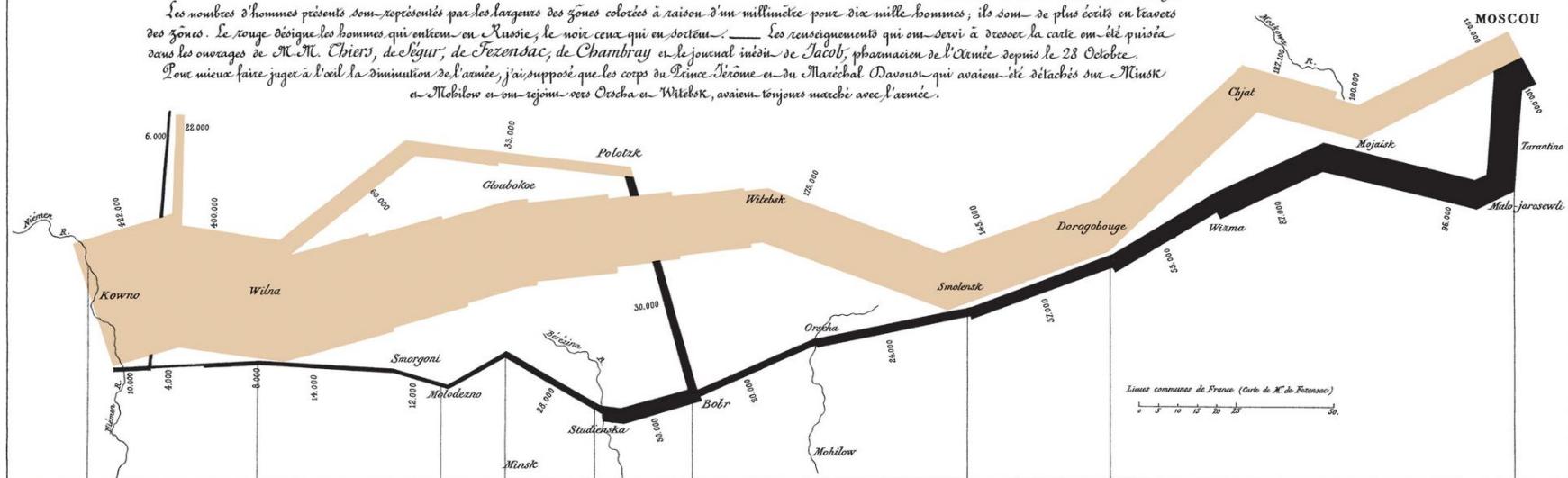
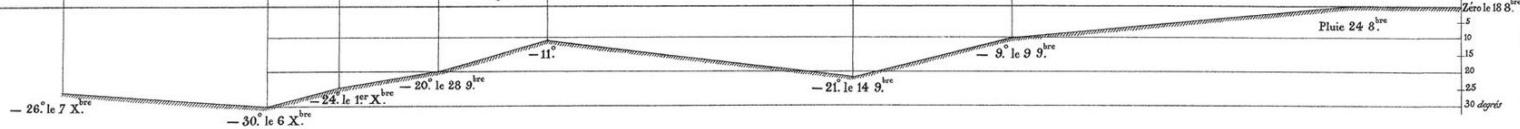


TABLEAU CRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop
le Niemen gelé.



Population de Paris en 1874

Vauthier 1874

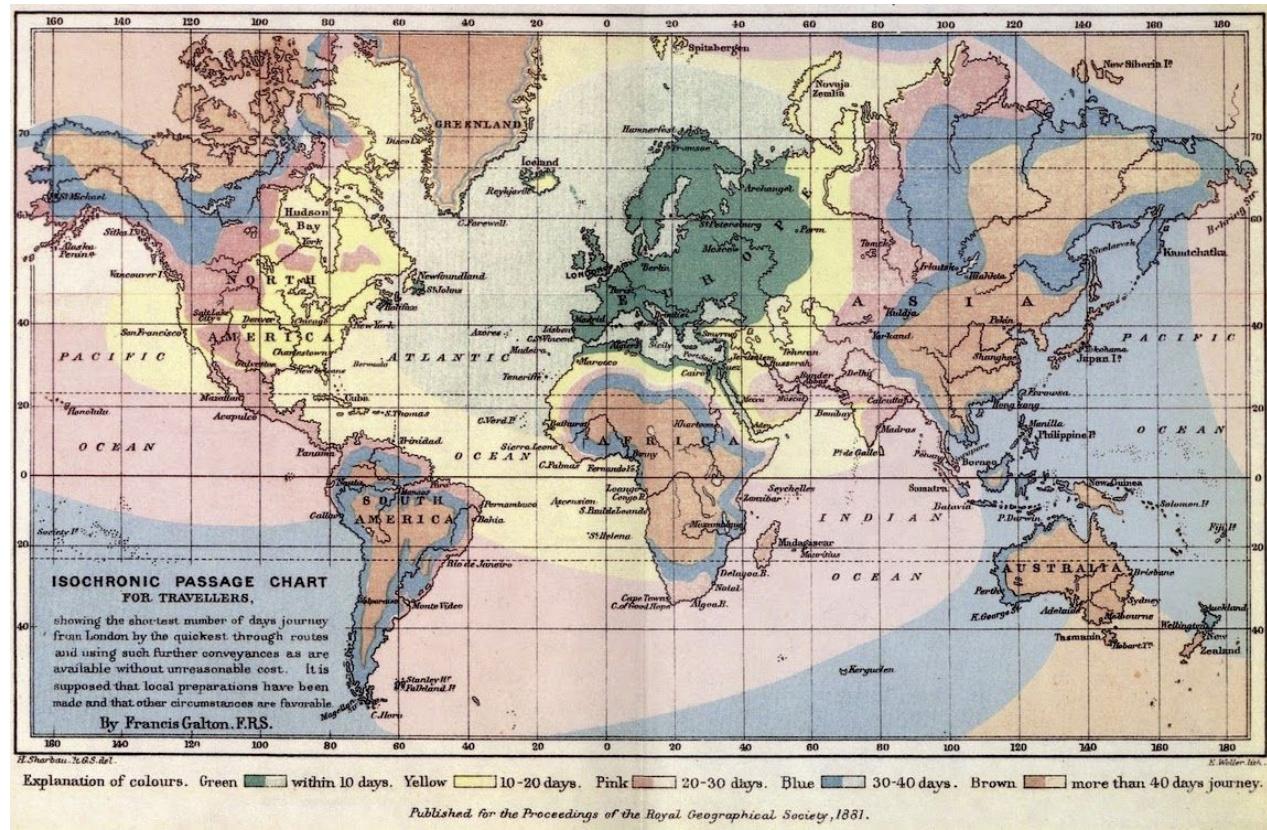
Première carte avec un tracé de données non géographiques



