

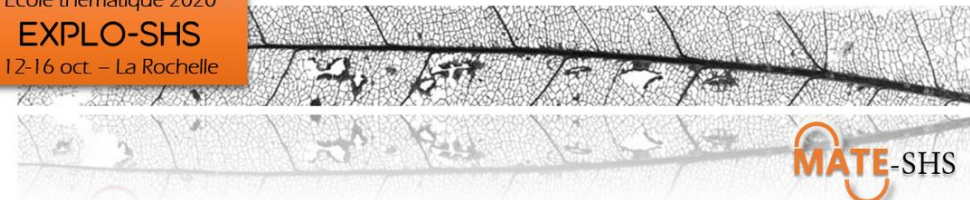
Principes de sémiologie graphique à des fins de visualisation

Najla Touati

Ingénieure géomaticienne

LISST UMR 5193 – UT2J-CNRS

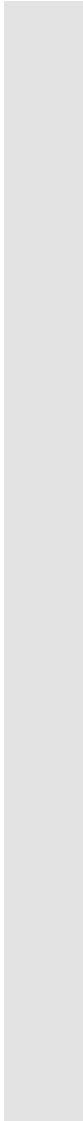
Ecole thématique 2020
EXPLO-SHS
12-16 oct. – La Rochelle



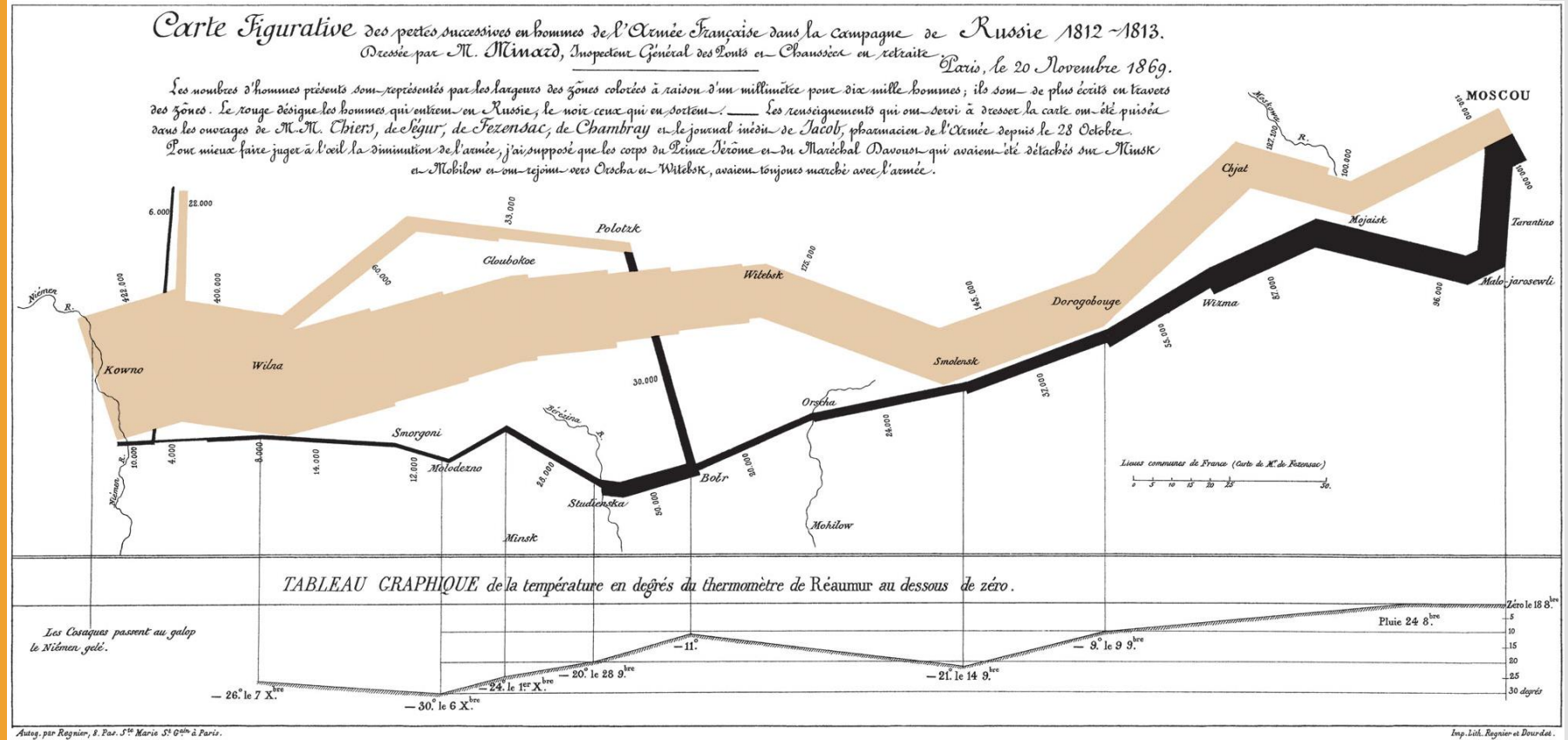


Introduction : un peu d'histoire...

Au commencement, une volonté de comprendre un phénomène
par l'exploration des données



