

DATA VISUALISATION

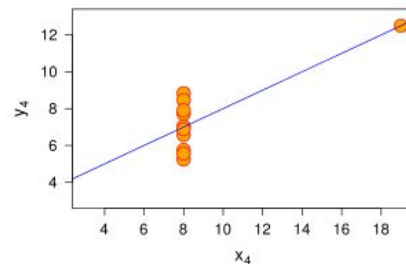
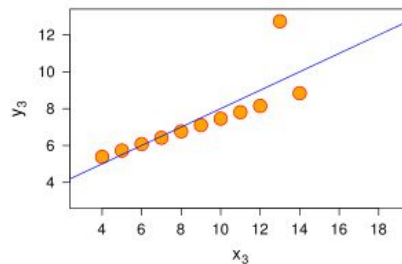
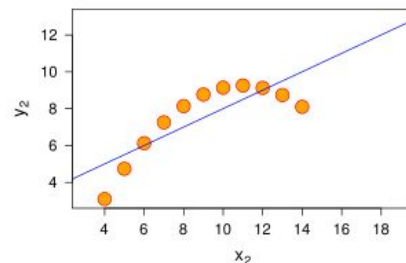
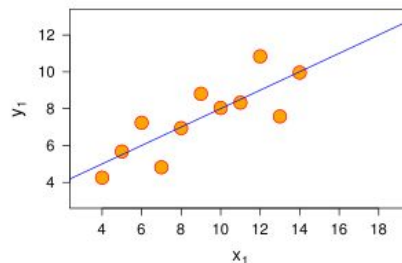
Quelques éléments de réflexion...

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



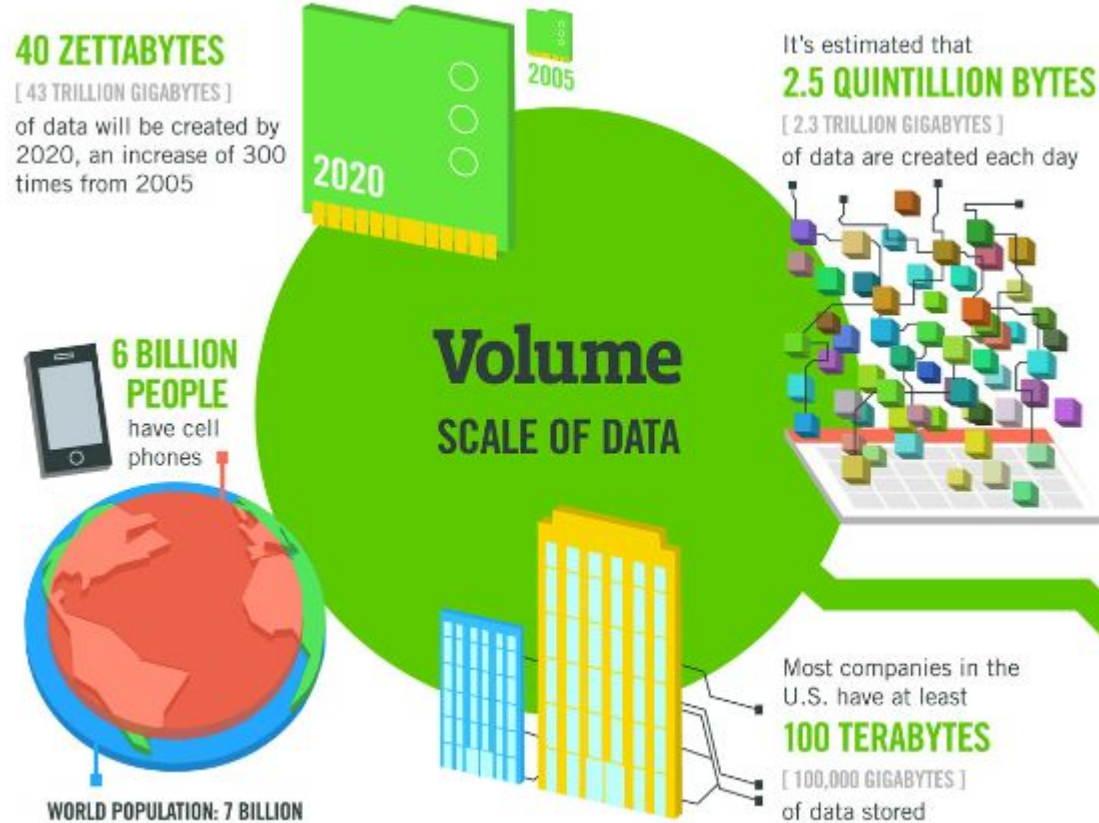
Moyenne, variance, corrélation, droite de régression identiques

Un volume toujours croissant...