Durée de voyage depuis Londres

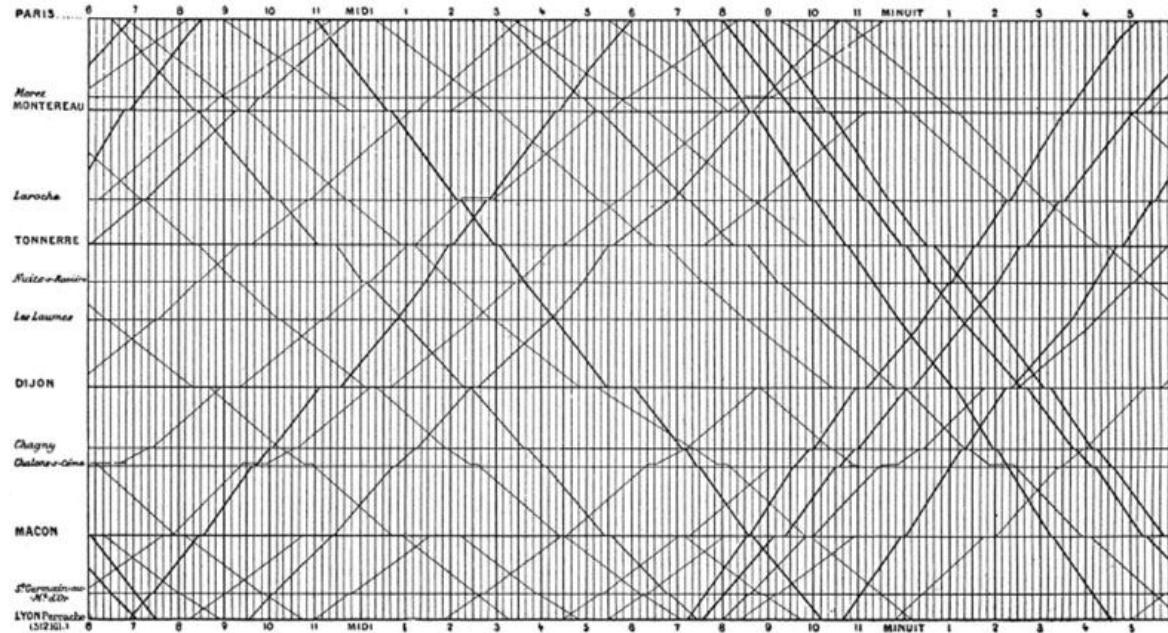
Galton 1881

Première carte isochrone



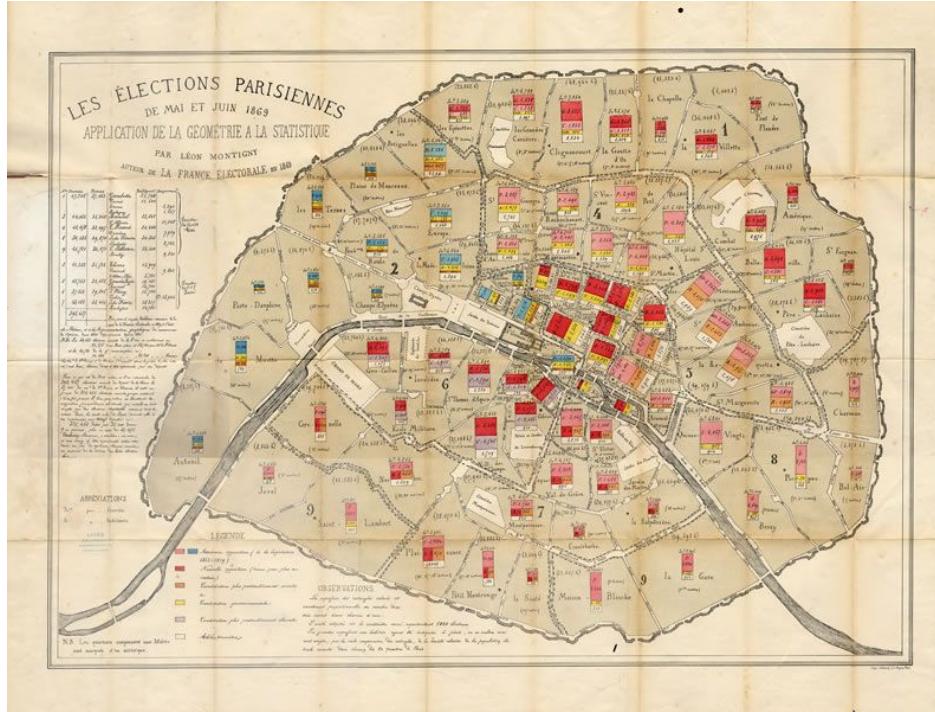
Marche des trains

Ibry (1885)



Plus d'histoire...

<http://www.ohmybox.info/dataviz/>



Éléments de perception visuelle

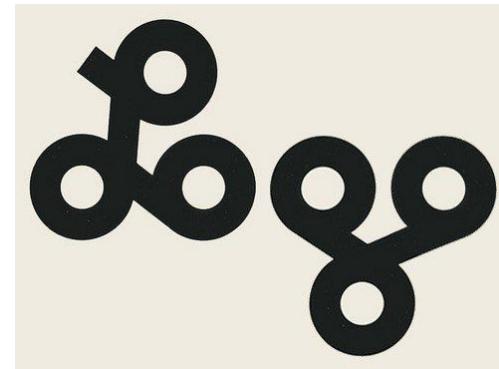
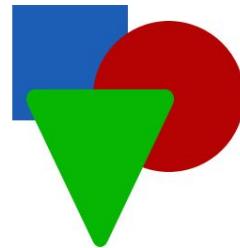
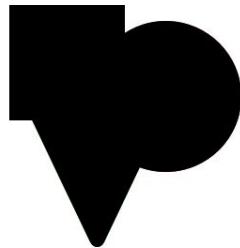
Introduction

La perception globale d'une forme précède les détails

“La psychologie de la forme ou gestaltisme (de l'allemand, Gestaltpsychologie) est une théorie psychologique, philosophique et biologique, selon laquelle les processus de la perception et de la représentation mentale traitent spontanément les phénomènes comme des ensembles structurés (les formes) et non comme une simple addition ou juxtaposition d'éléments” [[Wikipedia](#)]

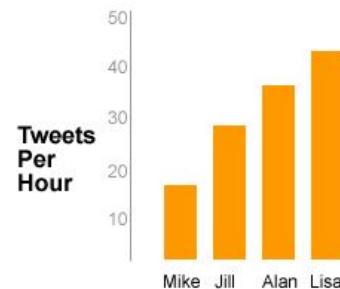
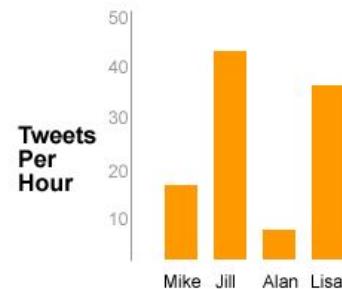
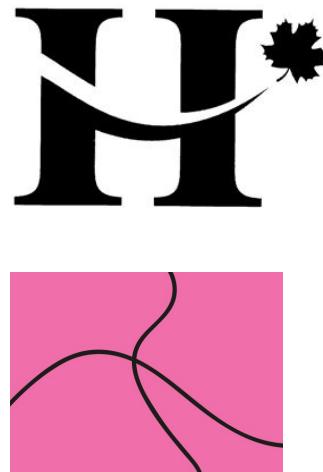
La loi de la bonne forme

Les formes complexes sont perçues comme une juxtaposition de formes simples...



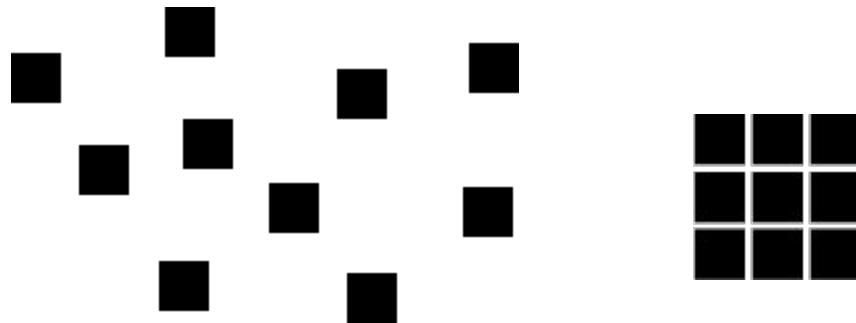
La loi de continuité

Les éléments disposés le long d'une courbe apparaissent liés entre eux. Sur un bar graph, l'information est plus facilement accessible...



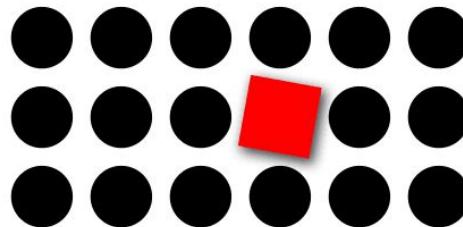
La loi de la proximité

Des choses proches sont perçues comme étant plus liées que si elles étaient distantes



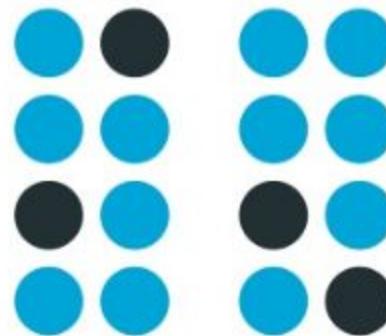
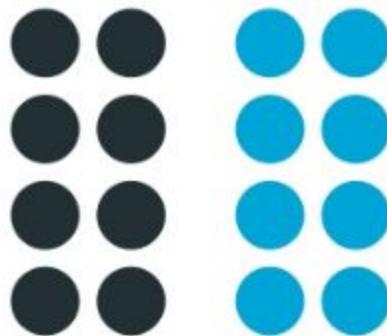
La loi de similitude / anomalie

Des choses similaires (forme, couleur, taille) sont perçues comme étant plus liées que si elles étaient différentes



Proximité et similarité

La proximité prédomine...