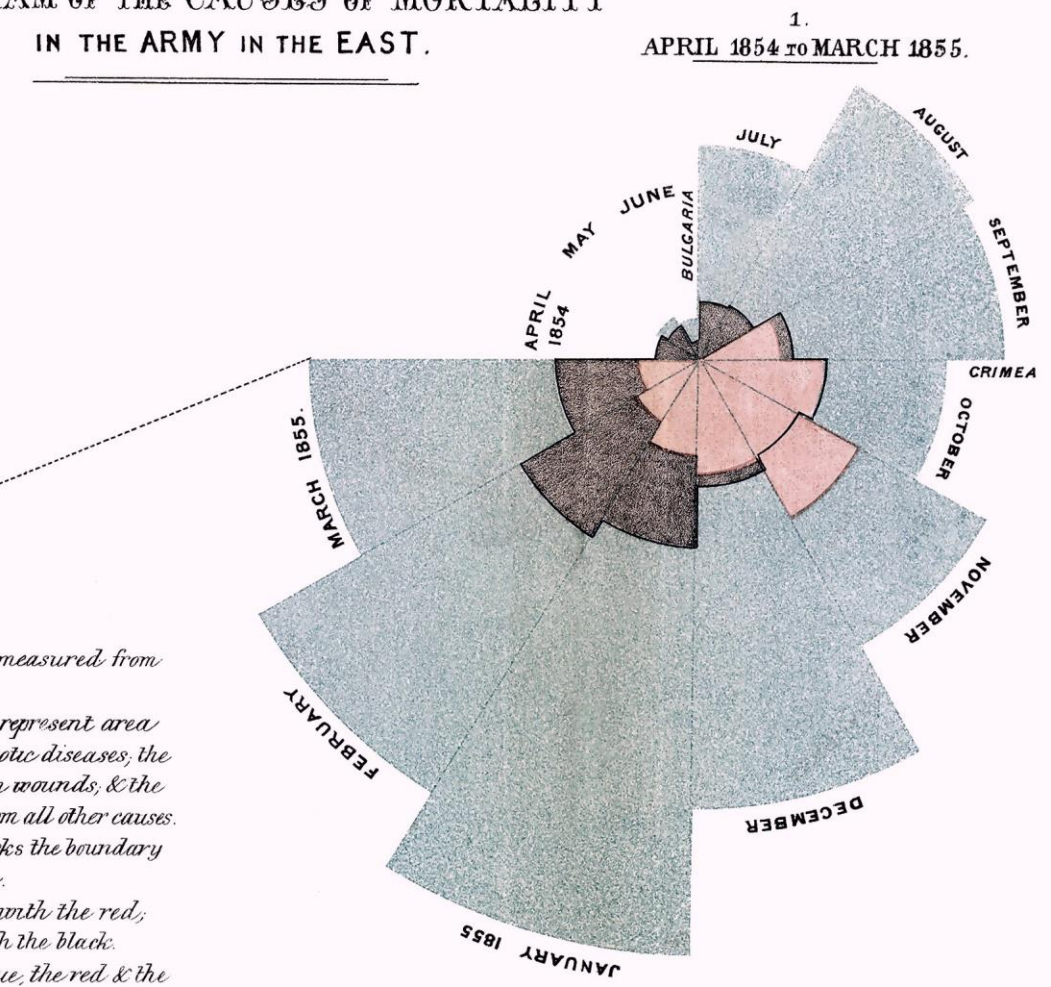
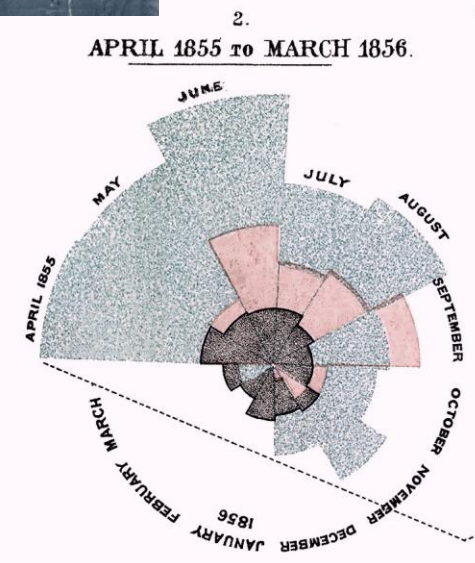
Exemple : La carte figurative des pertes successives en hommes de l'armée française dans la campagne de Russie 1812-1813
 Charles Minard, 1869



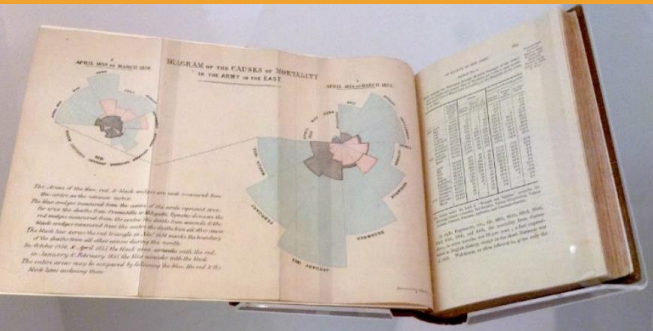
"Diagram of the causes of mortality in the army in the East", Florence Nightingale (1820-1910), 1858



DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.
 The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.
 The black line across the red triangle in Nov^r 1854 marks the boundary of the deaths from all other causes during the month.
 In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.
 The entire areas may be compared by following the blue, the red & the black lines enclosing them.





Explorer les données par la visualisation

Pourquoi ?

Méthodes et outils

Explorer les données par la visualisation

POURQUOI VISUALISER ?

Les avantages de la visualisation

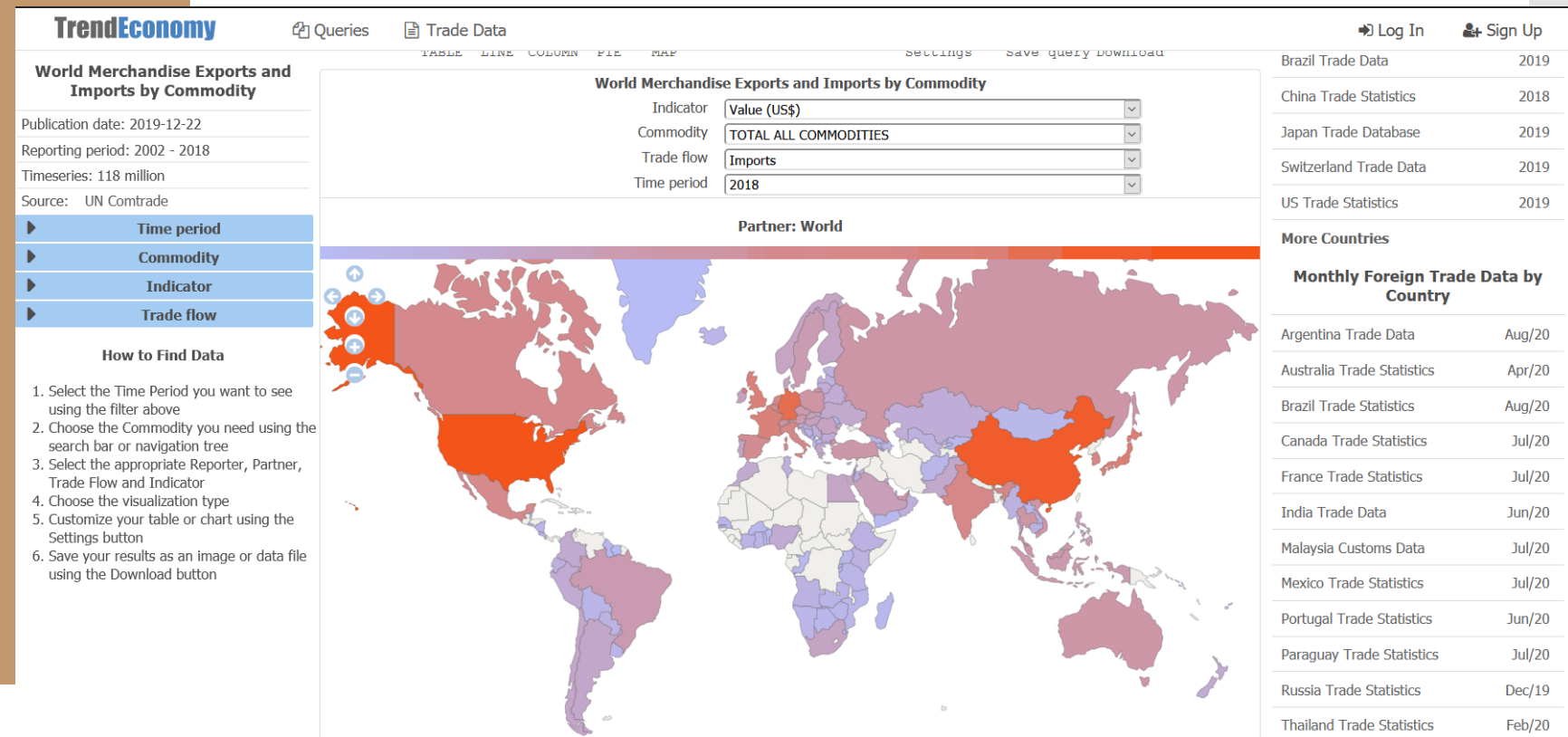
(C. Ware, 2000; 2005) :