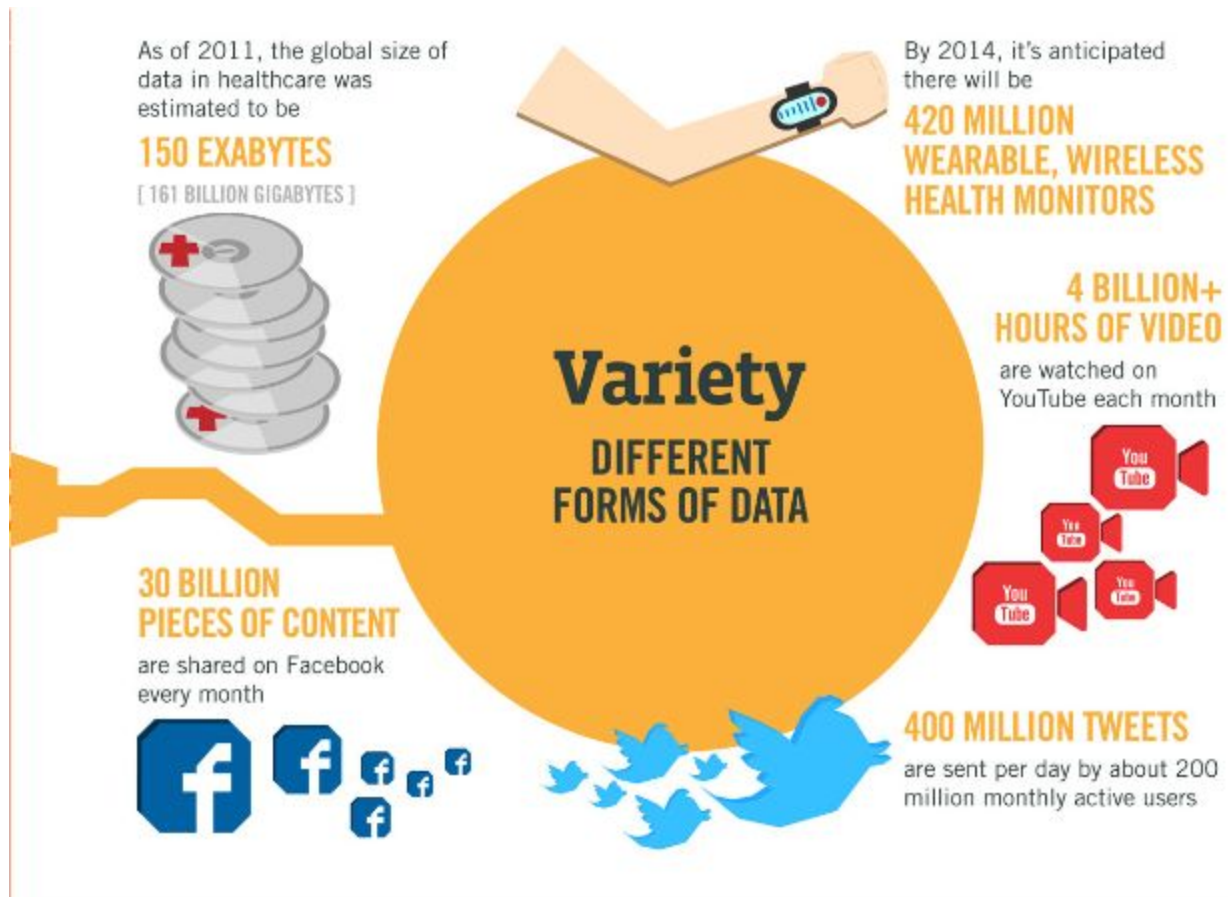


Volume

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



Variété



Vitesse

The New York Stock Exchange captures

1 TB OF TRADE INFORMATION

during each trading session



By 2016, it is projected there will be

18.9 BILLION NETWORK CONNECTIONS

– almost 2.5 connections per person on earth



Modern cars have close to

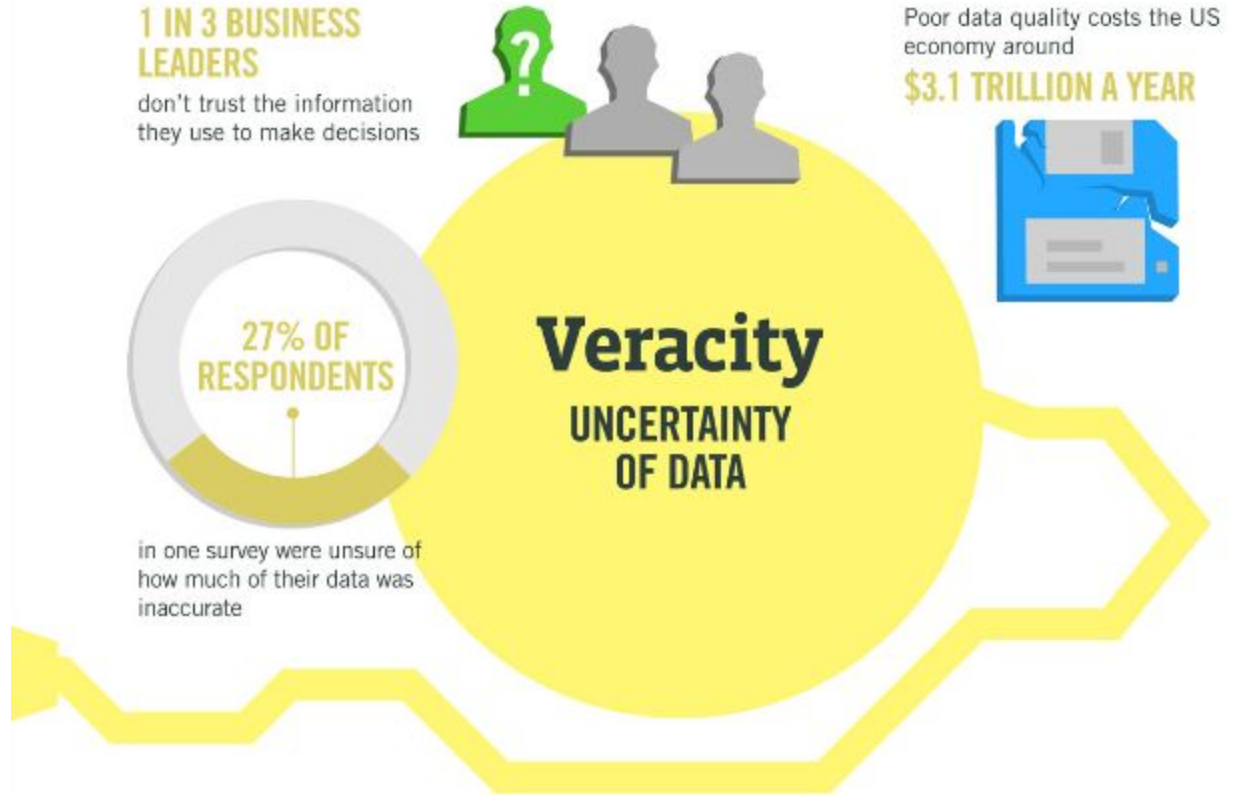
100 SENSORS

that monitor items such as fuel level and tire pressure

Velocity
ANALYSIS OF
STREAMING DATA



Véracité



Types de données

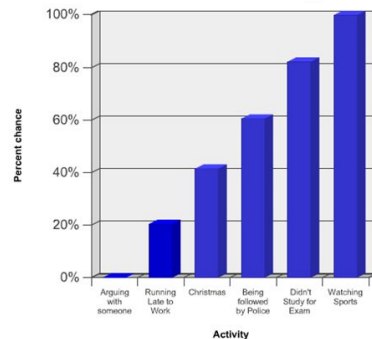
Données catégorielles

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

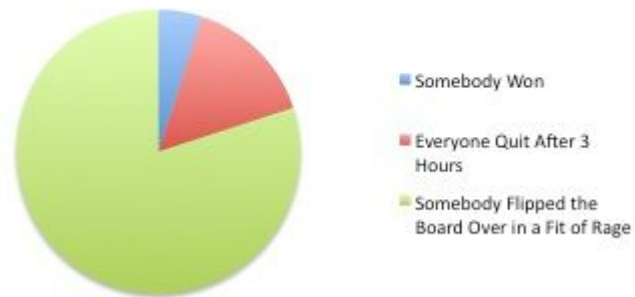
Opérations : = , ≠

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

Chances of an Atheist believing in God



Reasons for Ending a Game of Monopoly



Données ordinales

variable catégorielle numérique qui possède une propriété d'ordonnancement

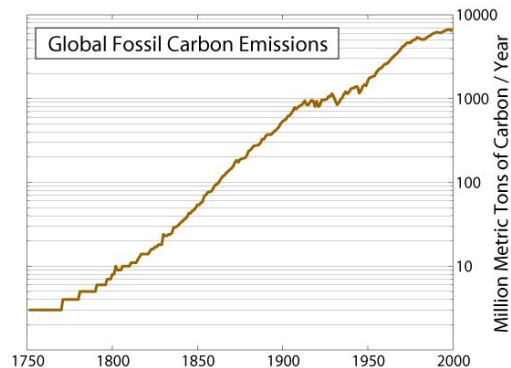
opérations : =, ≠, <, >



Données numériques

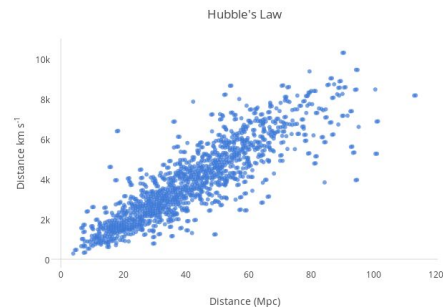
Données discrètes

- 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/Fahrenheit, dates, heures, QI, etc...

Ratio

Données numériques avec une référence absolue

Opérations : $=$, \neq , $+$, $-$, \times , $/$

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

Synthèse

	Catégorielle	Ordinale	Intervalle	Ratio
Ordonnement		✓	✓	✓
Fréquence de distribution	✓	✓	✓	✓
Mode	✓	✓	✓	✓
Médiane		✓	✓	✓
Moyenne			✓	✓
+ , -			✓	✓
x , /				✓
Référence absolue				✓

Example : 100 m Londres 2012

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

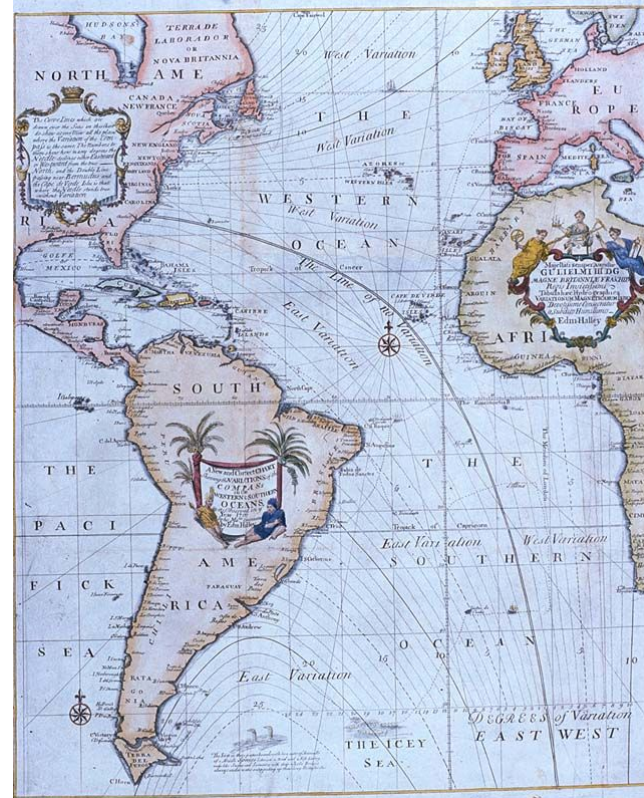
Rang	Ligne	Nom	Pays	Réaction	Temps
1	7	Usain Bolt	Jamaica	0.165	9.63
2	5	Yohan Blake	Jamaica	0.179	9.75
3	6	Justin Gatlin	United States	0.178	9.79
4	8	Ryan Bailey	United States	0.176	9.88
5	9	Churandy Martina	Netherlands	0.139	9.94
6	2	Richard Thompson	Trinidad and Tobago	0.160	9.98
7	3	Asafa Powell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

Historique

Champ magnétique terrestre

Halley 1701

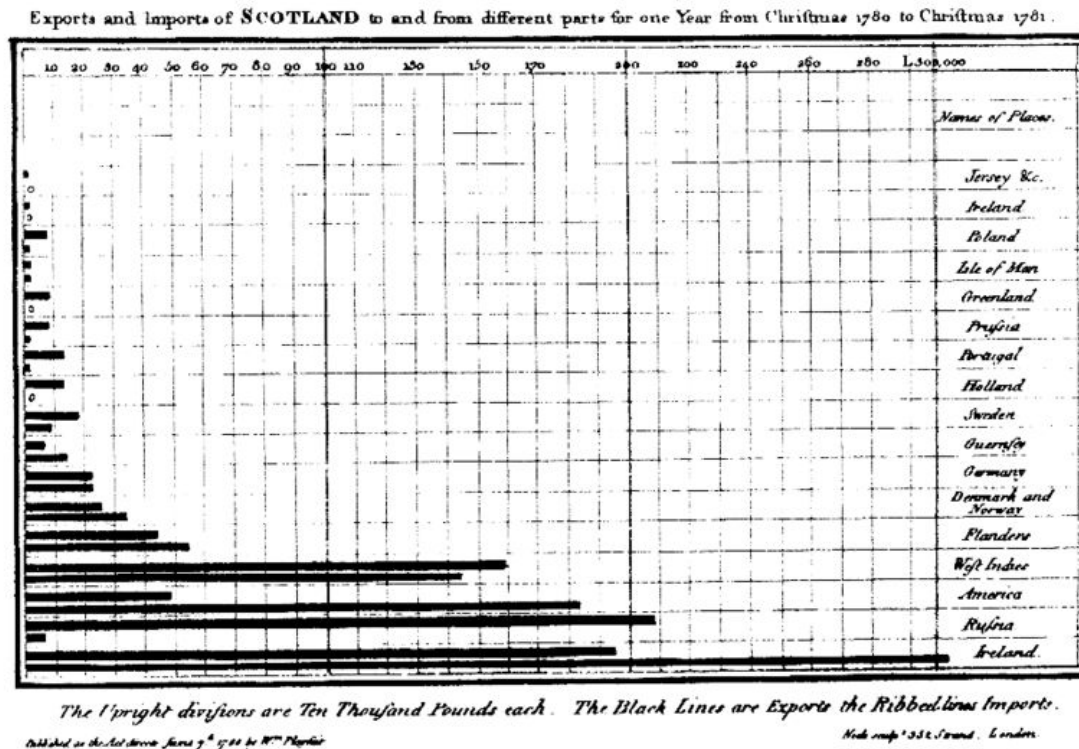
Première carte isogone (relie des points de même valeur)



Exportations de l'Ecosse

[Playfair](#) 1786

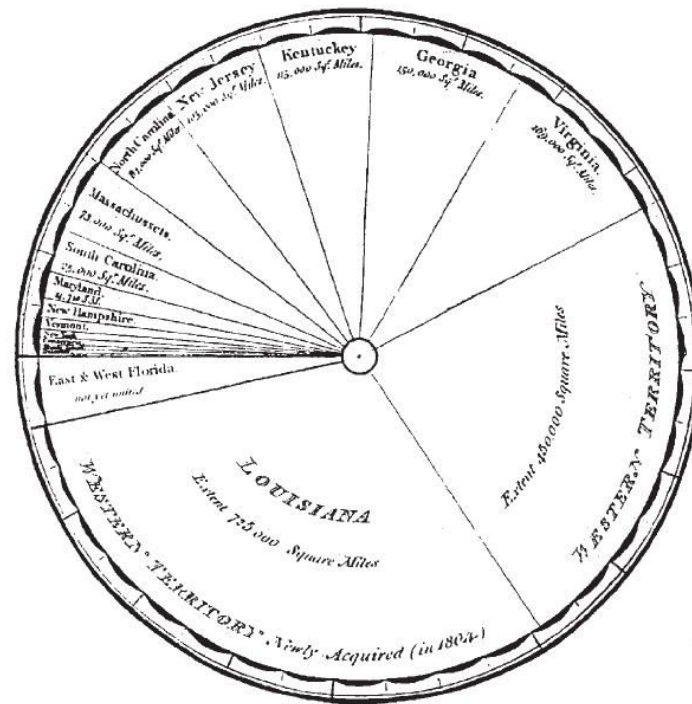
Premier bar graph



Commerce aux USA

Playfair 1805

Première "pie chart"



STATISTICAL REPRESENTATION of the UNITED STATES of AMERICA .

BY W. PLAYFAIR

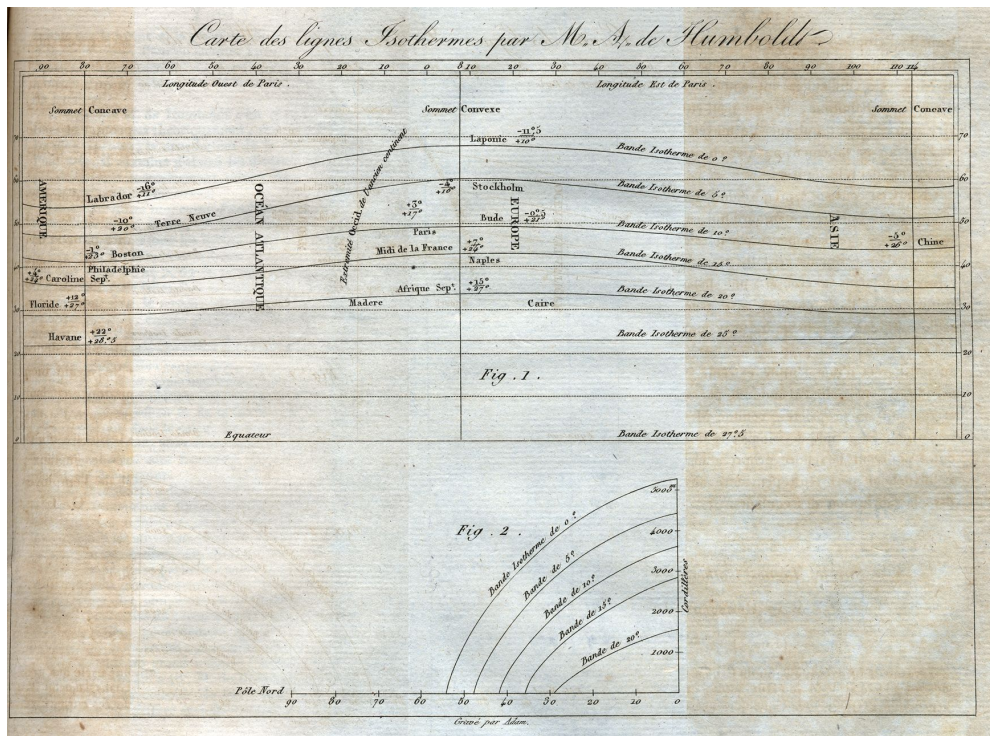
The Newly invented Method is intended to show the Proportions between the divisions in a striking Manner

Total Extent 1,275,000 Square Miles or 422 1/2 Millions of Acres.