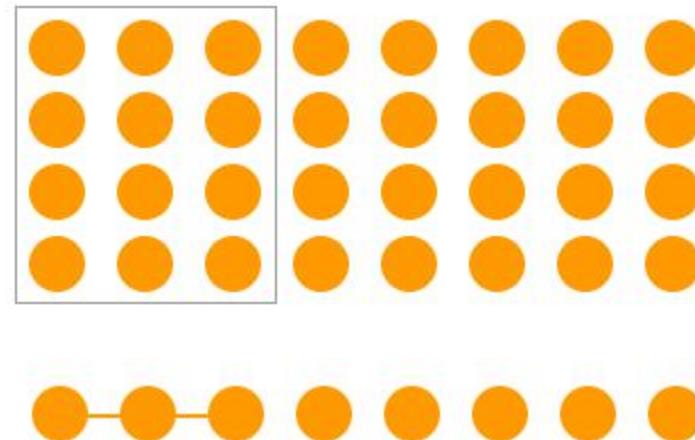
Loi de la symétrie

Des choses symétriques sont perçues comme étant liées. La symétrie prédomine sur la proximité



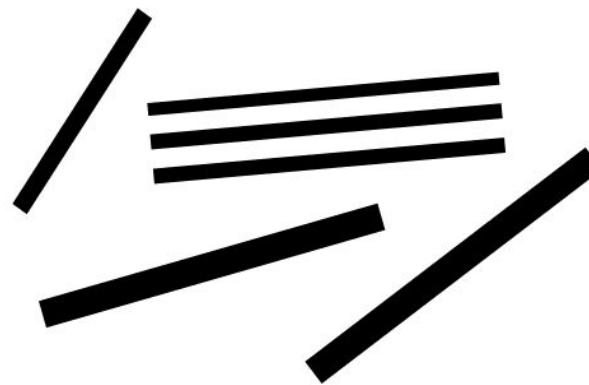
Loi de la connectivité

Des éléments connectés sont perçus comme étant liés...



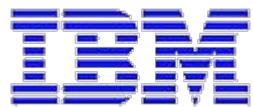
La loi de destin commun

Des choses partageant un mouvement commun ou une orientation commune sont perçues comme étant plus liées...

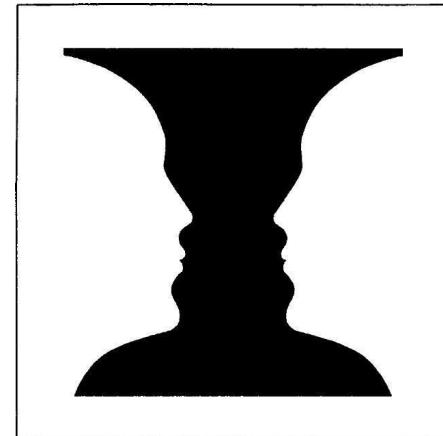


La loi de la clôture

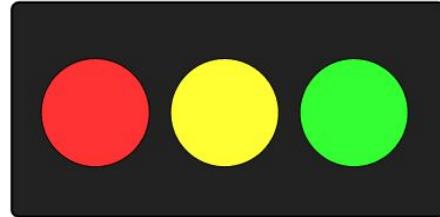
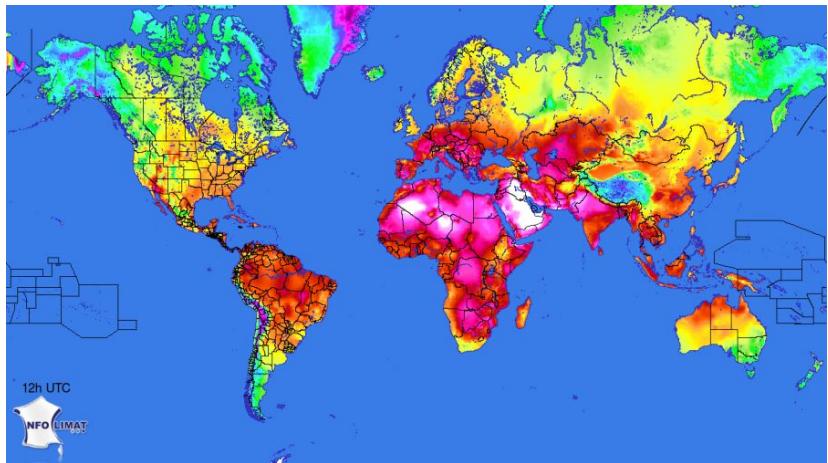
Un arrangement complexe de formes individuelles tendent à être perçues comme une seule forme simple à identifier...



La distinction figure - fond



L'expérience du passé

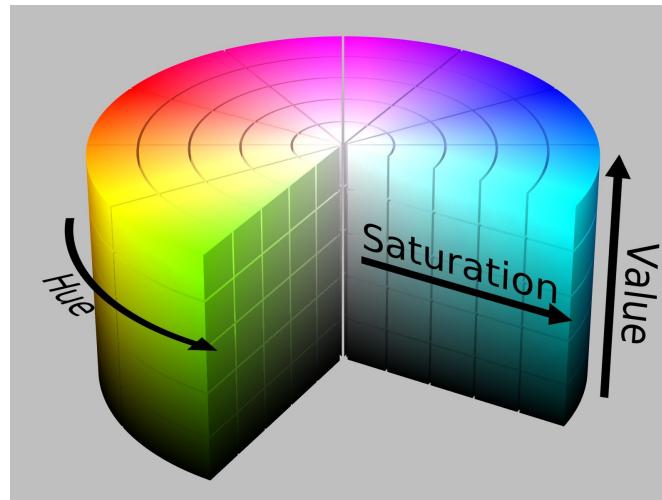


Les couleurs

Codage des couleurs

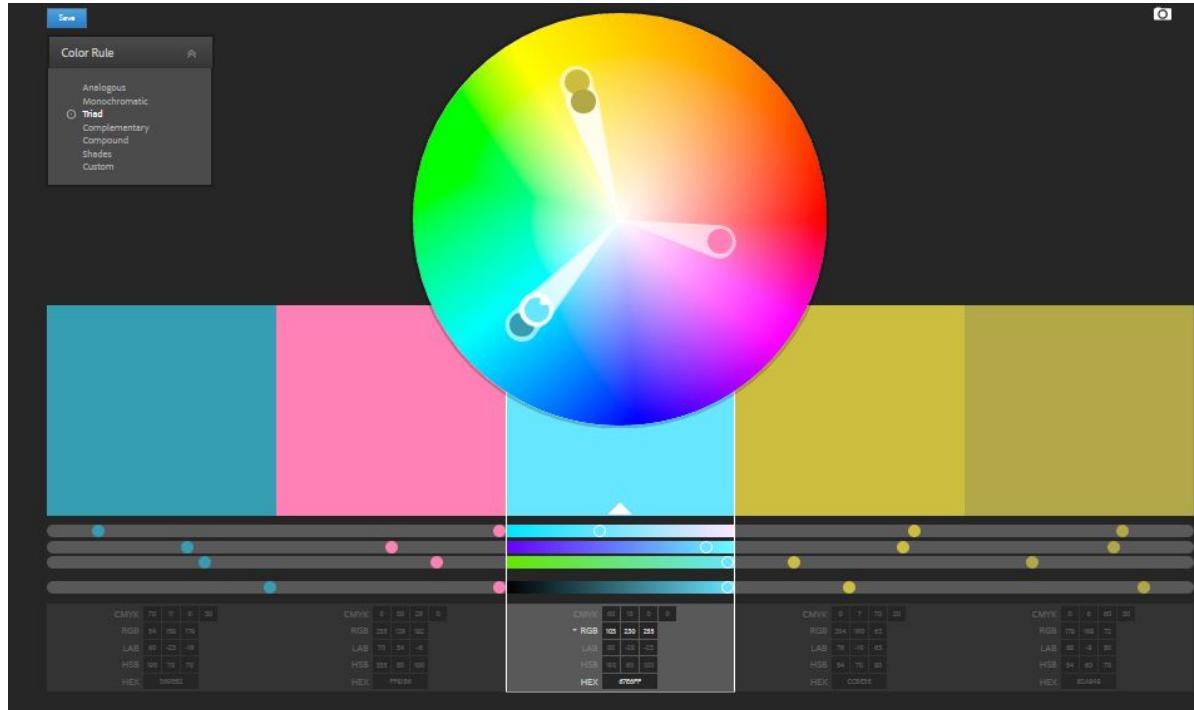
RGB (red, green, blue) : codage informatique, proche du matériel