- Manipuler des structures complexes ;
- Percevoir l'émergence de propriétés « inconnues » ;
- Mettre en évidence des problèmes dans les données, dans leur collecte ;
- Percevoir simultanément des propriétés à grande et à petite échelle sur les données ;
- Acquérir de nouvelles connaissances ;
- Faciliter la formation d'hypothèses sur les données ;
- Mieux identifier et comprendre les relations entre elles ;
- Prendre des décisions/transmettre l'information ;

Explorer les données par la visualisation

EXPLORER PAR LA VISUALISATION POUR EXTRAIRE DE L'INFORMATION

TrendEconomy



Explorer les données par la visualisation

EXPLORER PAR LA
VISUALISATION POUR EXTRAIRE
DE L'INFORMATION

Exploring Trade Data

A place to showcase innovative and experimental uses of UN Comtrade data. Explore visualizations of huge volume of data and metadata, cutting-edge data extraction tools, and alternative dissemination platforms.

Les outils de visualisation et d'exploration interactive offrent au chercheur des moyens plus complets pour la compréhension des différentes dimensions de ces mondes.

- **La dimension spatiale** : extension, diffusion, échelles imbriquées...
- **La dimension d'organisation résiliable** : hiérarchies, réseaux, centralités...
- **La dimension temporelle** : évolution, construction, diffusion...

(M. Maisonobe, L. Jegou, 2016)

Explorer les données par la visualisation

EXPLORER EN

VISUALISANT NOS DONNÉES

L'exploration visuelle permet de ré-investir des méthodes connues :

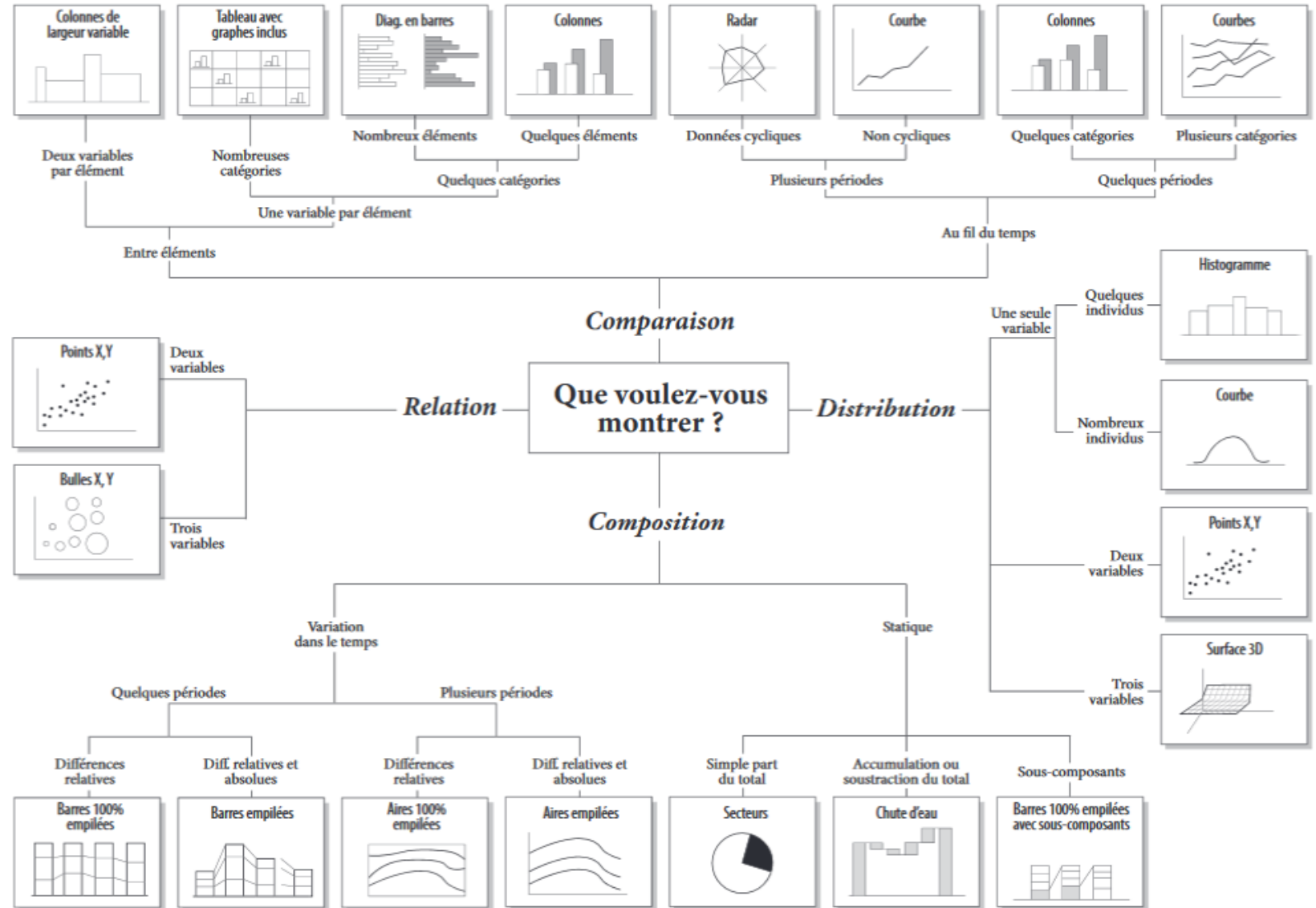
- L'analyse statistique des données (Tukey, 1977):
 - Par réduction des dimensions : ACP, AFC, etc.)
 - Par classification : CAH, etc
 - Par régressions
- Le choix d'un graphique adapté à ce que l'on souhaite montrer
- La compréhension des règles de construction visuelle (J. Bertin, 1977) : **la sémiologie graphique.**