Isothermes

[Humboldt](#) 1817

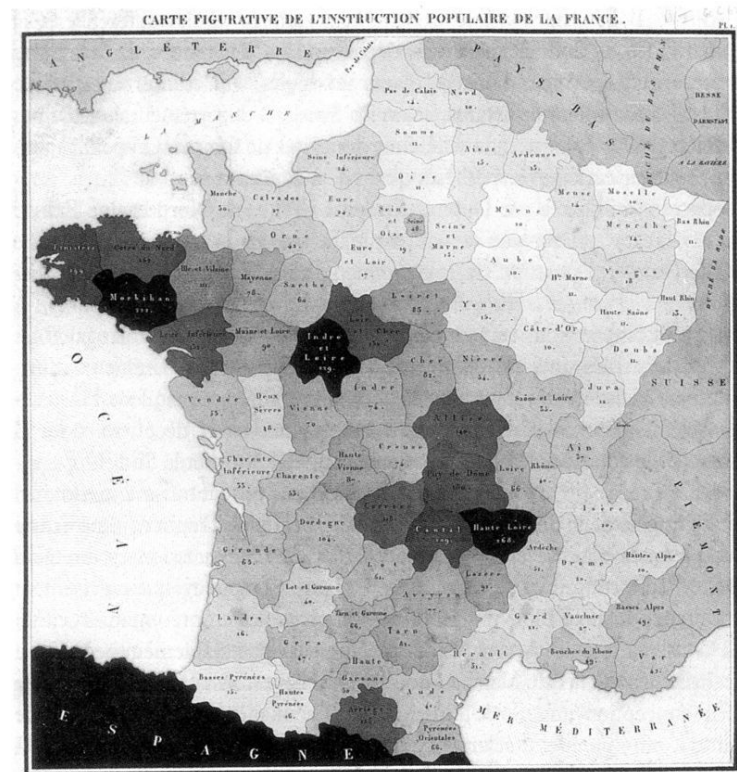
Premier graphe isotherme



L'instruction en France

[Dupin](#) 1826

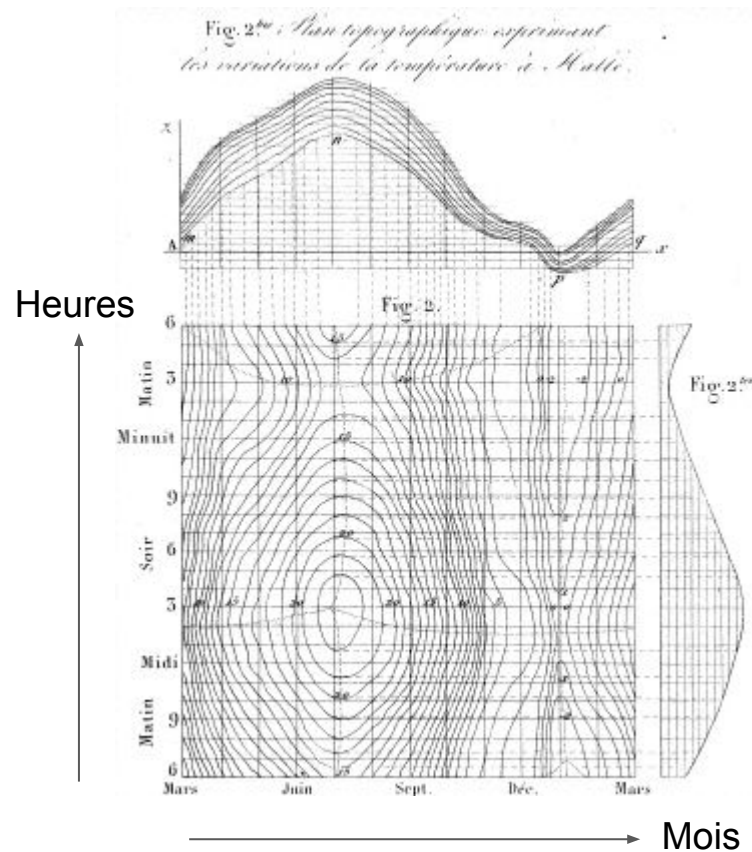
Première carte thématique choroplèthe



Contours de température

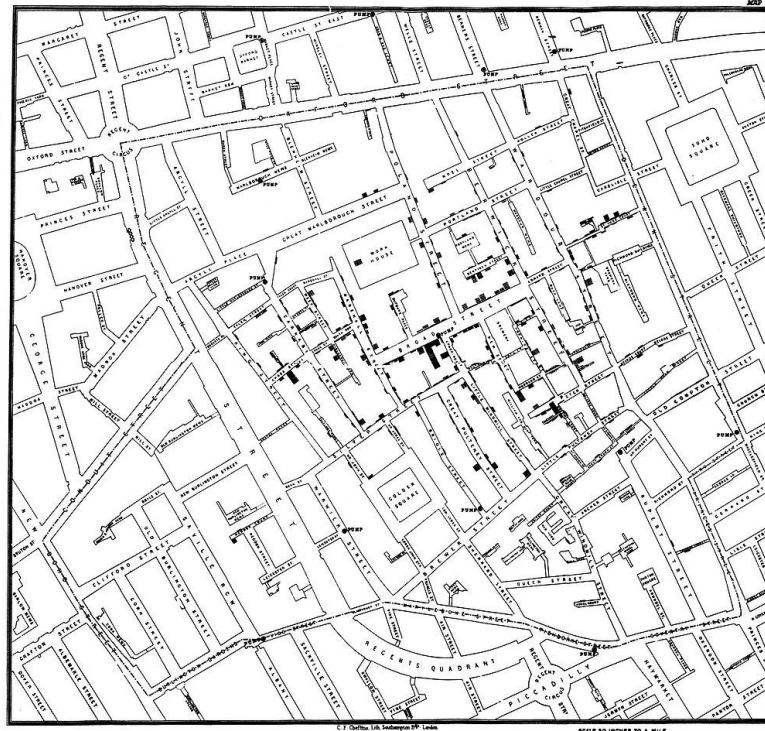
Lalanne 1843

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



Epidémie de Choléra à Londres

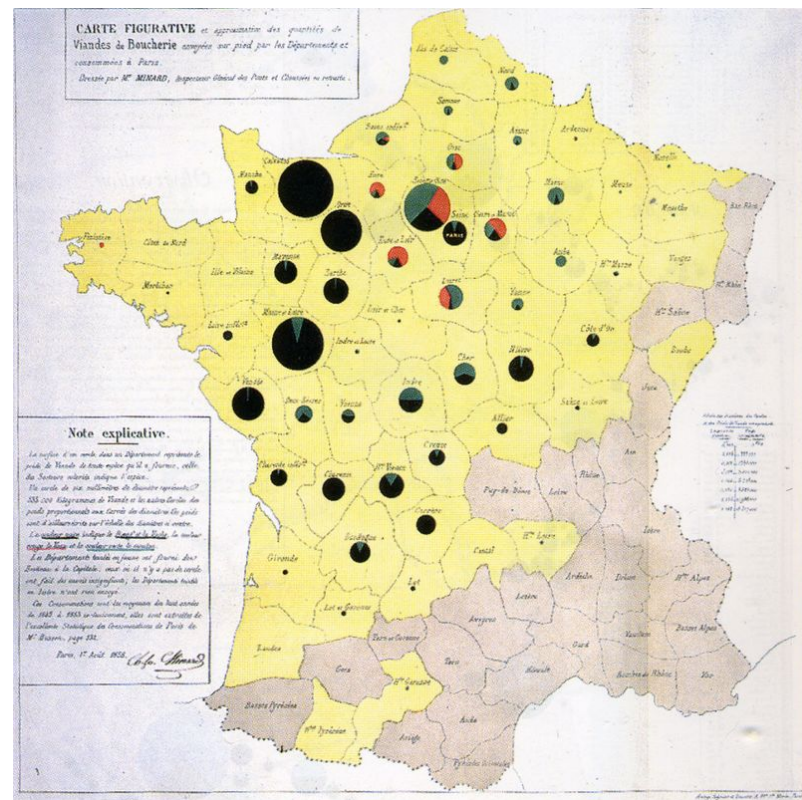
John Snow (1855)