HSV (hue, saturation, value) : plus proche de la perception humaine



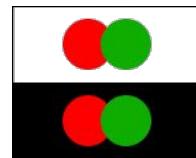
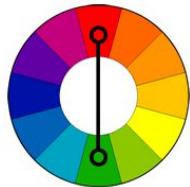
Choix d'une palette de couleur

<https://color.adobe.com/create/color-wheel>

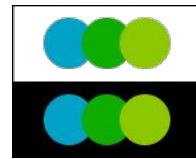
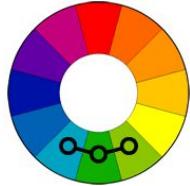


Combinaisons harmonieuses

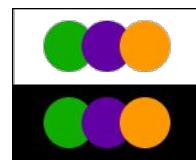
Complementary



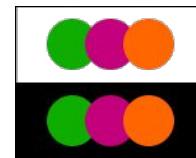
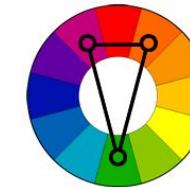
Analogous



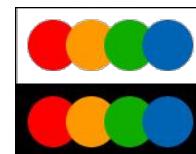
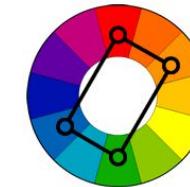
Triadic



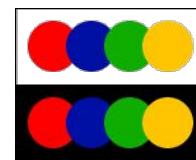
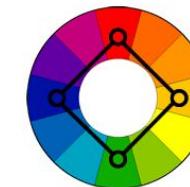
Split complementary



Tetradic



Square



Les 10 commandements...

THE 10 COMMANDMENTS OF COLOR THEORY

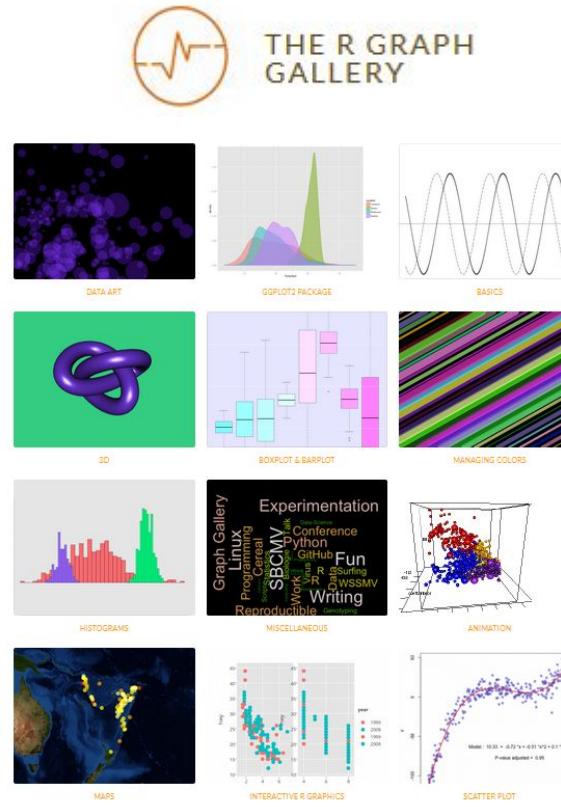
<p>1 KNOW THE COLOR WHEEL WELL! DO YOU KNOW WHAT EACH COLOR SIGNIFIES?</p>  <p>RED LIVE, ENERGY, INTENSITY.</p>  <p>YELLOW INFLUENCE, ATTENTION.</p>  <p>GREEN FRESHNESS, SAFETY, GROWTH.</p>  <p>BLUE STABILITY, TRUST, SERENITY.</p>  <p>PURPLE ROYALTY, WEALTH, FEMININITY.</p> 	<p>2 MATCH IT. DO NOT OVERLOOK THE AUSTERITY OF ANALOG COLORS!</p>  <p>Barber</p>  <p>wheelchair</p>  <p>CupCake</p>  <p>Sandals</p>  <p>Television</p> 	<p>3 CAN'T MATCH IT? CLASH IT WITH COMPLEMENTARY COLORS!</p>  <p>Barber</p>  <p>wheelchair</p>  <p>CupCake</p>  <p>Sandals</p>  <p>Television</p> 	<p>4 IS CONTRAST TOO INTENSE? THEN, SPLIT IT!</p>  <p>Barber</p>  <p>wheelchair</p>  <p>CupCake</p>  <p>Sandals</p>  <p>Television</p> 	<p>5 NEED MORE VARIATIONS? GO DOUBLE COMPLEMENTARY!</p>  <p>Barber</p>  <p>wheelchair</p>  <p>CupCake</p>  <p>Sandals</p>  <p>Television</p> 	<p>6 GO TRIAD WITH 3 DIFFERENT HUES... CHOOSE FROM A GREATER VARIETY!</p>  <p>Barber</p>  <p>wheelchair</p>  <p>CupCake</p>  <p>Sandals</p>  <p>Television</p> 	<p>7 SOMETIMES, MONOCHROME IS THE WAY TO GO...</p>  <p>Barber</p>  <p>wheelchair</p>  <p>CupCake</p>  <p>Sandals</p>  <p>Television</p> 	<p>8 OTHER TIMES, AN ACHROMATIC SCHEME SERVES BEST!</p>  <p>Barber</p>  <p>wheelchair</p>  <p>CupCake</p>  <p>Sandals</p>  <p>Television</p> 	<p>9 KNOW YOUR HUES, TINTS, SHADES AND TONES... WHAT WORKS WHERE?</p>  <p>Barber</p>  <p>wheelchair</p>  <p>CupCake</p>  <p>Sandals</p>  <p>Television</p> 	<p>10 AND LASTLY, RGB, CMYK AND PANTONE ARE NOT THE SAME!</p>  <p>PANTONE</p>  <p>CupCake</p>  <p>CMYK</p>  <p>RGB</p>  <p>CupCake</p> 
---	--	--	---	---	---	--	---	---	---

Principaux types de graphiques

Graph galleries



The Data Visualisation Catalogue



Histogrammes et densité

Représentation de données numériques

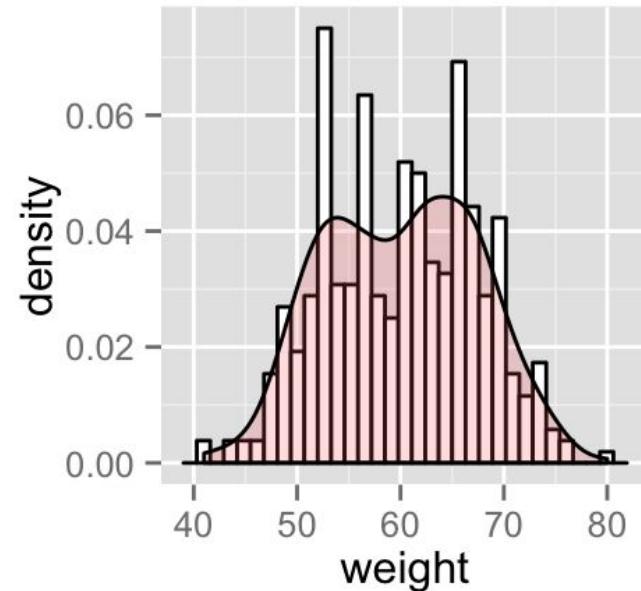
Permet d'estimer la distribution d'une variable continue