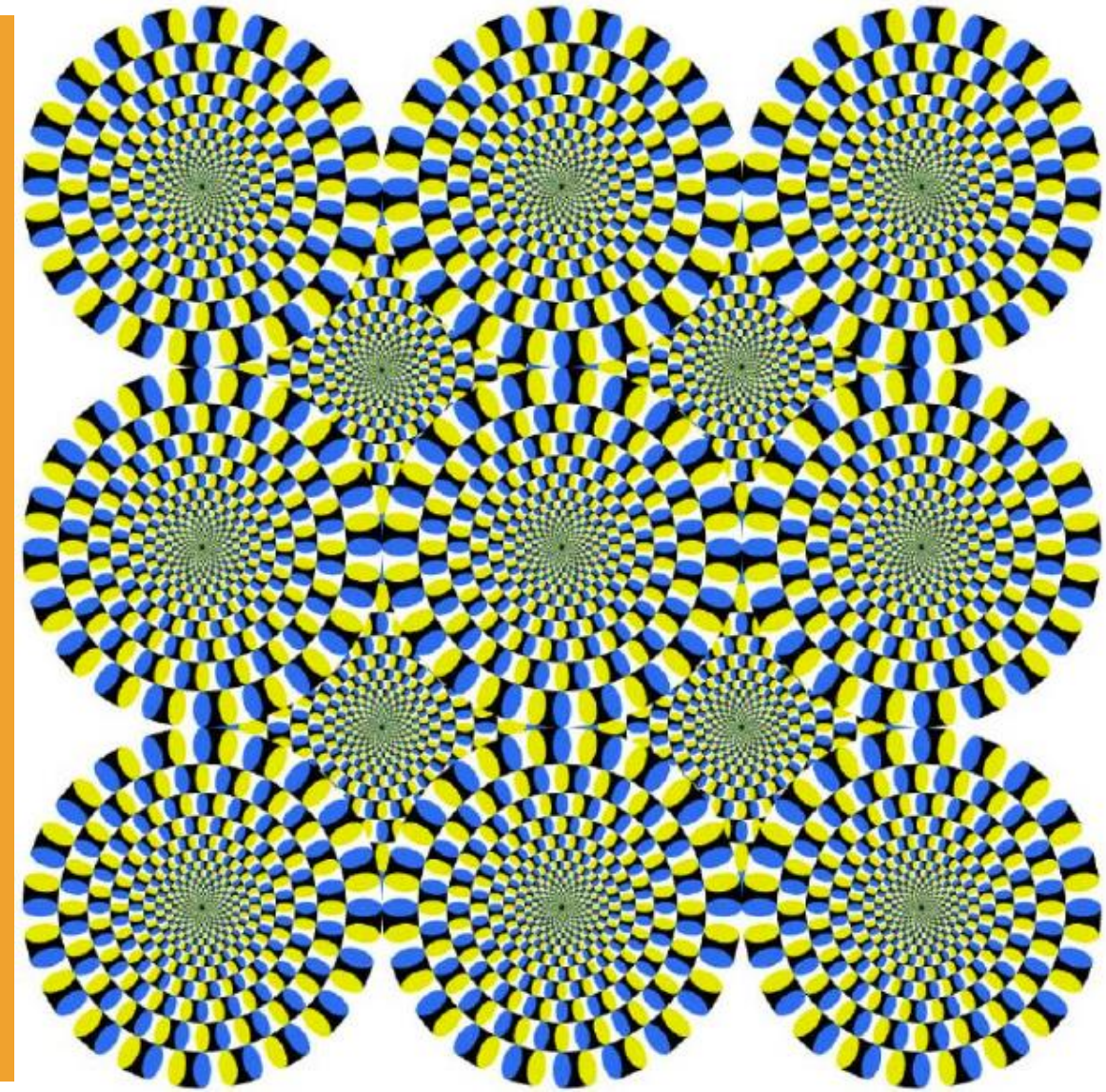
Cela implique alors :

Plusieurs **allers-retours, plusieurs manipulations/itérations** en respectant les règles statistiques et ...graphiques !

Choix
du graphique
le plus adapté
au jeu de données
et à la problématique

Arbre de décision des graphiques





La perception visuelle :
un « sens actif et
puissant »

Comment lit-on un graphique ?

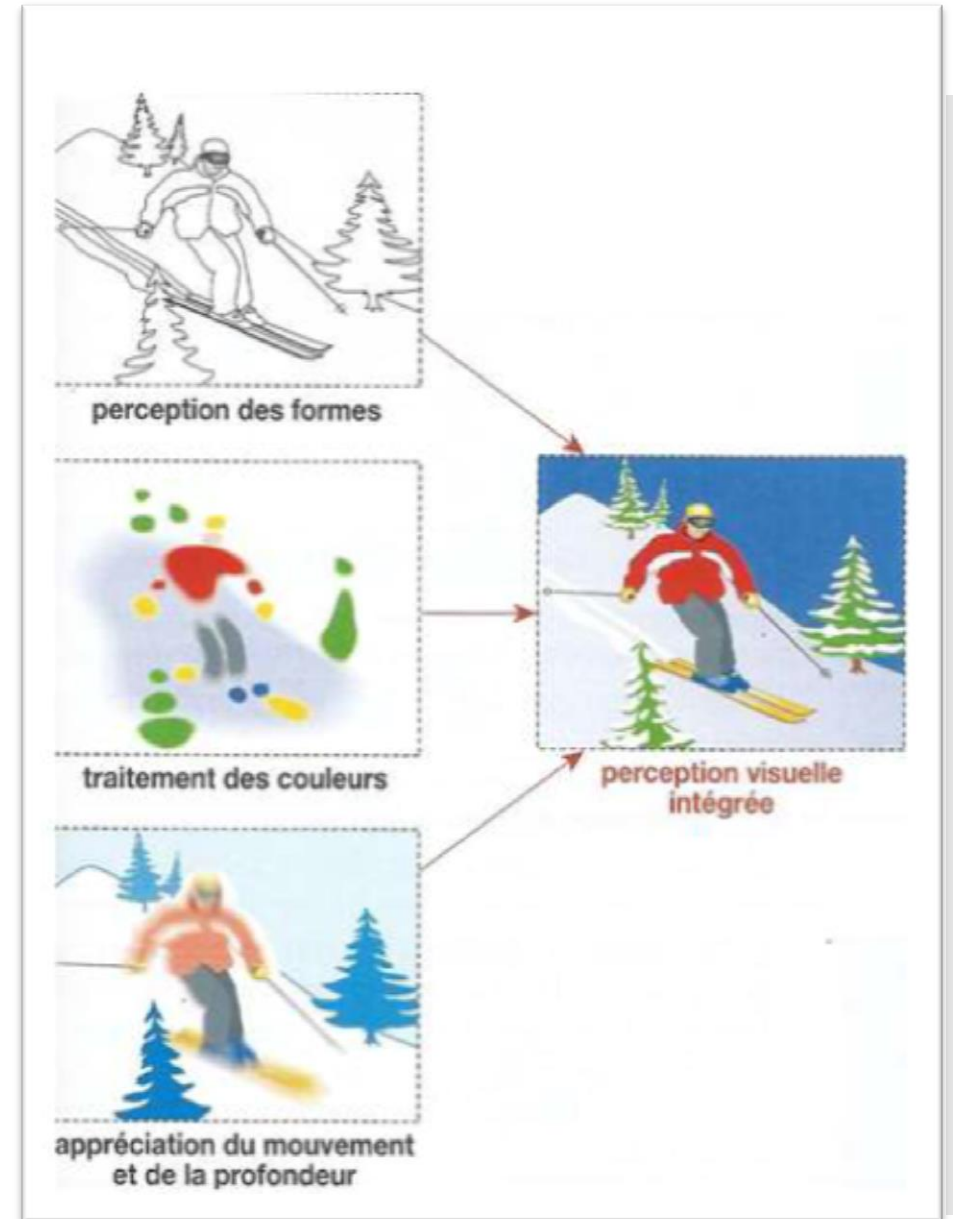


- ✓ **Une assimilation du cerveau quasi instantanée** et sans effort d'un grand nombre d'informations représentées graphiquement.
- ✓ Il est fortement influencé par certaines **variations visuelles**, car elles lui permettent de comprendre son environnement.
- ✓ Travail cognitif facilité par la représentation : **Un graphique est efficace lorsque la représentation est bien choisie.**