Quantité de viande consommée à Paris

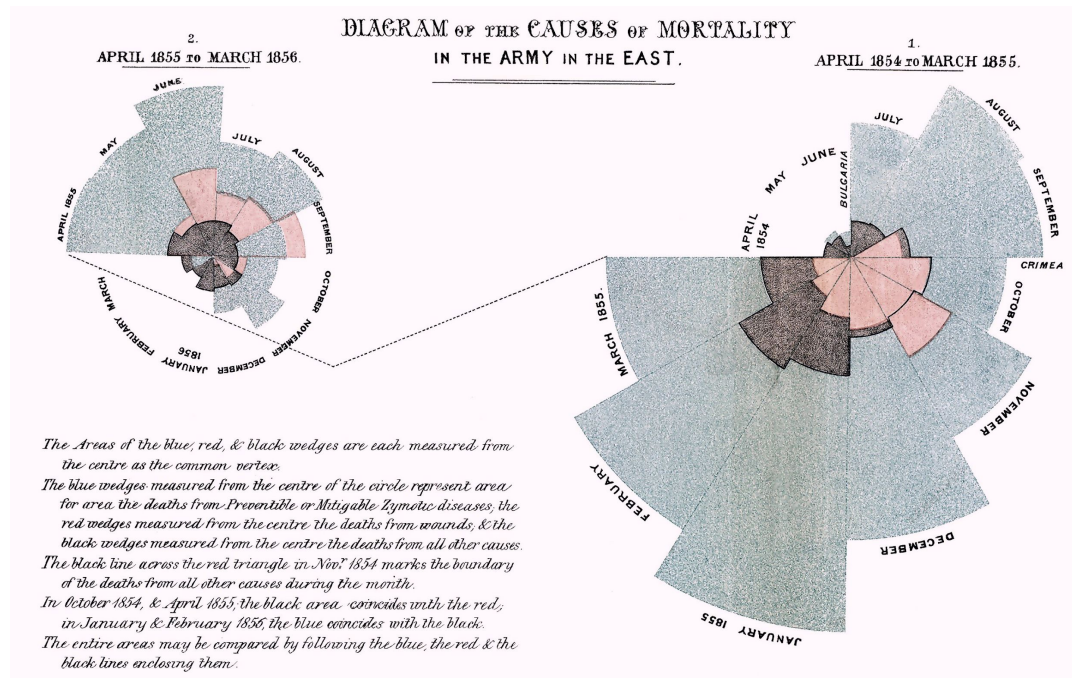
Minard 1858

Première carte incorporant des diagrammes



Mortalité dans l'armée anglaise

[Nightingale](#) 1858



Données météorologiques

Galton 1863

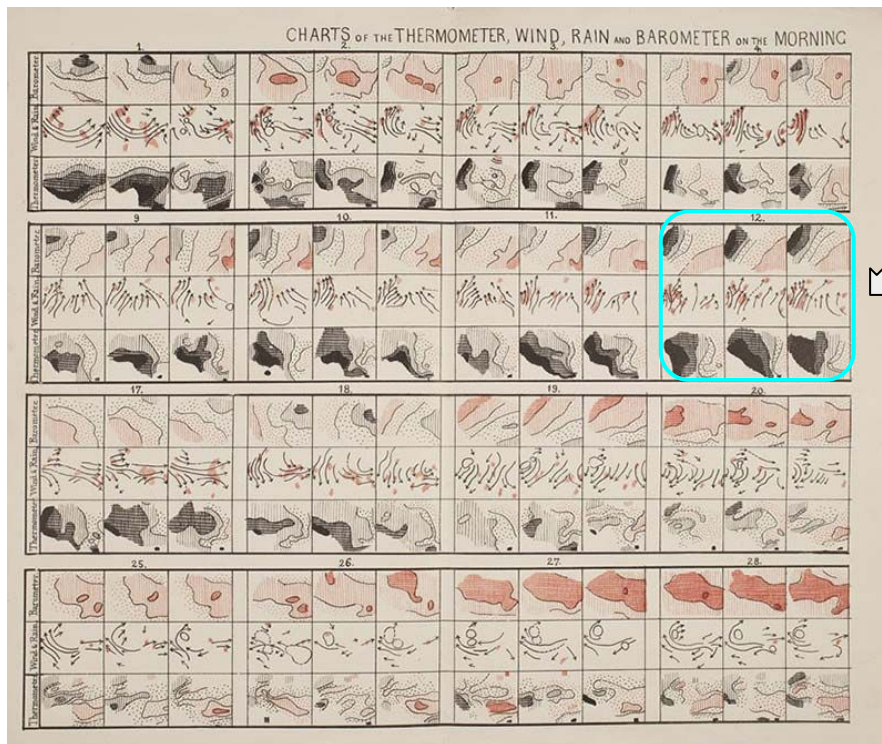
Matin - Midi - Soir



Pression atmosphérique
Vent et pluie
Température



Jour



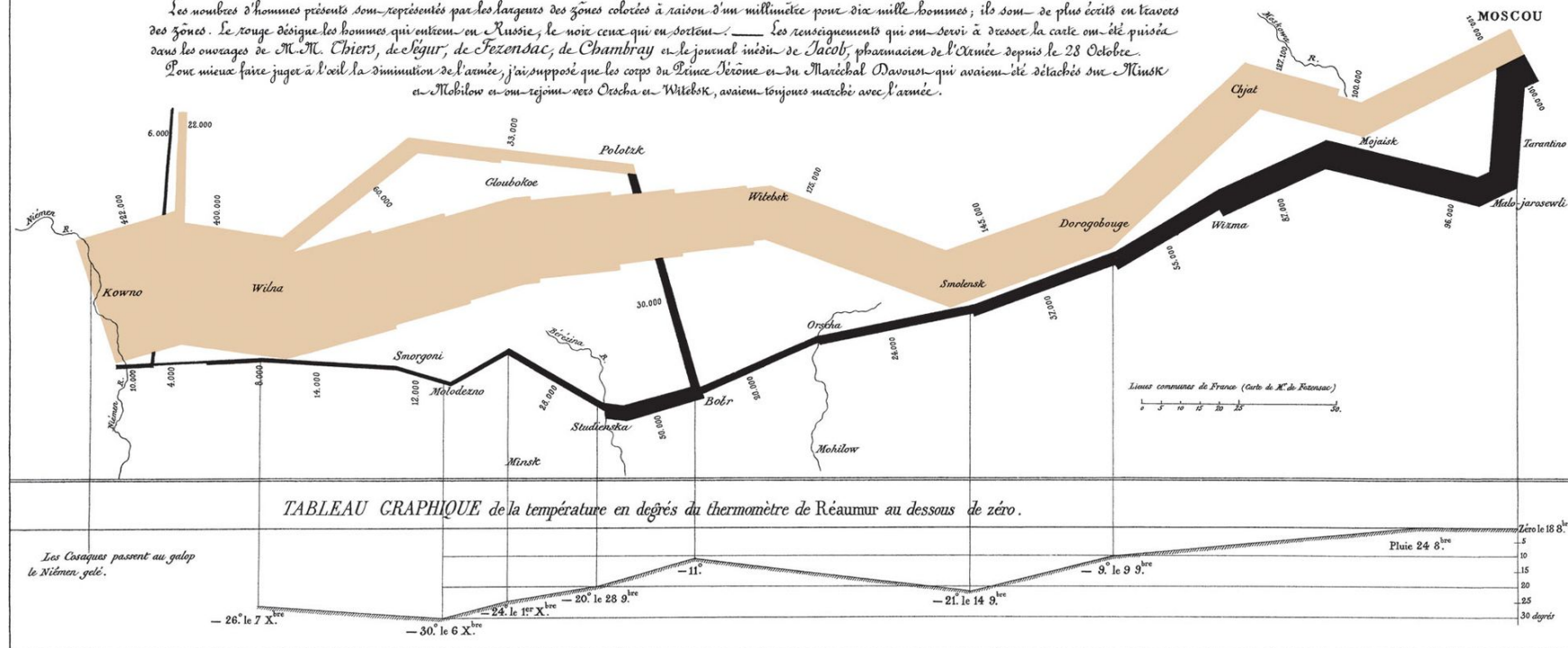
La retraite de Russie (Minard 1869)

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

Révisée 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. M. Chiers, de Fézard, de Chambray et le journal inédit 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 détachés sur Minsk et Mohilew et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.



Population de Paris en 1874

Vauthier 1874

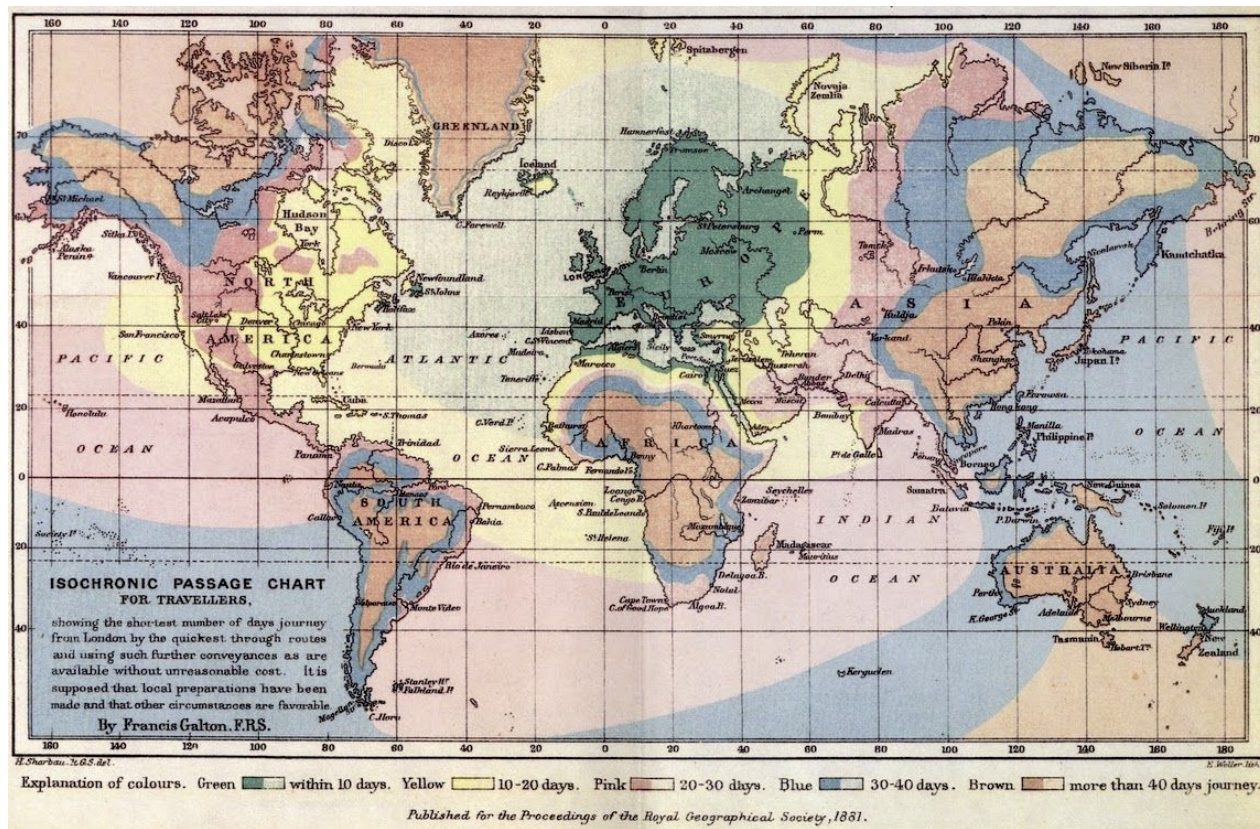
Première carte avec un tracé de données non géographiques



Durée de voyage depuis Londres

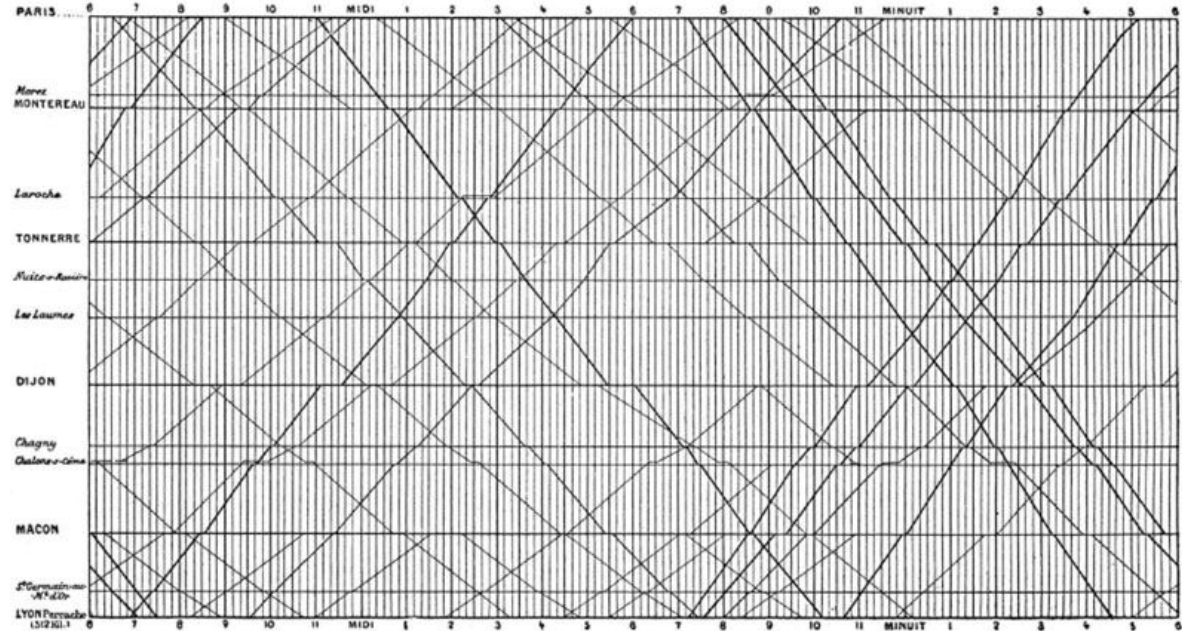
Galton 1881

Première carte isochrone



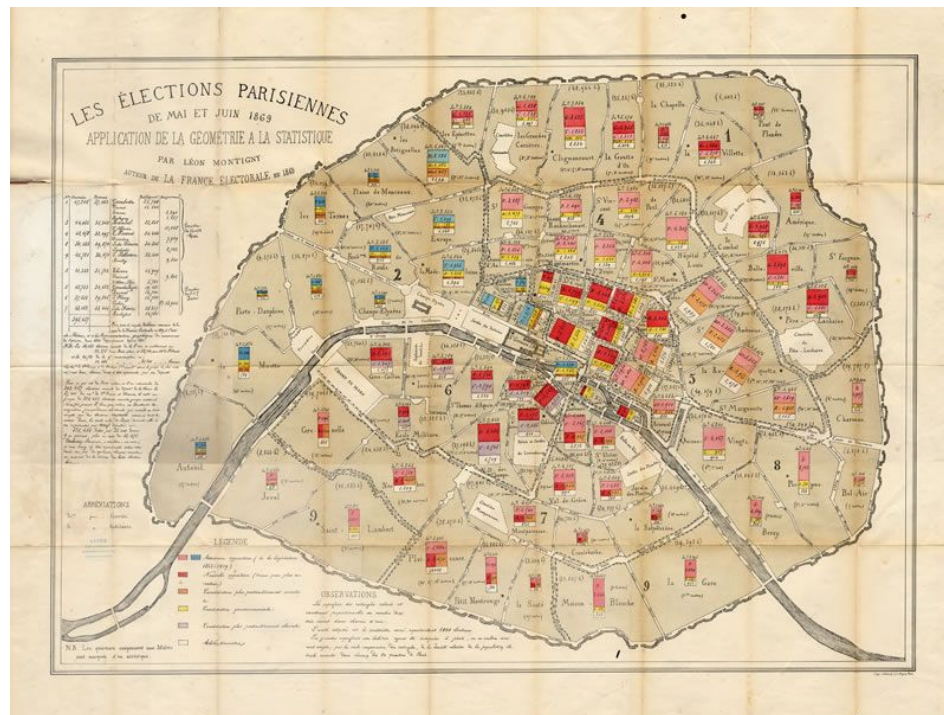
Marche des trains

Ibry (1885)



Plus d'histoire...