Définir la largeur des classes

Répartir les valeurs numériques dans chacune des classes

Compter le nombre de valeurs dans chaque classe

Normaliser pour superposer une densité (diviser par la surface)



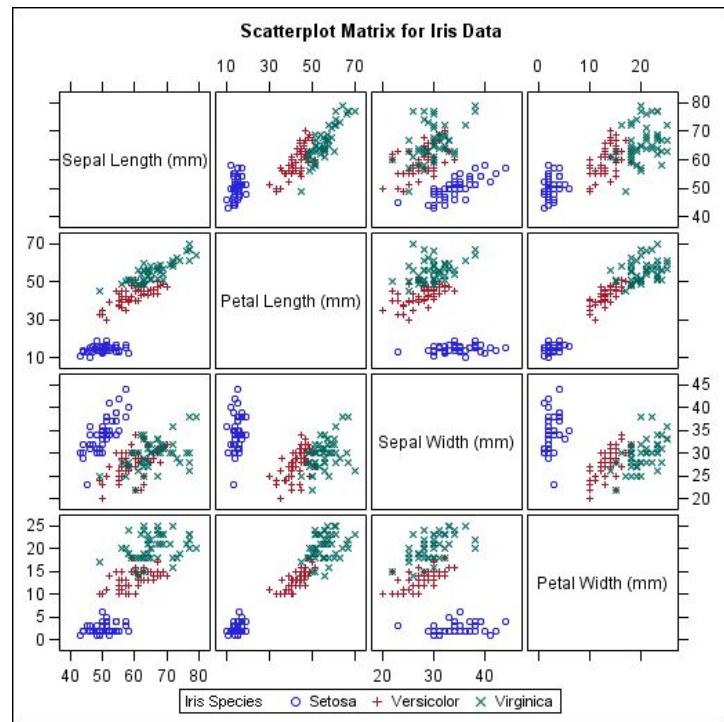
<https://tinlizzie.org/histograms>

Scatter plot matrix

Pour un ensemble de données $\{X_1, \dots, X_n\}$, affiche sous forme matricielle l'ensemble des scatter plot X_i vs X_j

X_i vs X_j se trouve en (ligne i ; colonne j)

La matrice possède n lignes et n colonnes



Coordonnées parallèles

permet de visualiser un grand nombre de données et d'extraire des corrélations

Chaque axe vertical est une dimension

Parallel coordinate plot, Fisher's Iris data

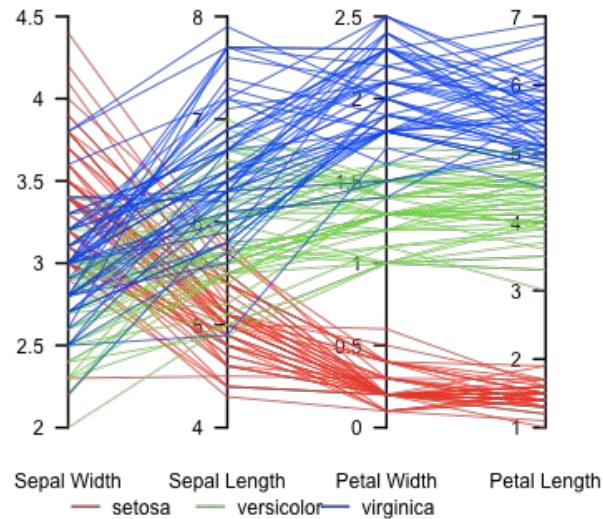


Table de contingence

Représentation des données numériques relatives à 2 ou plusieurs variables catégorielles

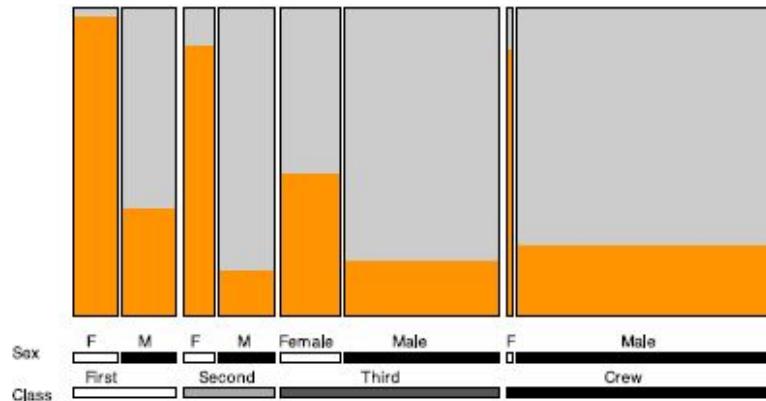
Class:	First		Second		Third		Crew	
Gender:	Female	Male	Female	Male	Female	Male	Female	Male
Survived:								
yes	141	62	93	25	90	88	20	192
no	4	118	13	154	106	422	3	670
Survival Rate (in %)	97	34	88	14	46	17	87	22

Titanic data - Source : Handbook of Data Visualization [Chen, Härdle, Unwin]

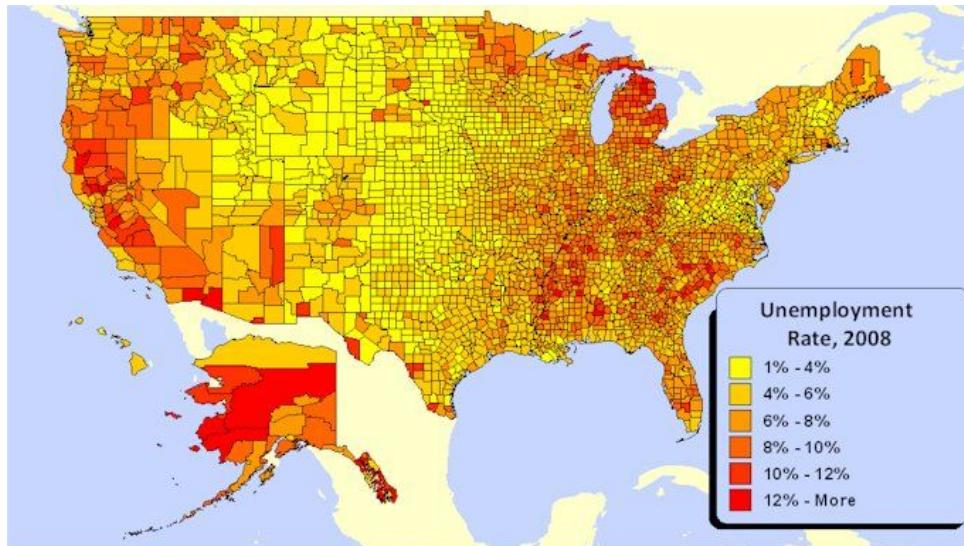
Mosaic plot

Utilisé pour la représentation des tables de contingence

Class:	First		Second		Third		Crew	
Gender:	Female	Male	Female	Male	Female	Male	Female	Male
Survived:								
yes	141	62	93	25	90	88	20	192
no	4	118	13	154	106	422	3	670
Survival Rate (in %)	97	34	88	14	46	17	87	22



Carte choroplète

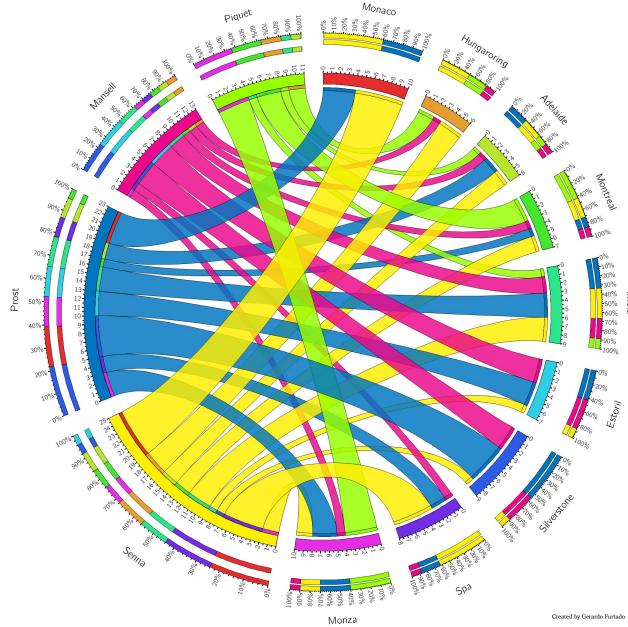


Chord diagram

4 CHAMPIONS, 9 TRACKS

4 champions from the 80's and early 90's (Senna, Piquet, Prost and Mansell) and 9 tracks: 2 low-speed (Monaco, Hungaroring and Adelaide), 3 medium-speed (Montreal, Imola and Estoril) and 3 high-speed tracks (Silverstone, Spa and Monza). How did they fare in each kind of track? The diagram shows the number of wins of each champion in each circuit. Yellow ribbons for Senna, Blue ribbons for Prost, red ribbons for Mansell and green ribbons for Piquet.