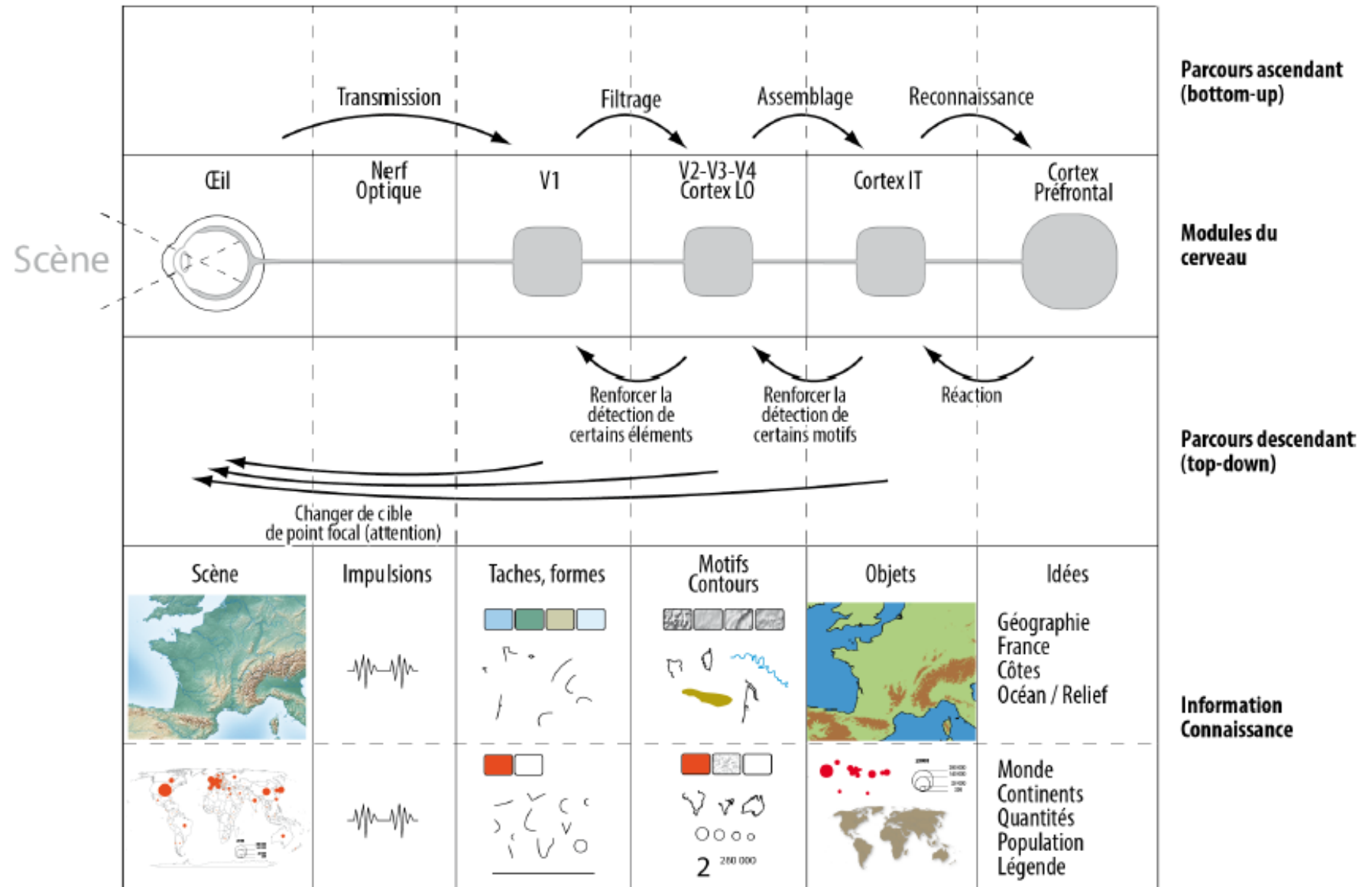
La perception visuelle

Ce qu'il faut savoir de la **perception visuelle** :

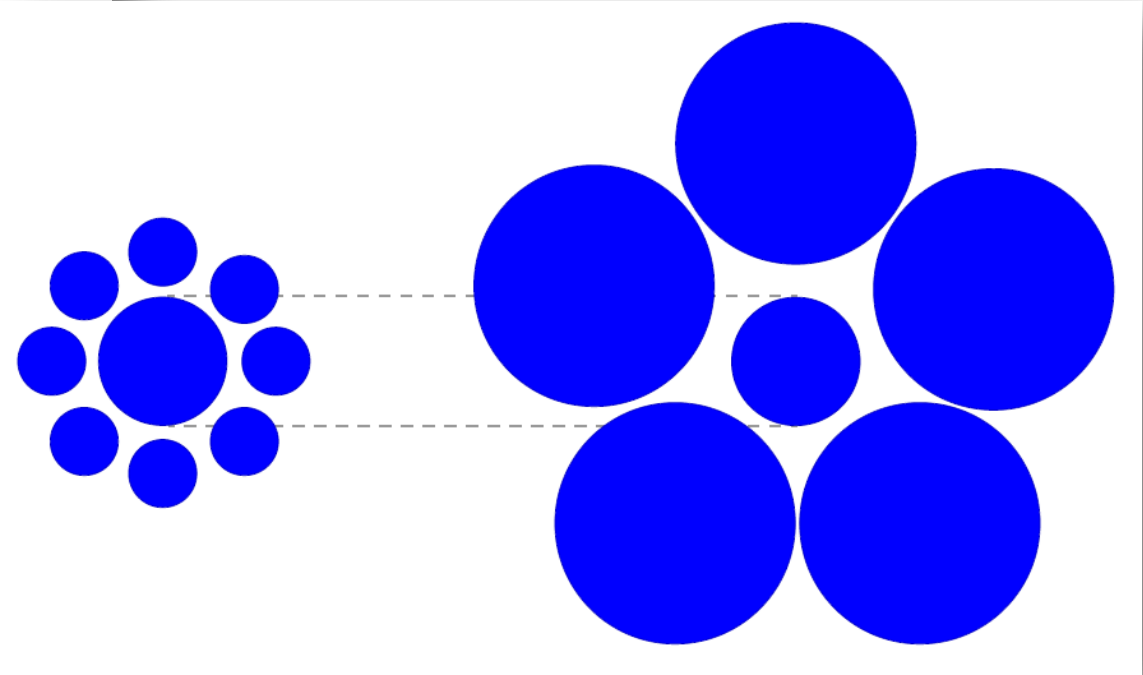
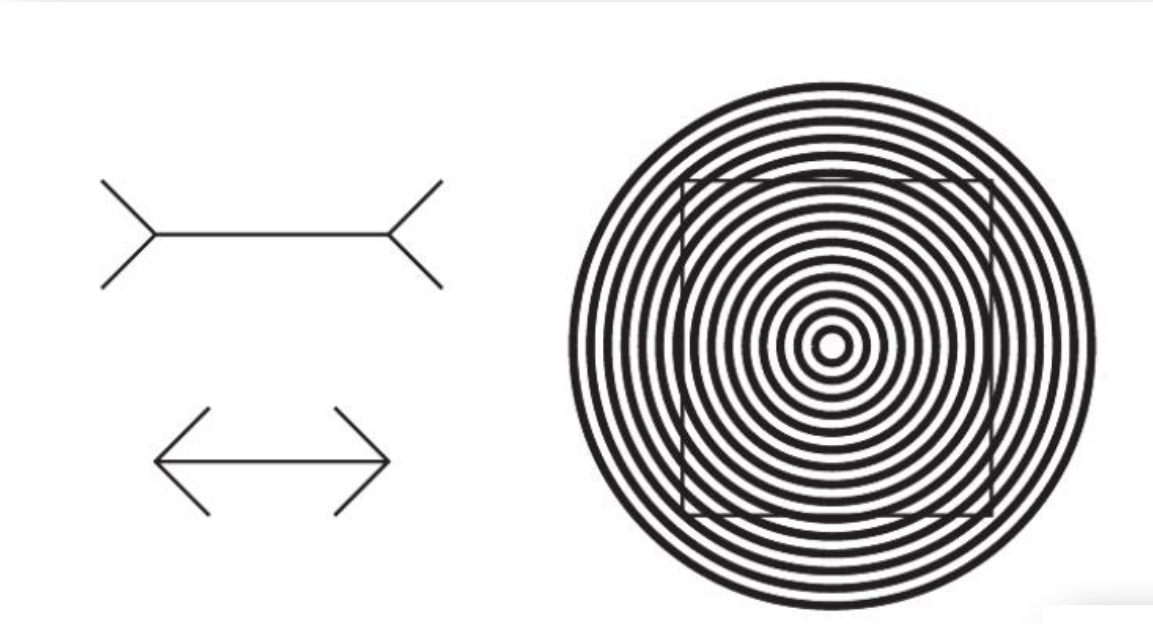
- Très partielle, incomplète et imprécise
- **Processus systémique**, pas instantané comme une photographie
- Joue sur l'expérience et la **cognition**
- Certaines formes visuelles sont parfois **détectées instantanément**