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



Éléments de perception visuelle

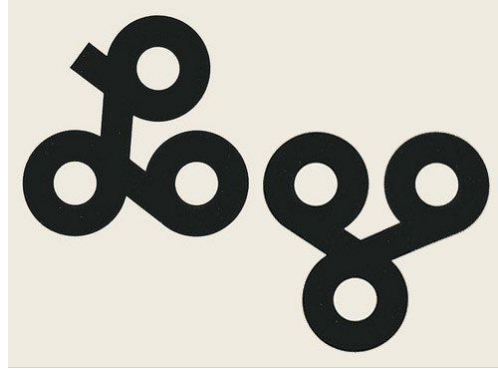
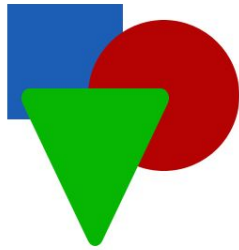
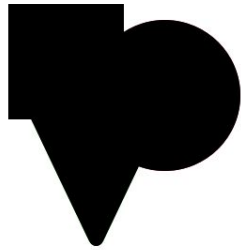
Introduction

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](#)]

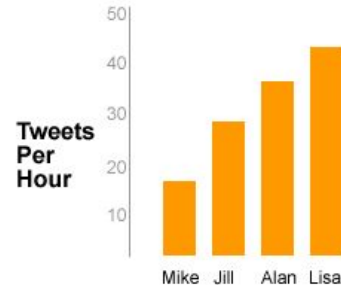
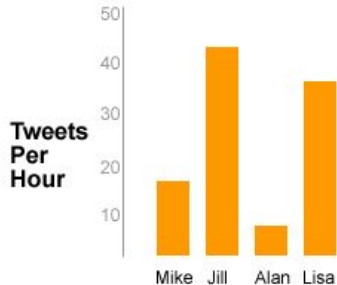
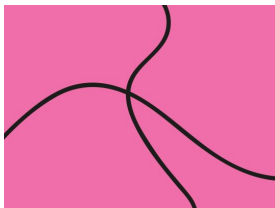
La loi de la bonne forme

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



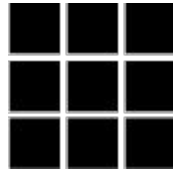
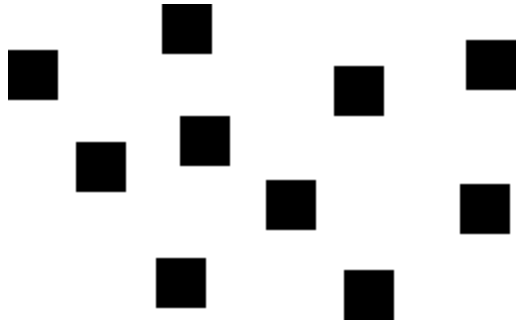
La loi de continuité

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



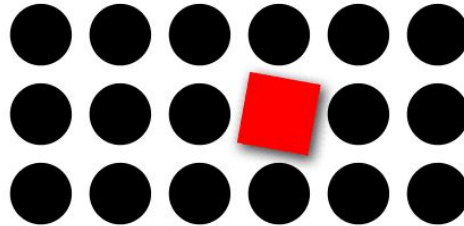
La loi de la proximité

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



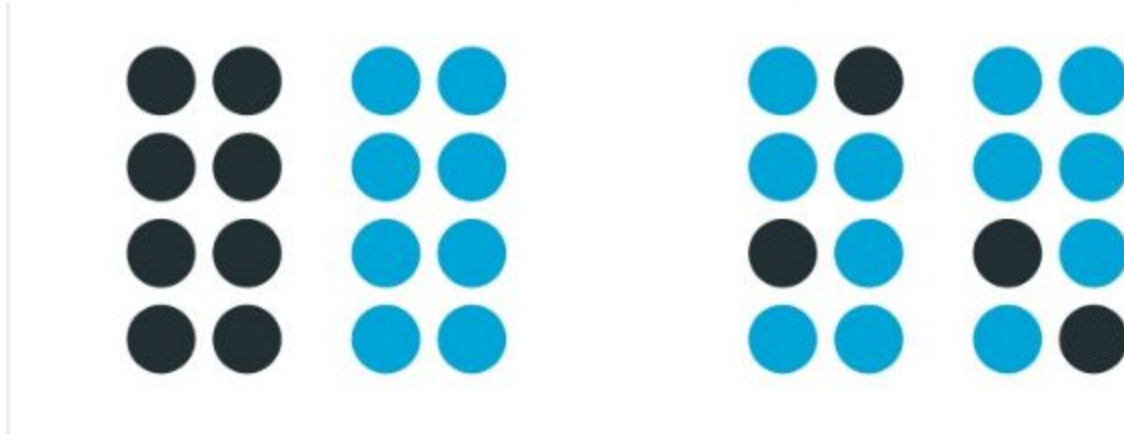
La loi de similitude / anomalie

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



Proximité et similarité

La proximité prédomine...



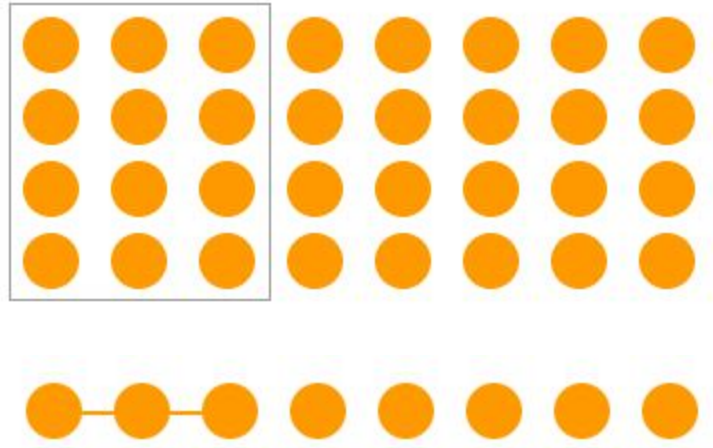
Loi de la symétrie

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



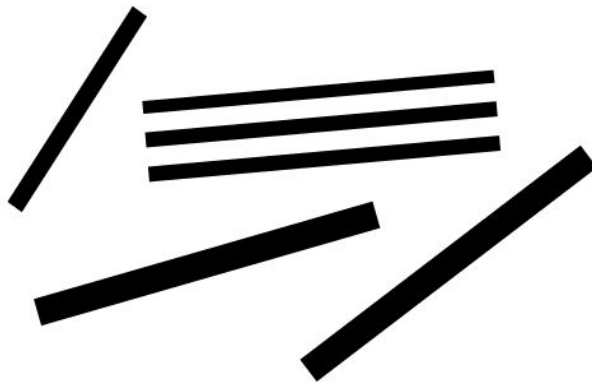
Loi de la connectivité

Des éléments connectés sont perçus comme étant liés...



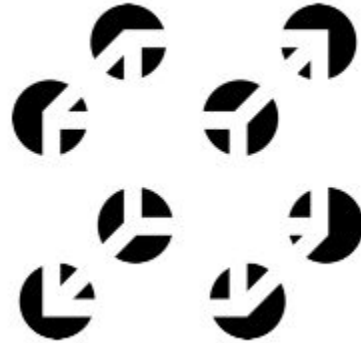
La loi de destin commun

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

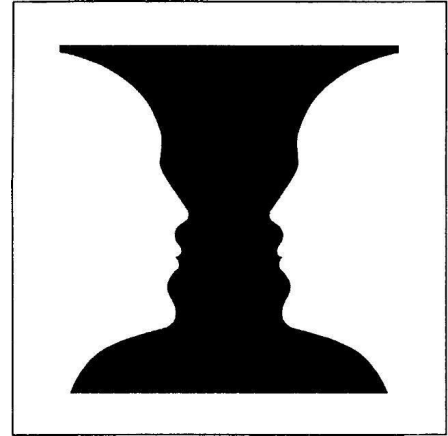


La loi de la clôture

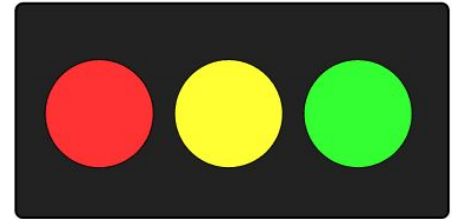
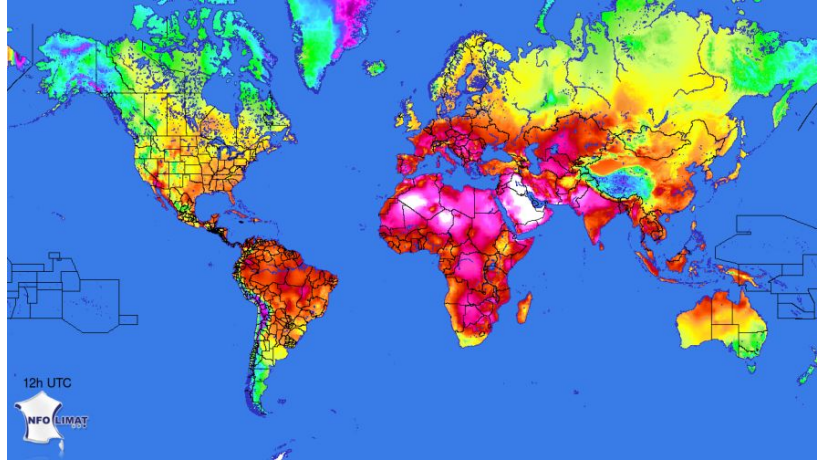
Un arrangement complexe de formes individuelles tendent à être perçues comme une seule forme simple à identifier...



La distinction figure - fond



L'expérience du passé

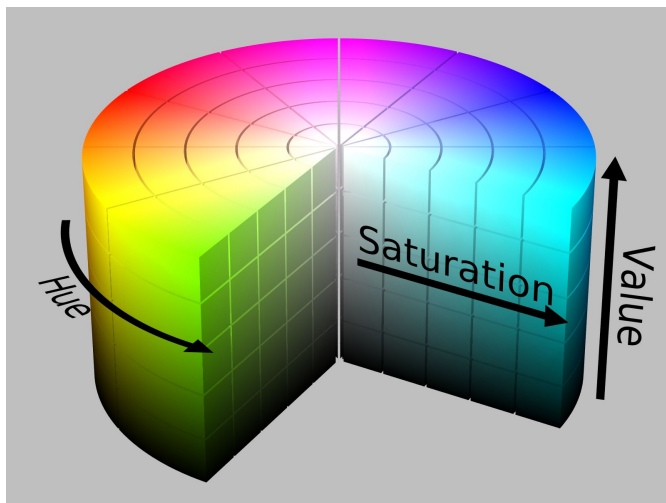


Les couleurs

Codage des couleurs

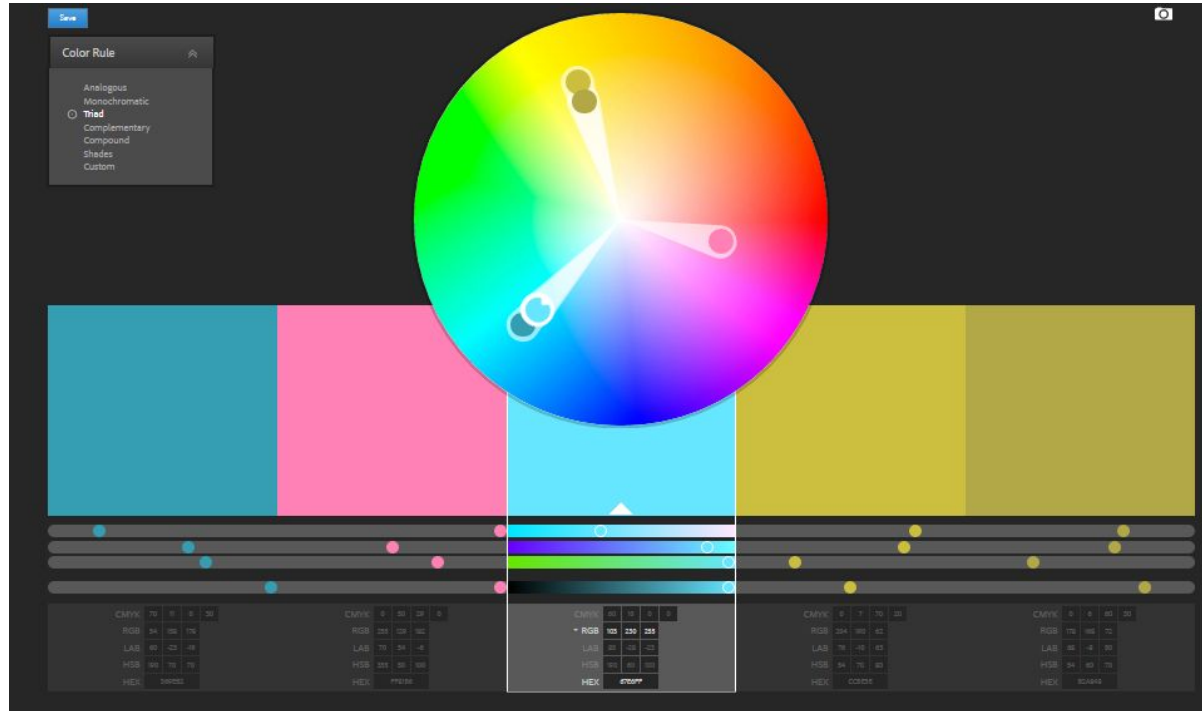
RGB (red, green, blue) : codage informatique, proche du matériel