The inner values are absolute number of wins (per driver and per circuit), and the outer values are percentages.



Created by Gerardo Furtado

Choix du type de graphique (1/3)

See relationships among data points

-  Scatterplot
-  Matrix Chart
-  Network Diagram

Track rises and falls over time

-  Line Graph
-  Stack Graph
-  Stack Graph for Categories

Analyze a text

-  Word Tree
-  Tag Cloud
-  Phrase Net
-  WC G Word Cloud Generator

Compare a set of values

-  Bar Chart
-  Block Histogram
-  Bubble Chart

See the parts of a whole

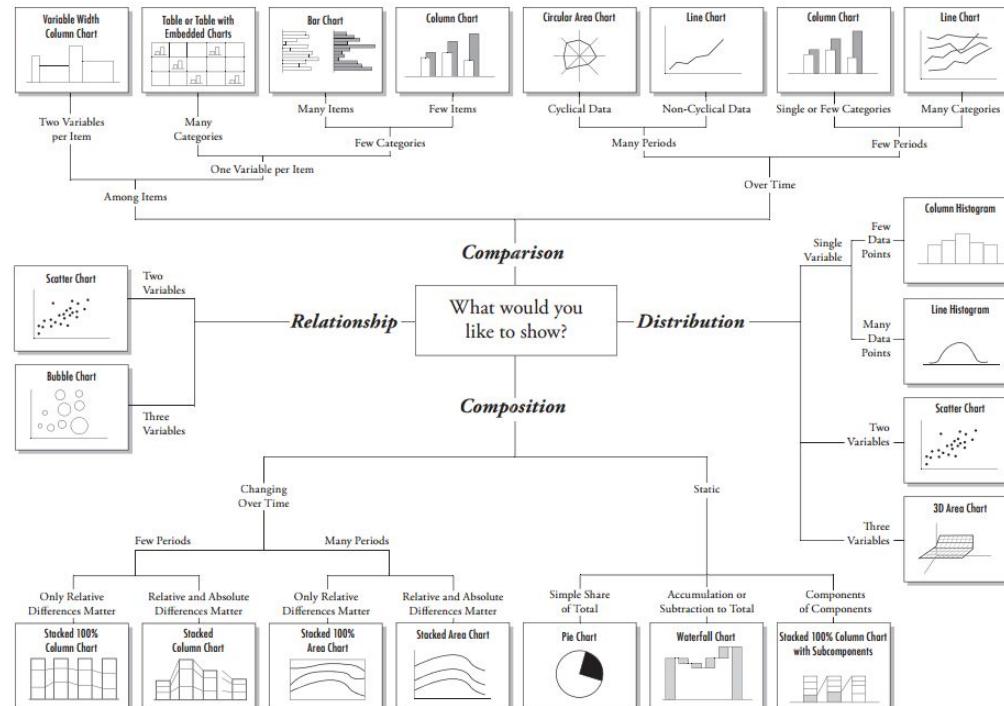
-  Pie Chart
-  Treemap
-  Treemap for Comparisons

See the world

-  Ottawa Neighbourhood Map
-  US County Map
-  World Map
-  Massachusetts Map
-  New Jersey Map

Choix du type de graphique (2/3)

Chart Suggestions—A Thought-Starter



Choix du type de graphique (3/3)

A PERIODIC TABLE OF VISUALIZATION METHODS

C continuum	D Data Visualization Visual representations of quantitative data in schematic form (either with or without axes)	S Strategy Visualization The systematic use of complementary visual representations in the analysis, development, formulation, communication, and implementation of strategies in organizations.	G graphic facilitation
Tb table	Ca cartesian coordinates	M Metaphor Visualization Visual Metaphors position information graphically to organize and structure information. They also convey an insight about the represented information through the key characteristics of the metaphor that is employed	Ct cartoon
Pi pie chart	L line chart	C Compound Visualization The complementary use of different graphic representation formats in one single schema or frame	E meeting trace
B bar chart	Ac area chart	F flight plan	Ri rich picture
Hi histogram	Sc scatterplot	Cs concept skeletons	Br bridge
Tk tally box plot	Sp spectrogram	Fu funnel	Kn knowledge map
Da data map	Tp treemap	Gc game chart	D dilemma diagram
Cn cone tree	Sy system dynamics simulation	Pm perspectives diagram	Pr parameter ruler
Ed edgeworth box	Pf portfolio diagram	Ti tension map	Kn learning map
Pc performance charting	St strategy map	Dt decision tree	Ic iceberg
St strategy map	Oc organization chart	Cp open critical path method	Lm learning map
Ho house of quality	Fd feedback diagram	Cf concept fan	
Ft failure tree	Mq magic quadrant	Co concept map	
Mq life-cycle diagram	Ld life-cycle diagram	Pe pert chart	
Po porter's five forces	Bm big matrix	Ev evocative knowledge map	
S i-Cycle	Stc strategy canvas	V Vee diagram	
Sm stakeholder map	Vc value chain	H hypercycle	
Is ishikawa diagram	Hy hypercycle	Sr stakeholder rating map	
Tc technology roadmap	Ta tape	Ta tape	
Sd spray diagram	Ed edge diagram	Ed edge diagram	

Process Visualization

Note: Depending on your location and connection speed it can take some time to load a pop-up picture.

© Ralph Lengler & Martin J. Eppler, www.visual-literacy.org

version 1.5

Structure Visualization

- Overview
- Detail
- Detail AND Overview
- < > Divergent thinking
- > < Convergent thinking

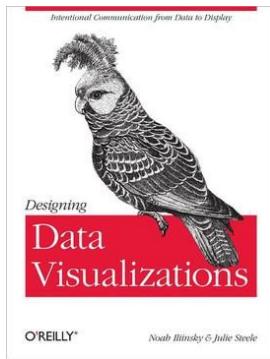
Su supply demand curve	Pc performance charting	St strategy map	Oc organization chart	Ho house of quality	Fd feedback diagram	Ft failure tree	Mq magic quadrant	Ld life-cycle diagram	Po porter's five forces	S i-Cycle	Sm stakeholder map	Is ishikawa diagram	Tc technology roadmap
Ed edgeworth box	Pf portfolio diagram	Sg strategic game board	Mz münchener's organiograph	Z zweig's morphological box	Ad affinity diagram	De decision discovery diagram	Bm big matrix	Stc strategy canvas	Vc value chain	Hy hypercycle	Sr stakeholder rating map	Ta tape	Sd spray diagram