Le parcours de la vision

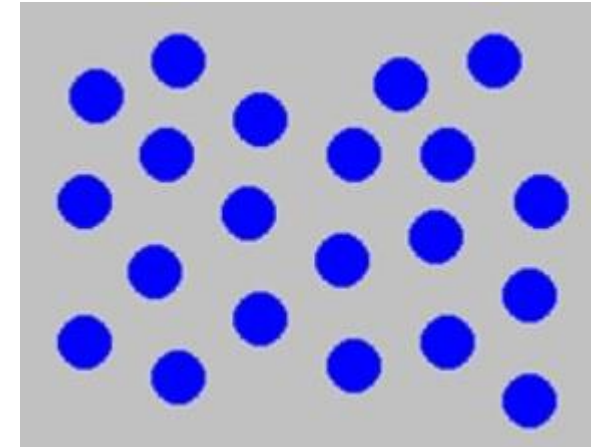
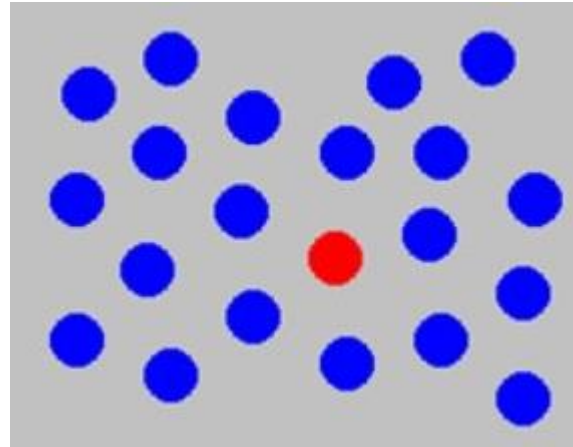


Illusion de Muller-Lyer

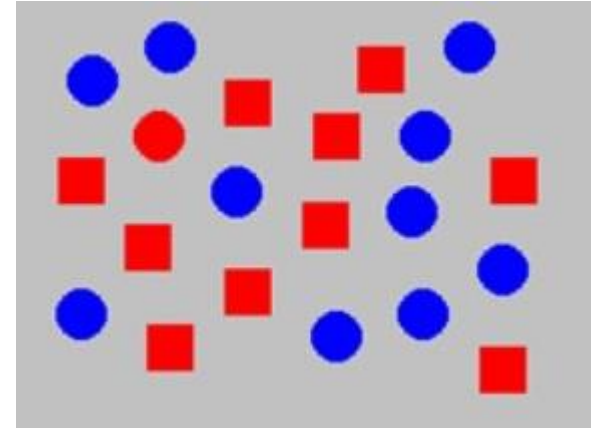
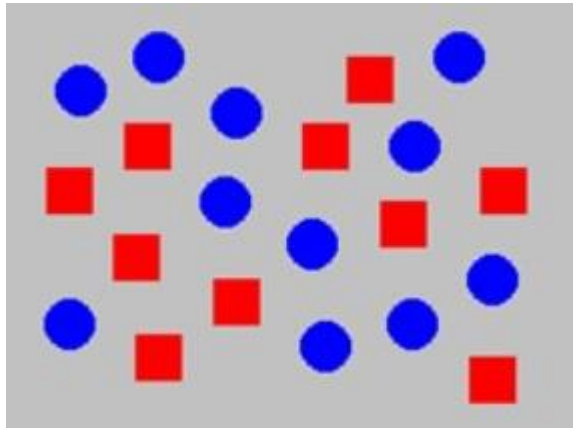


Par exemple

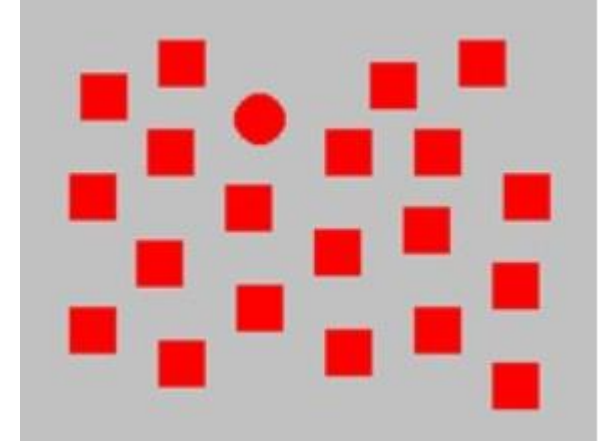
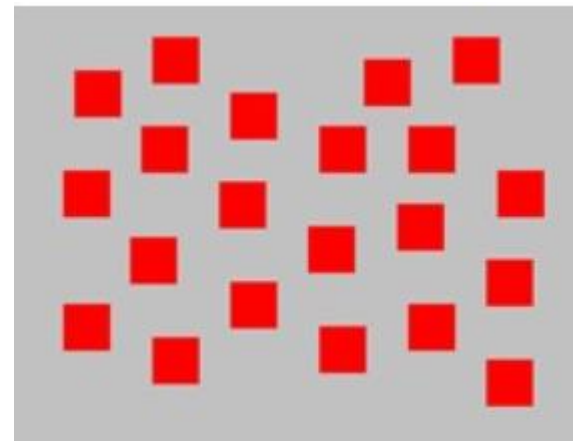
La teinte