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



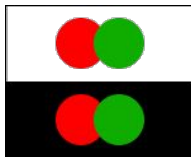
Choix d'une palette de couleur

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

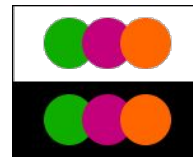
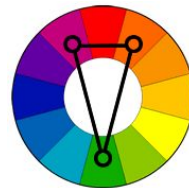


Combinaisons harmonieuses

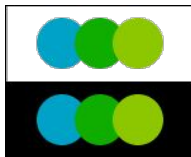
Complementary



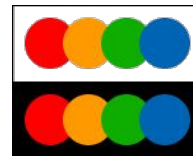
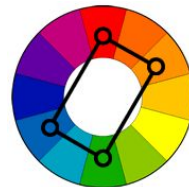
Split complementary



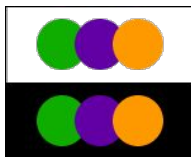
Analogous



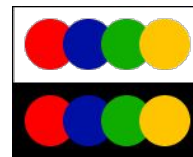
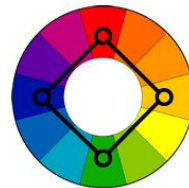
Tetradic



Triadic



Square



Les 10 commandements...

THE 10 COMMANDMENTS OF COLOR THEORY

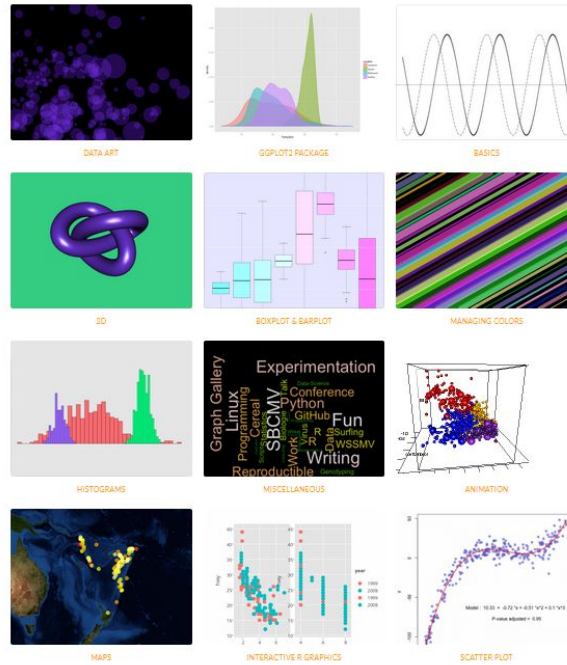
<p>1</p> <p>KNOW THE COLOR WHEEL WELL! DO YOU KNOW WHAT EACH COLOR SIGNIFIES?</p>  <p>RED LIVE, ENERGY, INTENSITY</p> <p>YELLOW JOY, INTELLECT, EXTENSION</p> <p>GREEN FRESHNESS, SAFETY, GROWTH</p> <p>BLUE STABILITY, TRUST, SERENITY</p> <p>PURPLE ROYALTY, WEALTH, FEMININITY</p>	<p>2</p> <p>MATCH IT. DO 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</p> <p>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</p> <p>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</p> <p>NEED MORE VARIATIONS? GO DOUBLE COMPLEMENTARY!</p>  <p>Barber</p> <p>wheelchair</p> <p>CupCake</p> <p>Sandals</p> <p>Television</p>	<p>6</p> <p>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</p> <p>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</p> <p>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</p> <p>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</p> <p>AND LASTLY, RGB, CMYK AND PANTONE ARE NOT THE SAME!</p>  <p>CupCake</p>  <p>CupCake</p>  <p>CupCake</p>
--	---	---	--	--	---	---	--	--	---

Principaux types de graphiques

Graph galleries



The Data Visualisation Catalogue



Histogrammes et densité

Représentation de données numériques

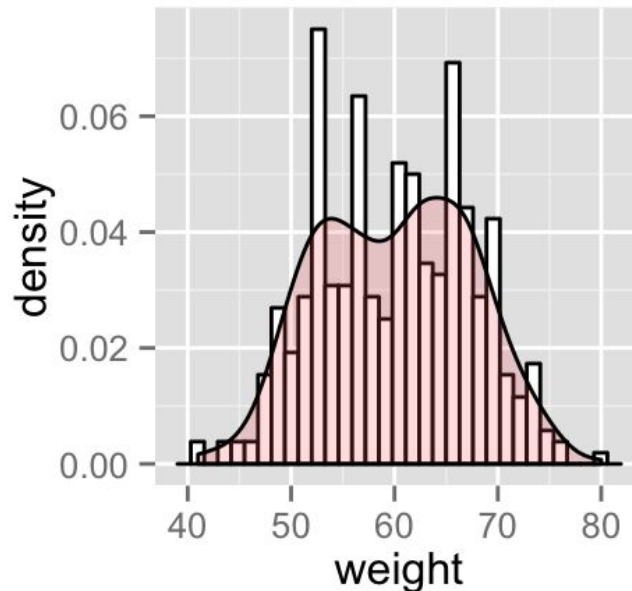
Permet d'estimer la distribution d'une variable continue

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)



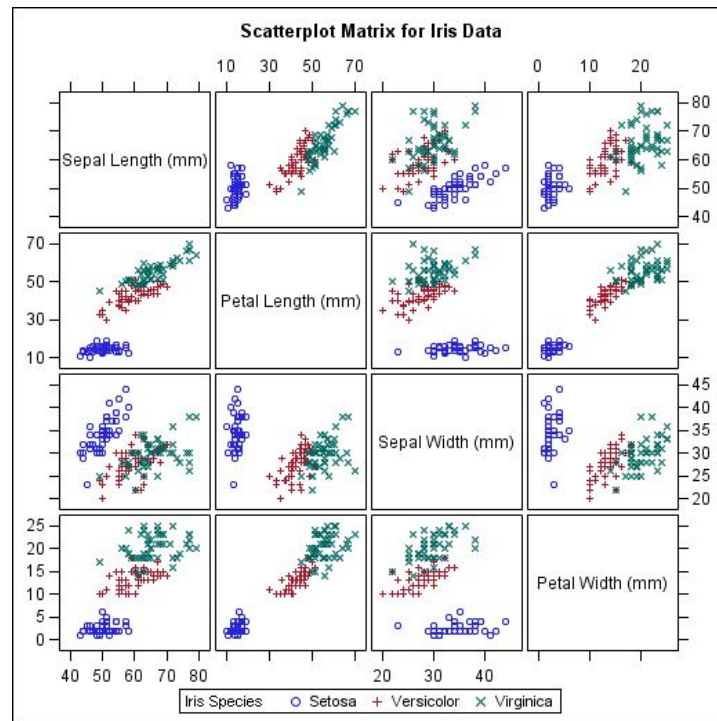
<https://tinlizzie.org/histograms>

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

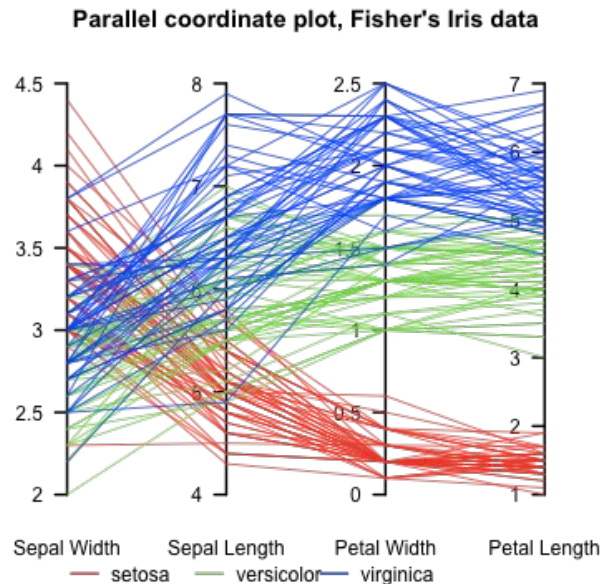


Table de contingence

Représentation des données numériques relatives à 2 ou plusieurs variables catégorielles

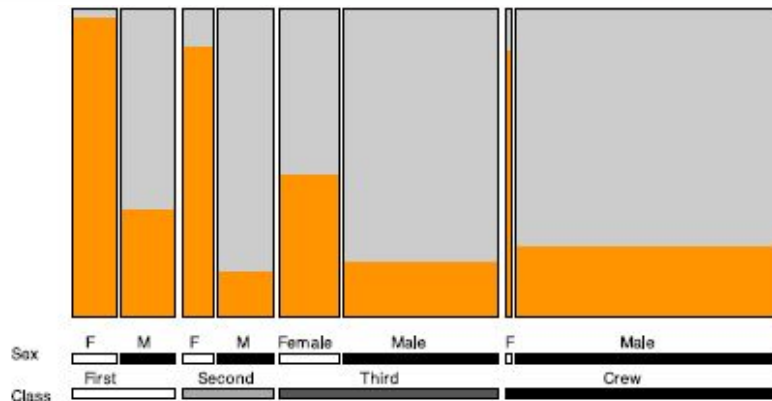
Class:	First		Second		Third		Crew	
Gender:	Female	Male	Female	Male	Female	Male	Female	Male
Survived:								
yes	141	62	93	25	90	88	20	192
no	4	118	13	154	106	422	3	670
Survival Rate (in %)	97	34	88	14	46	17	87	22

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

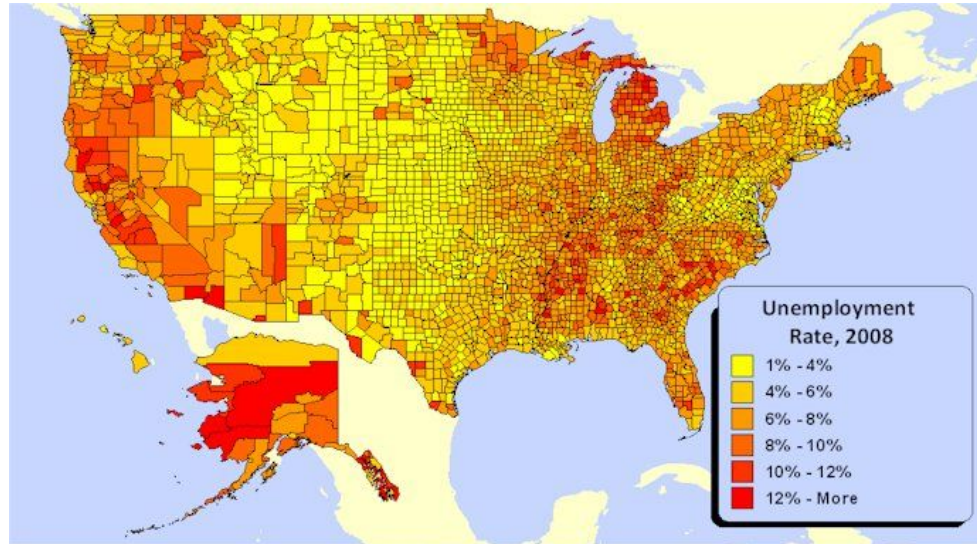
Mosaic plot

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



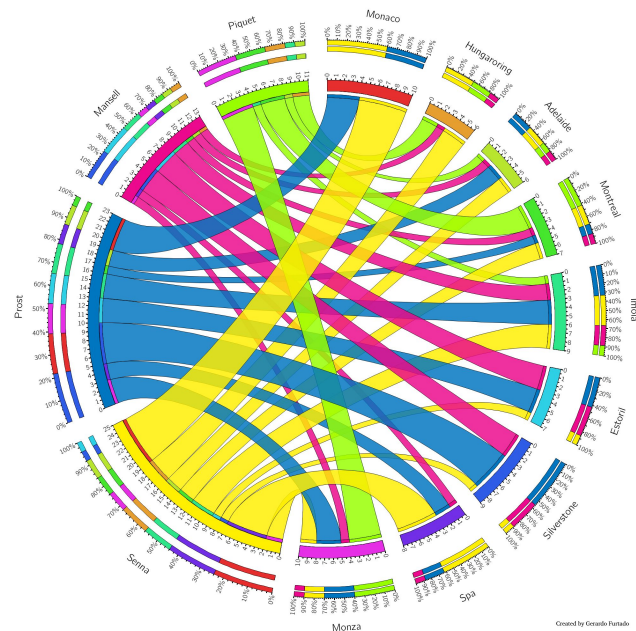
Carte choroplèthe



Chord diagram

4 CHAMPIONS, 9 TRACKS

4 champions from the 80's and early 90's (Senna, Piquet, Prost and Mansell) and 9 tracks: 3 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.