Bonnes pratiques

Codes visuels

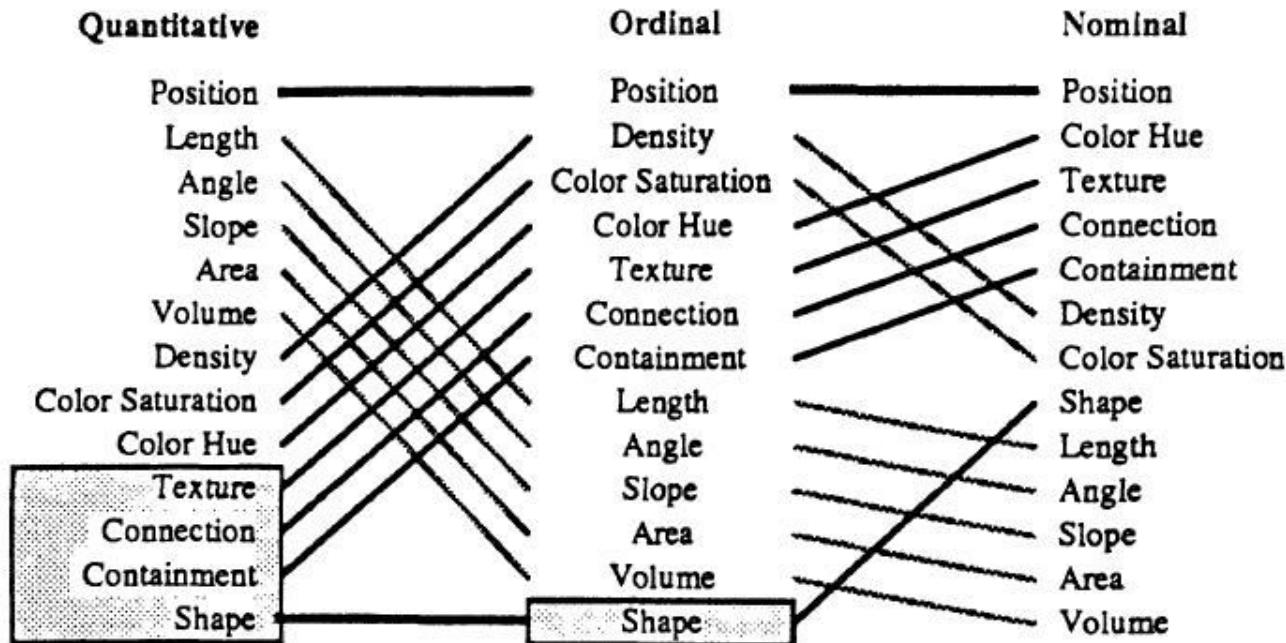
Designing Data Visualizations

Noah Iliinsky...



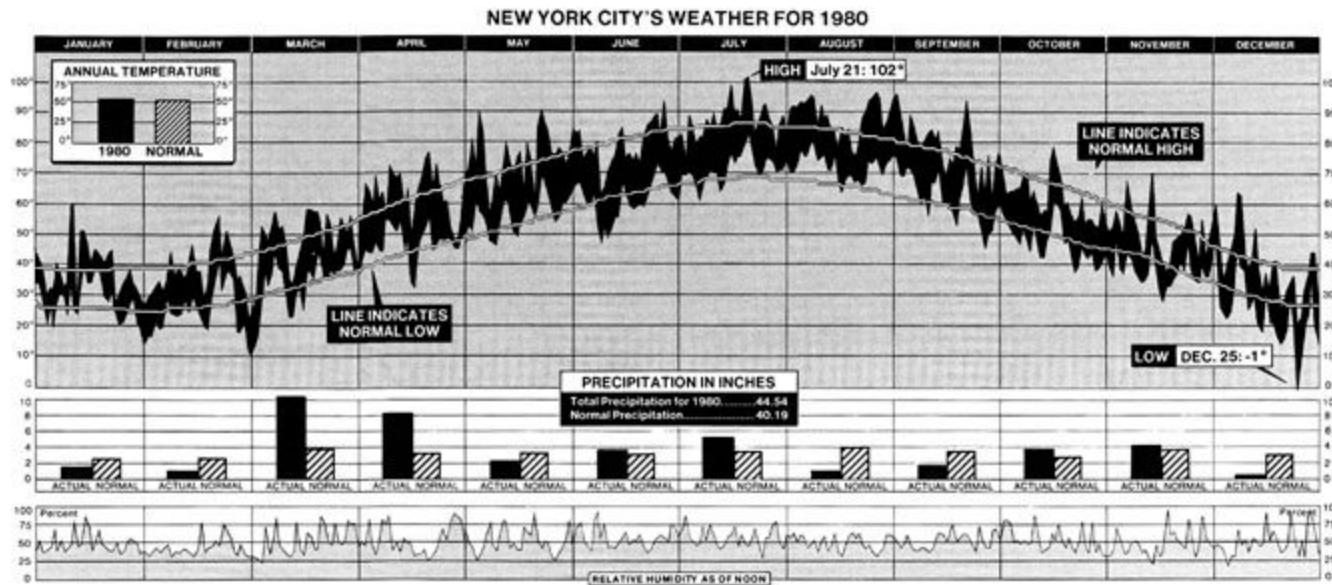
Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
position, placement	yes	infinite	Good	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite	Good	Good	Good	Good
length	yes	many	Good	Good			
size, area	yes	many	Good	Good			
angle	yes	medium	Good	Good			
pattern density	yes	few	Good	Good			
weight, boldness	yes	few		Good			
saturation, brightness	yes	few		Good			
color	no	few (<20)				Good	
shape, icon	no	medium				Good	
pattern texture	no	medium				Good	
enclosure, connection	no	infinite				Good	Good
line pattern	no	few					Good
line endings	no	few					Good
line weight	yes	few		Good			

Classement des perceptions visuelles



Météo à New York

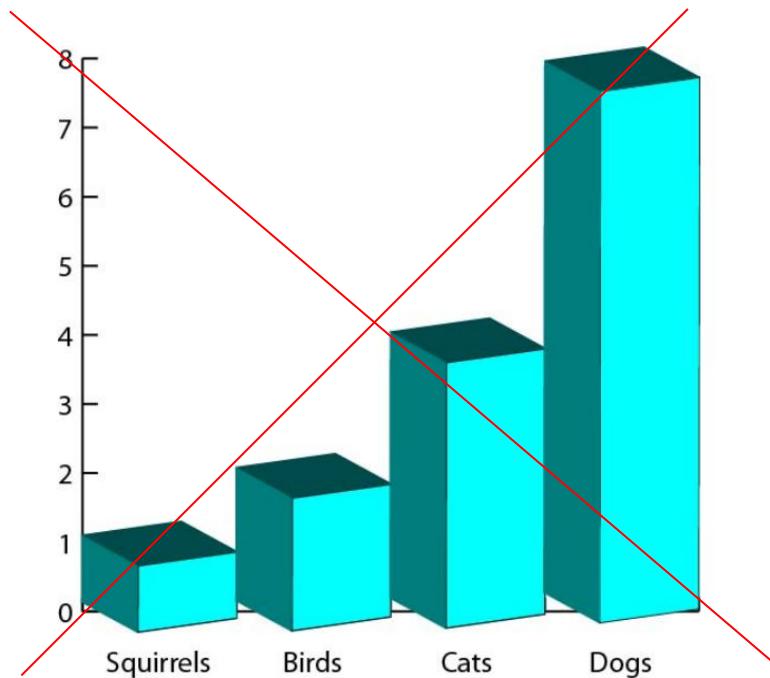
2200 données numériques représentées !



Exemples à éviter

Principes de Tufte

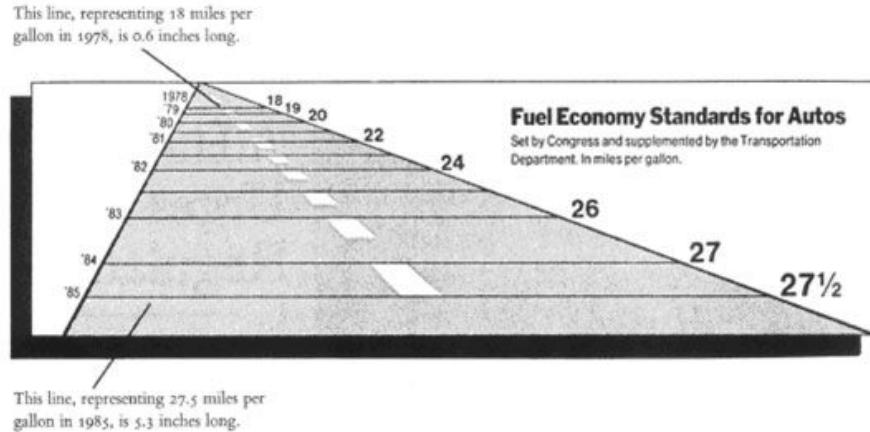
1. La représentation des grandeurs (ou des variations de ces grandeurs) doit être directement proportionnelle aux valeurs numériques
2. Éviter toute ambiguïté et toute distorsion graphique
3. La dimension du graphique ne doit pas être supérieure à la dimension des données
4. Maximiser le “data ink ratio”