Teinte et forme



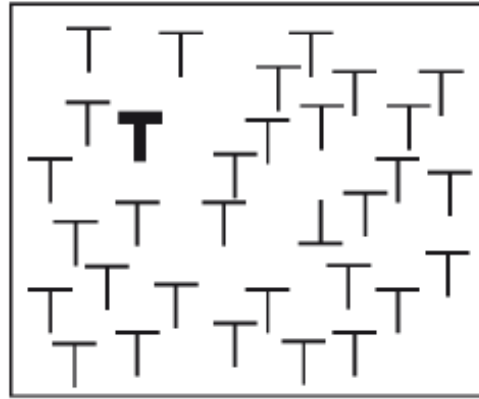
La forme



Par exemple

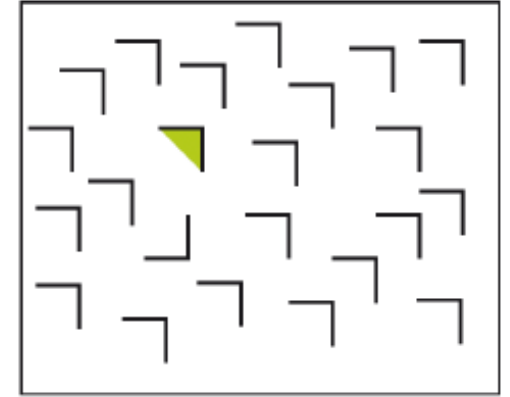
┌
difficile

┐
facile



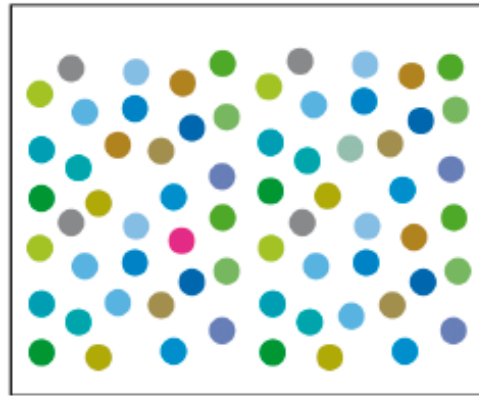
└
difficile

┘
facile



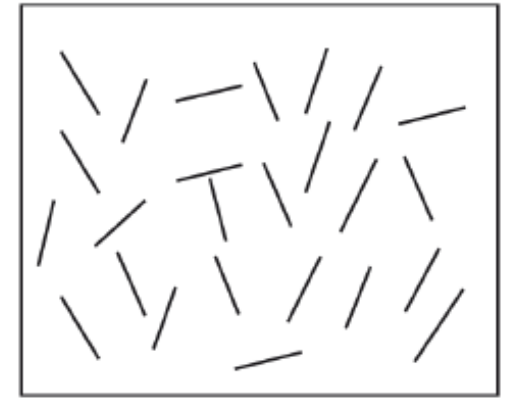
●
difficile

●
facile



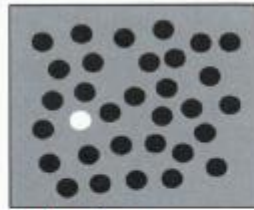
／
difficile

└
facile

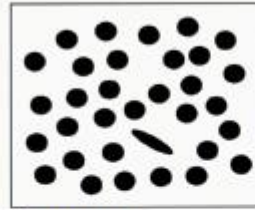


Traduit par L. Jegou, d'après C. Ware

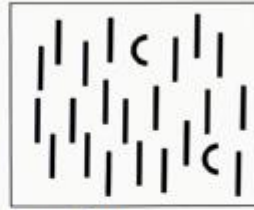
L'effet "pop", les éléments basiques de la perception visuelle



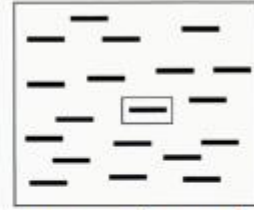
Valeur de gris



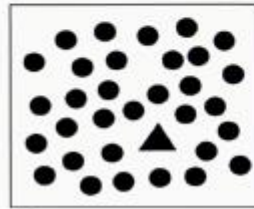
Étirement



Courbure



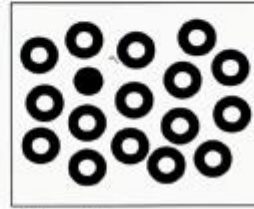
Présence d'un encadré



Forme



Surlignage



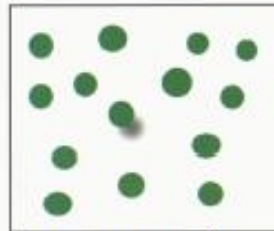
Remplissage



Netteté

Les éléments saillants, qui sautent au yeux par rapport au contexte/fond

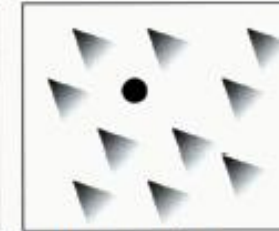
Autres exemples



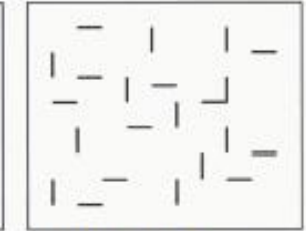
Projette une ombre



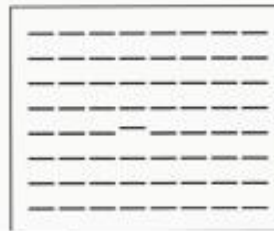
Convexité / concavité



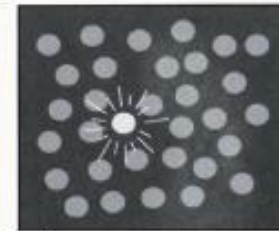
Coins anguleux



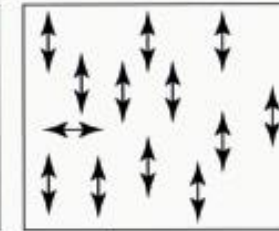
Lignes jointives



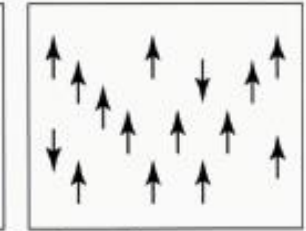
Désalignement



Clignotement



Sens du mouvement

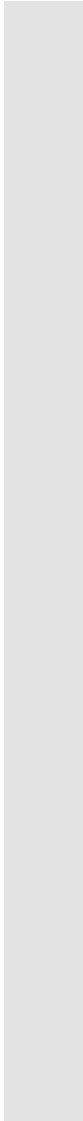


Phase du mouvement



Représenter correctement l'information

Bases de la représentation des données et
de la sémiologie graphique



Principes généraux de représentation des données

- **Objectif** : simplifier pour rendre l'information plus lisible tout en respectant la donnée d'origine
- Dans le cas d'une variable quantitative continue : Il faut **réduire** le nombre de cas possibles (de valeurs de la variable) à quelques-uns seulement, pour avoir moins de couleurs différentes sur le graphique.