Choix du type de graphique (1/3)

See relationships among data points

-  Scatterplot
-  Matrix Chart
-  Network Diagram

Compare a set of values

-  Bar Chart
-  Block Histogram
-  Bubble Chart

Track rises and falls over time

-  Line Graph
-  Stack Graph
-  [Stack Graph for Categories](#)

See the parts of a whole

-  Pie Chart
-  Treemap
-  Treemap for Comparisons

Analyze a text

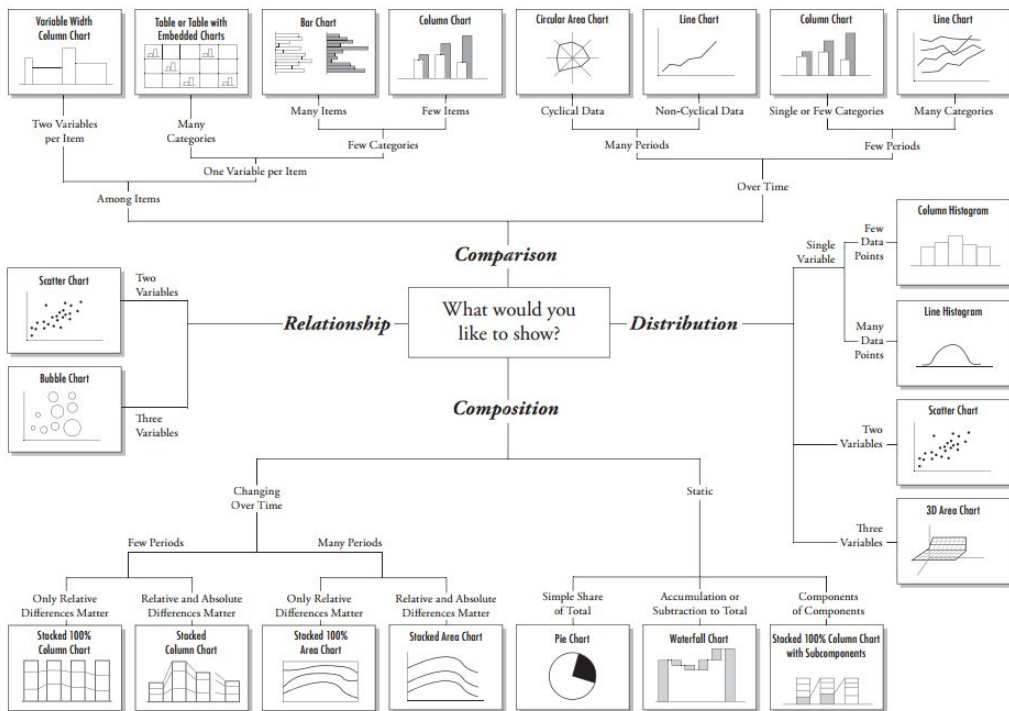
-  Word Tree
-  Tag Cloud
-  Phrase Net
-  Word Cloud Generator

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 cartouche	 Data Visualization Visual representations of quantitative data in schematic form (either with or without axes)	 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	 Metaphor Visualization Visual Metaphor pattern information graphically to organize and structure information. This also conveys an insight about the represented information through the key characteristics of the metaphor that is employed	 Gt cartoon
 Pi pie chart	 L line chart	 Compound Visualization The complementary use of different graphic representation formats in one single scheme or frame	 Gt rich picture
 B bar chart	 Ae area chart		 Kn knowledge map
 Hi histogram	 Sc scatterplot	 R radar chart	 D diagram
 Tk box plot	 Sp spectrogram	 Pa parallel coordinates	 Pr parameter ruler
 Da data map	 Tp treemap	 Hy hypertext	 Cf concept fan
 Gn cone tree	 Sy system dyn./simulation	 Cy cycle diagram	 Co concept map
 Df data flow diagram	 Se semantic network	 T timeline	 Ic iceberg
 So self system modeling	 Sn synergy map	 Ve ven diagram	 Lm learning map
 Fo force field diagram	 Ib argumentation map	 Mi mind map	
 Pr process event chain	 Pe petri chart	 Sq square oppositions	
		 Ce cause-effect chain	
		 Ar argument slide	
		 Sw swim lane diagram	
		 Gc game chart	
		 Pm perspective diagram	
		 Dt decision tree	
		 Bp open critical path method	
		 Tl toolmap	
		 Gt graph	
		 Fp flight plan	
		 Me meeting trace	
		 Tm temple	
		 St story template	
		 Tr tree	
		 Br bridge	
		 Fu funnel	
		 Cs concept cocoon	
		 Gt graph	
		 Co communication diagram	
		 Fp flight plan	

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

Cy	Process Visualization
Hy	Structure Visualization
	Overview
	Detail
	Detail AND Overview
< >	Divergent thinking
> <	Convergent thinking

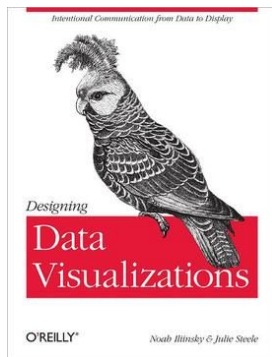
☼☼ Su supply demand curve	☼☼ Pe performance	☼☼ St strategy map	☼☼ Oc cost organization	☼☼ Ho history of quality	☼☼ Fd feedback circuit	☼☼ Ft failure tree	☼☼ Mq margin management	☼☼ Ld life-cycle product	☼☼ Bo partner's strategy	☼☼ Cs s-cycle	☼☼ Sm stockholder model	☼☼ Is ishikawa diagram	☼☼ Tc technology roadmap
☼☼ Ed edge warrior box	☼☼ Pf portfolio graph	☼☼ Sg strategic game board	☼☼ Mz marketing's graph	☼☼ Z zwoy's morphological	☼☼ Ad advertising strategy	☼☼ De decision discovery heuristic	☼☼ Bm big matrix	☼☼ Stc strategy canvas	☼☼ Ve value chain	☼☼ Hy hyper-cycle	☼☼ Sr shareholder rating map	☼☼ Ta type	☼☼ Sd spread diagram

Bonnes pratiques

Codes visuels

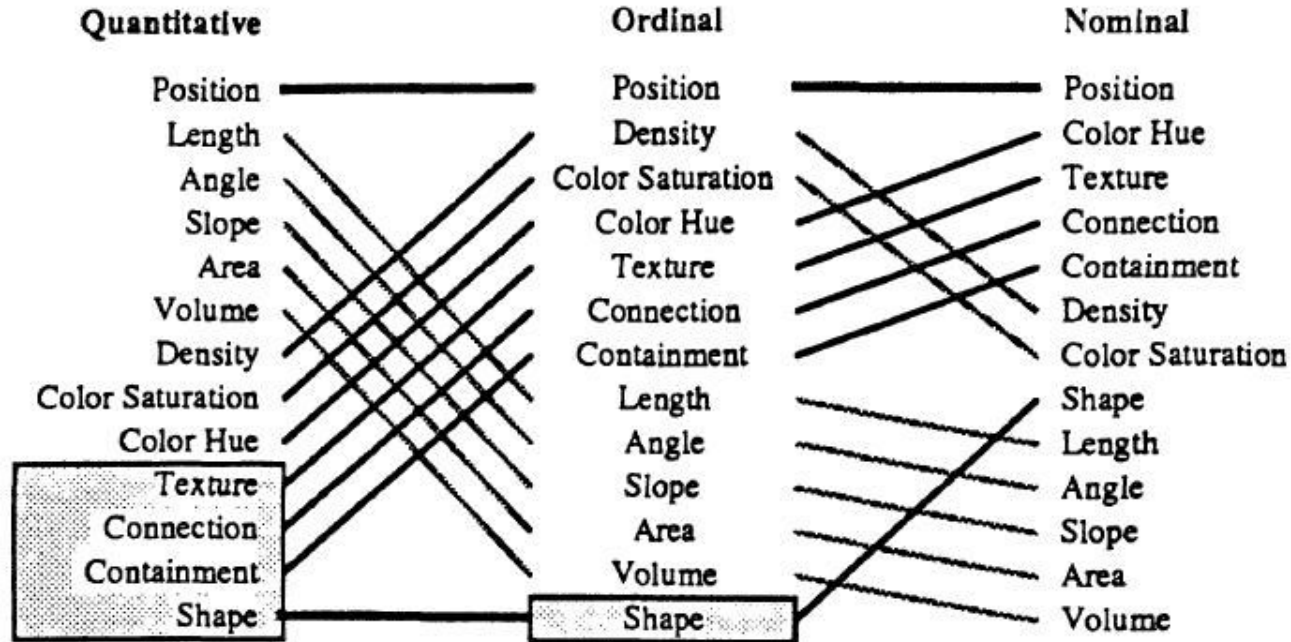
Designing Data Visualizations

Noah Iliinsky...



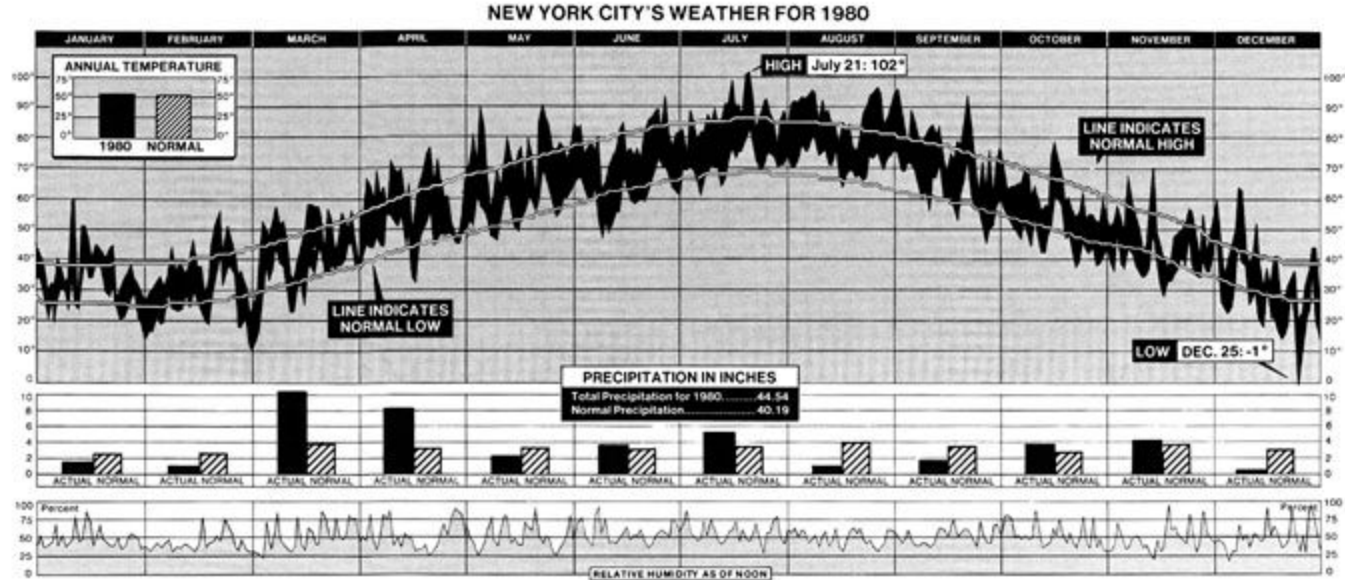
Example	Encoding	Ordered	Useful values	Quantitative	Ordinal	Categorical	Relational
	position, placement	yes	infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	text labels	optional alpha or num	infinite	Good	Good	Good	Good
	length	yes	many	Good	Good		
	size, area	yes	many	Good	Good		
	angle	yes	medium	Good	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