Attention au “lie factor”

LF = surface graphique / valeur de la donnée

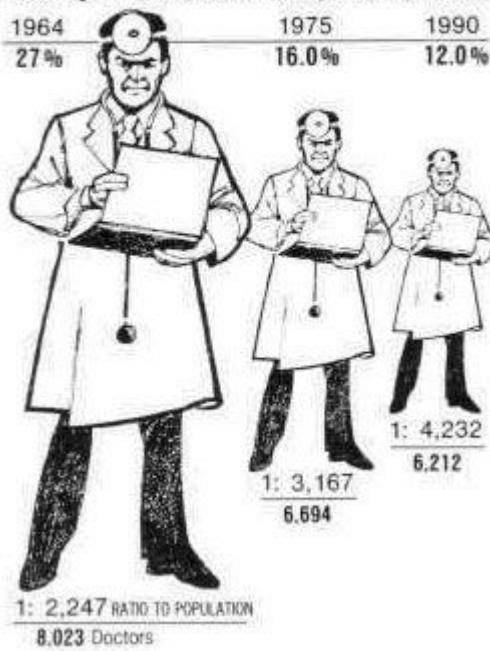
[Tufte 1991]



THE SHRINKING FAMILY DOCTOR In California

Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27%	16.0%	12.0%

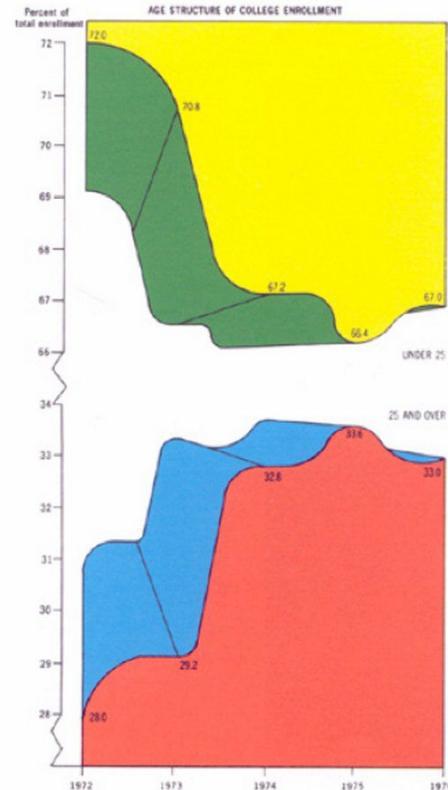


The worst chart ever...

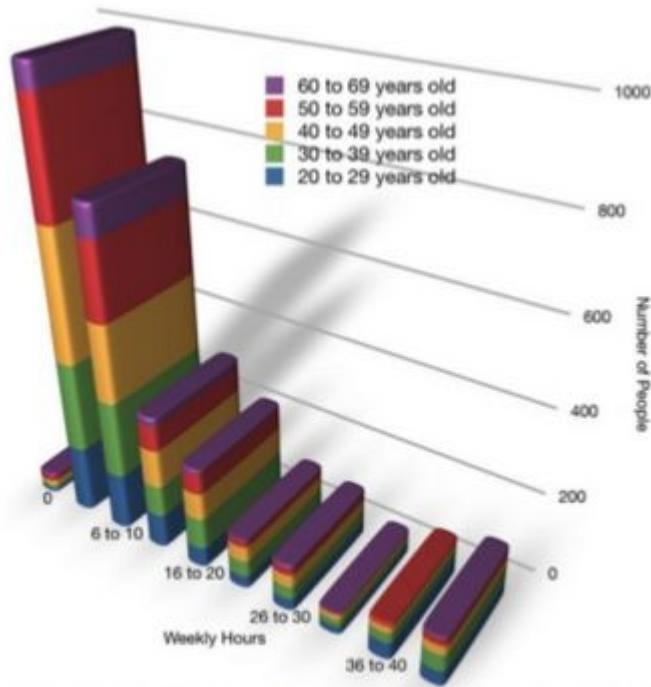
La couleur et la 3D ne servent à rien

L'axe des Y n'est pas continu

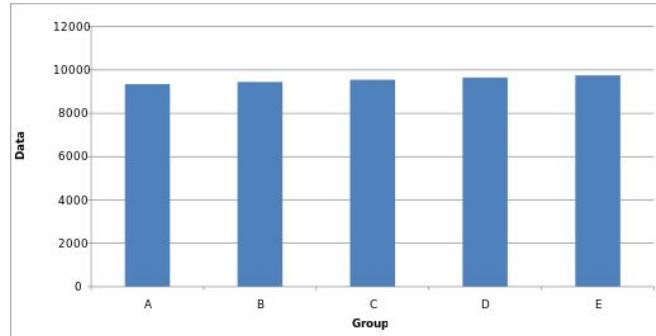
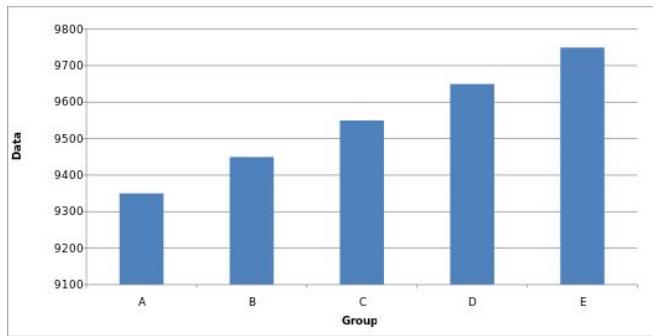
Seulement 5 chiffres à communiquer



Data ink ratio

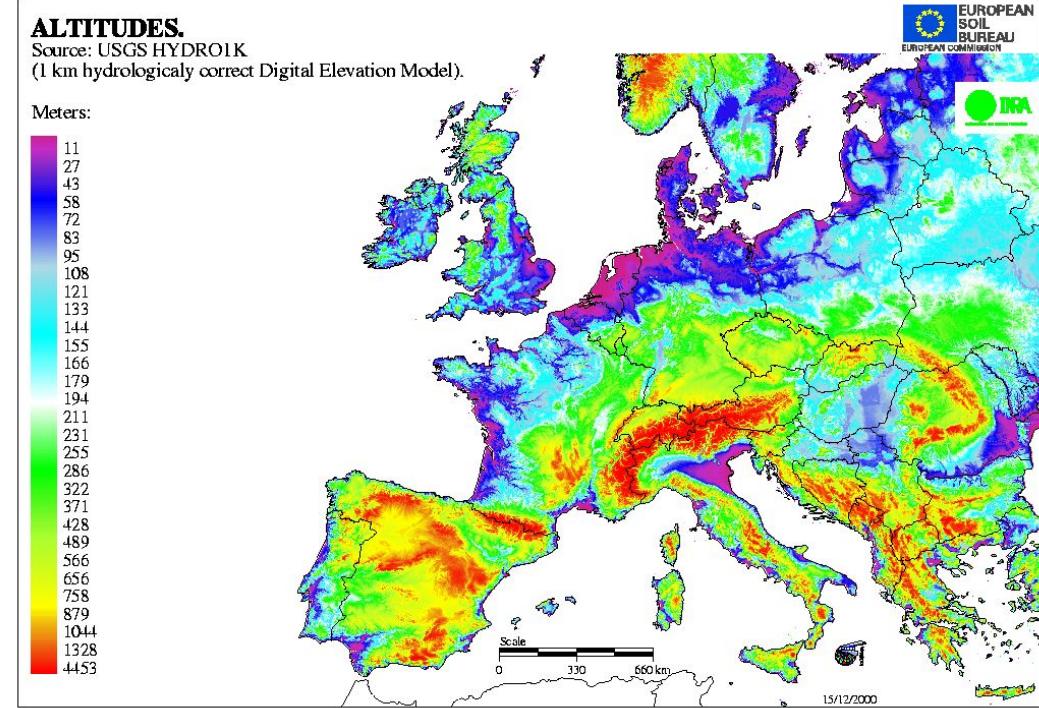


Graduation des axes



Mauvais choix de couleurs

Echelle des couleurs inversée. Les altitudes élevées devraient être en haut



D'autres exemples...