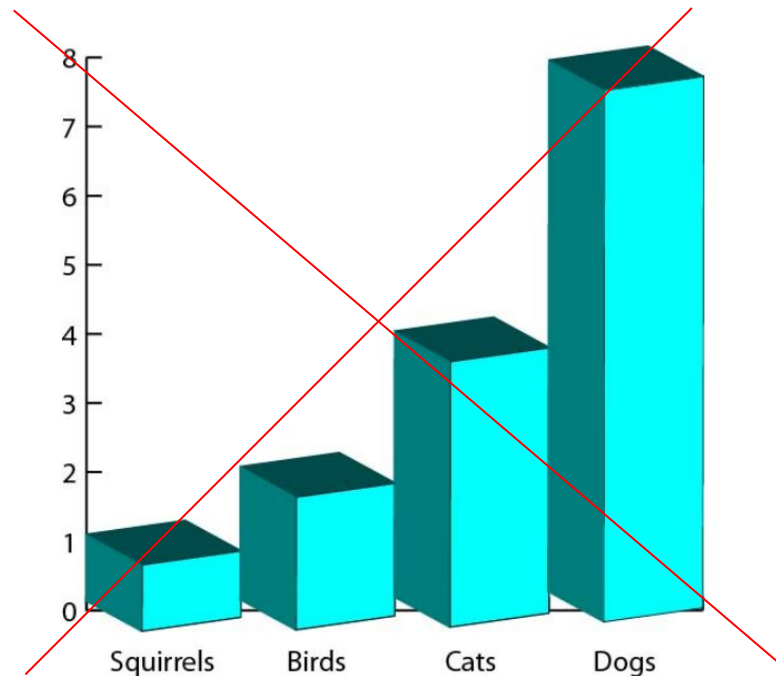
2200 données numériques représentées !



Exemples à éviter

Principes de Tufte

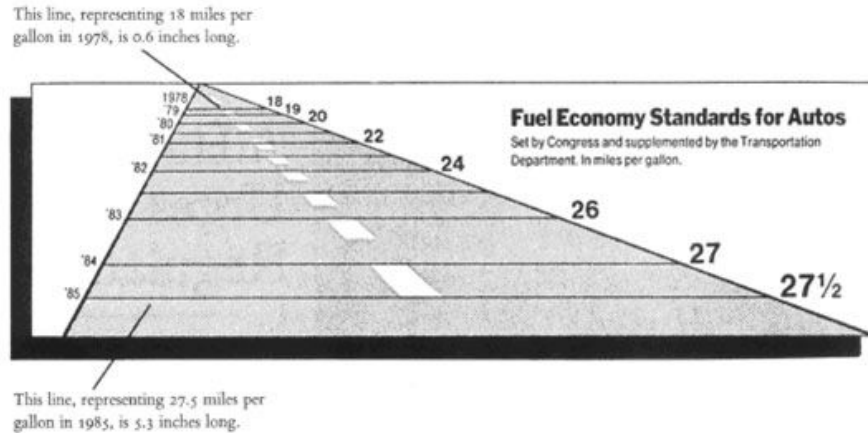
1. La représentation des grandeurs (ou des variations de ces grandeurs) doit être directement proportionnelle aux valeurs numériques
2. Éviter toute ambiguïté et toute distorsion graphique
3. La dimension du graphique ne doit pas être supérieure à la dimension des données
4. Maximiser le “data ink ratio”



Attention au “lie factor”

LF = surface graphique / valeur de la donnée

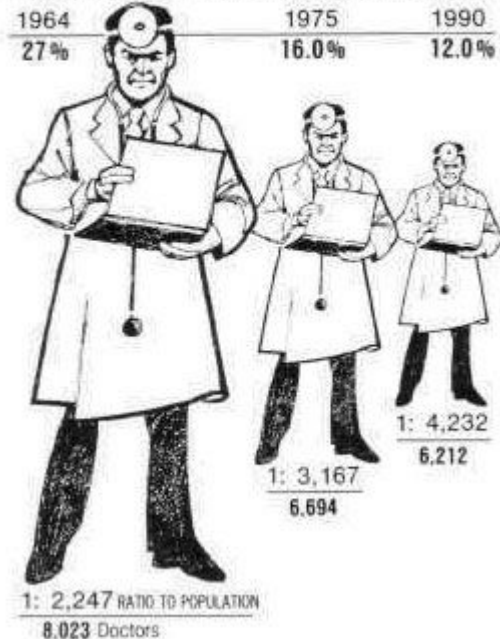
[Tufte 1991]



THE SHRINKING FAMILY DOCTOR In California

Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27%	16.0%	12.0%

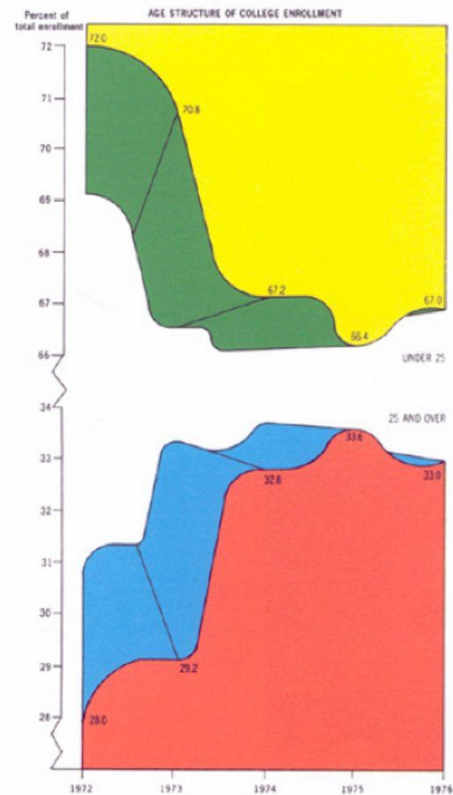


The worst chart ever...

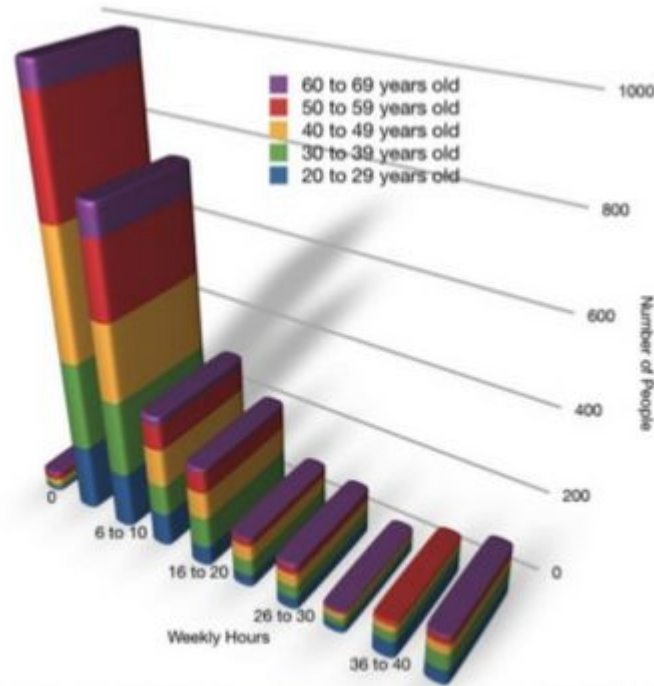
La couleur et la 3D ne servent à rien

L'axe des Y n'est pas continu

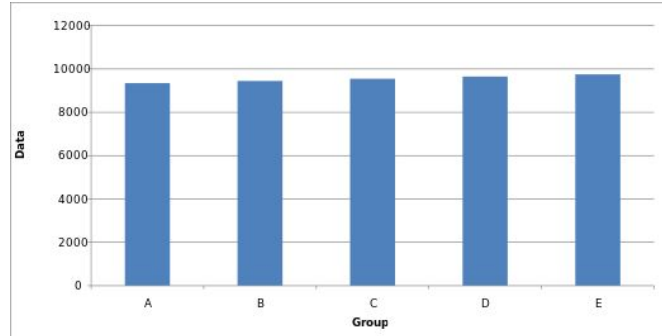
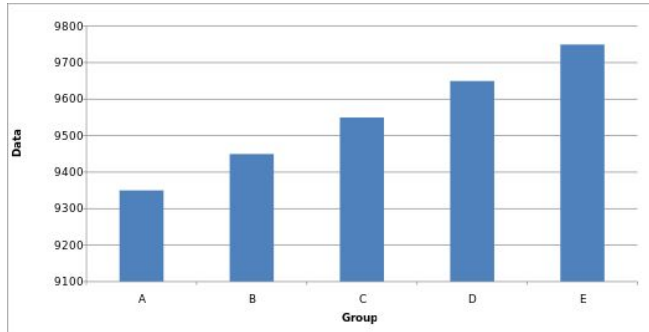
Seulement 5 chiffres à communiquer



Data ink ratio

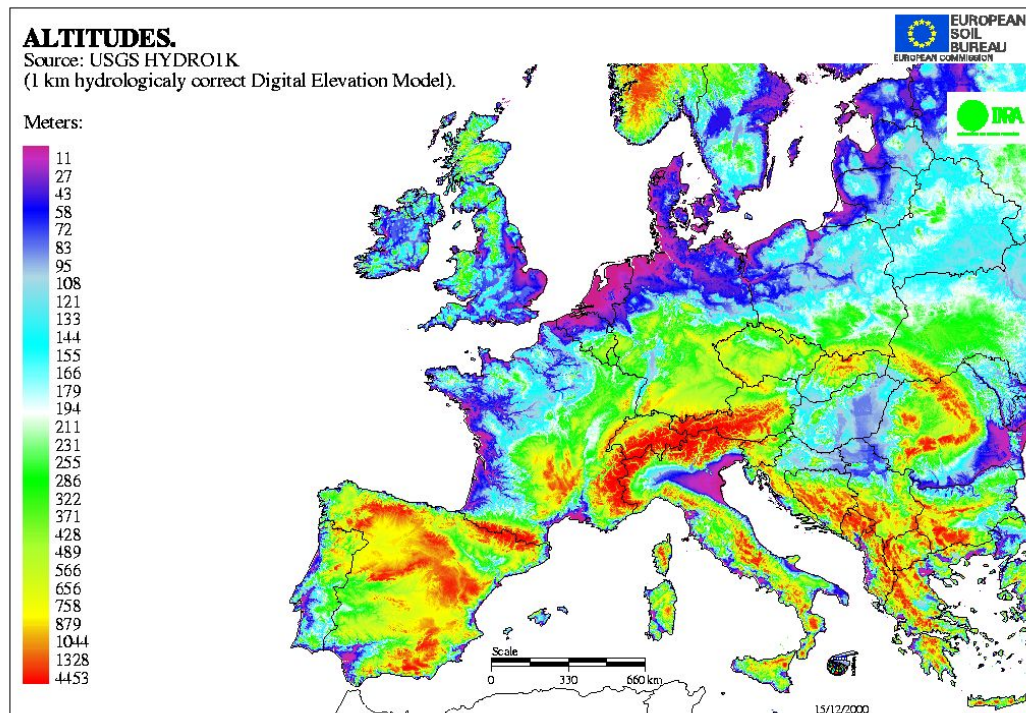


Graduation des axes



Mauvais choix de couleurs

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



D'autres exemples...