➡ En statistique, on parle de « réaliser un découpage en classes » ou en cartographie, on dit « discrétiser une variable ».

Principes généraux de représentation des données

L'image colorée que représente la visualisation doit donner une idée correcte et pertinente de l'information contenue dans la variable.

- Il ne doit pas y avoir de classe vide.
- Toutes les classes doivent être jointives (pas de manques entre deux classes successives).
- Les classes ne doivent pas se chevaucher (un individu ne peut appartenir qu'à une classe à la fois).
- Les couleurs choisies doivent évoquer un phénomène.

Découpage en classes : exemple

En statistique

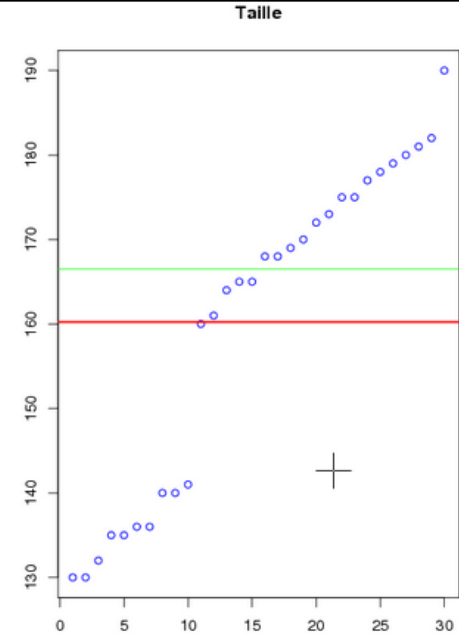
ID	TAILLE
P01	130
P02	140
G01	170
G02	160
P03	136
G03	165
P04	130
P05	135
P06	140
P07	135
G04	161
P08	136
G05	180
G06	190
P09	141
P10	132
G07	165
G08	168
G09	182
G10	177
G11	172
G12	168
G13	175
G14	181
G15	173
G16	169
G17	178
G18	179
G19	175
G20	164

Caractéristiques de la variable Taille

Taille	30 individus
Moyenne	160.233 cm
Ecart-type	18.693 cm
Coef. de variation	12 %
1er Quartile	140 cm
Médiane	166.5 cm
3eme Quartile	175 cm
Minimum	130 cm
Maximum	190 cm

Diagramme en tige et feuilles

```
130 | 0025566
140 | 001
150 |
160 | 01455889
170 | 02355789
180 | 012
190 | 0
```



Classe 1 : [130,160]

Classe 2 :]160,190]



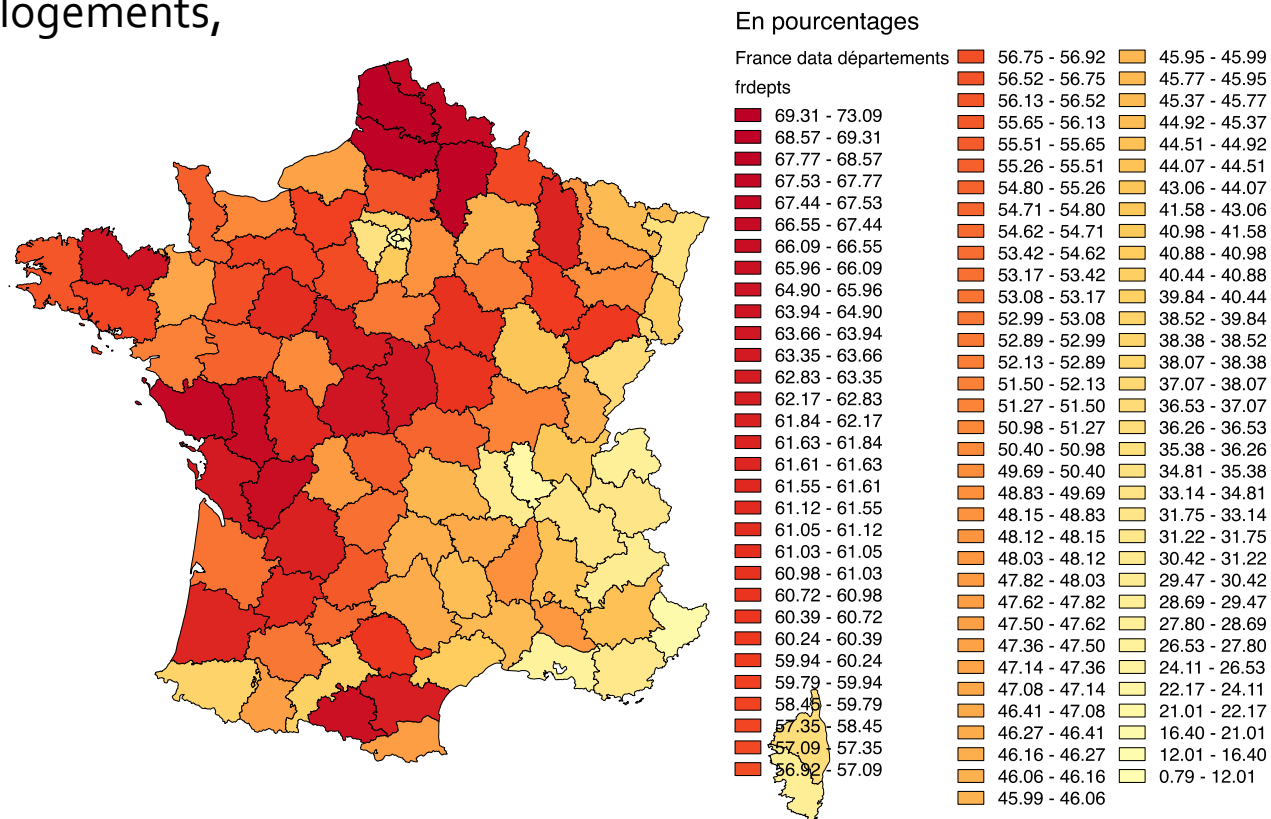
Numéro	Classe 1	Classe 2
01	P01 130	G04 161
02	P04 130	G20 164
03	P10 132	G03 165
04	P05 135	G07 165
05	P07 135	G08 168
06	P03 136	G12 168
07	P08 136	G16 169
08	P02 140	G01 170
09	P06 140	G11 172
10	P09 141	G15 173
11	G02 160	G13 175
12		G19 175
13		G10 177
14		G17 178
15		G18 179
16		G05 180
17		G14 181
18		G09 182
19		G06 190

Découpage en classes exemple

En cartographie

La situation la plus fréquente, en cartographie, est celle de la représentation d'une variable dont l'effectif est trop grand pour être représenté de manière exhaustive.

Part des maisons individuelles dans le total des logements, en 1990

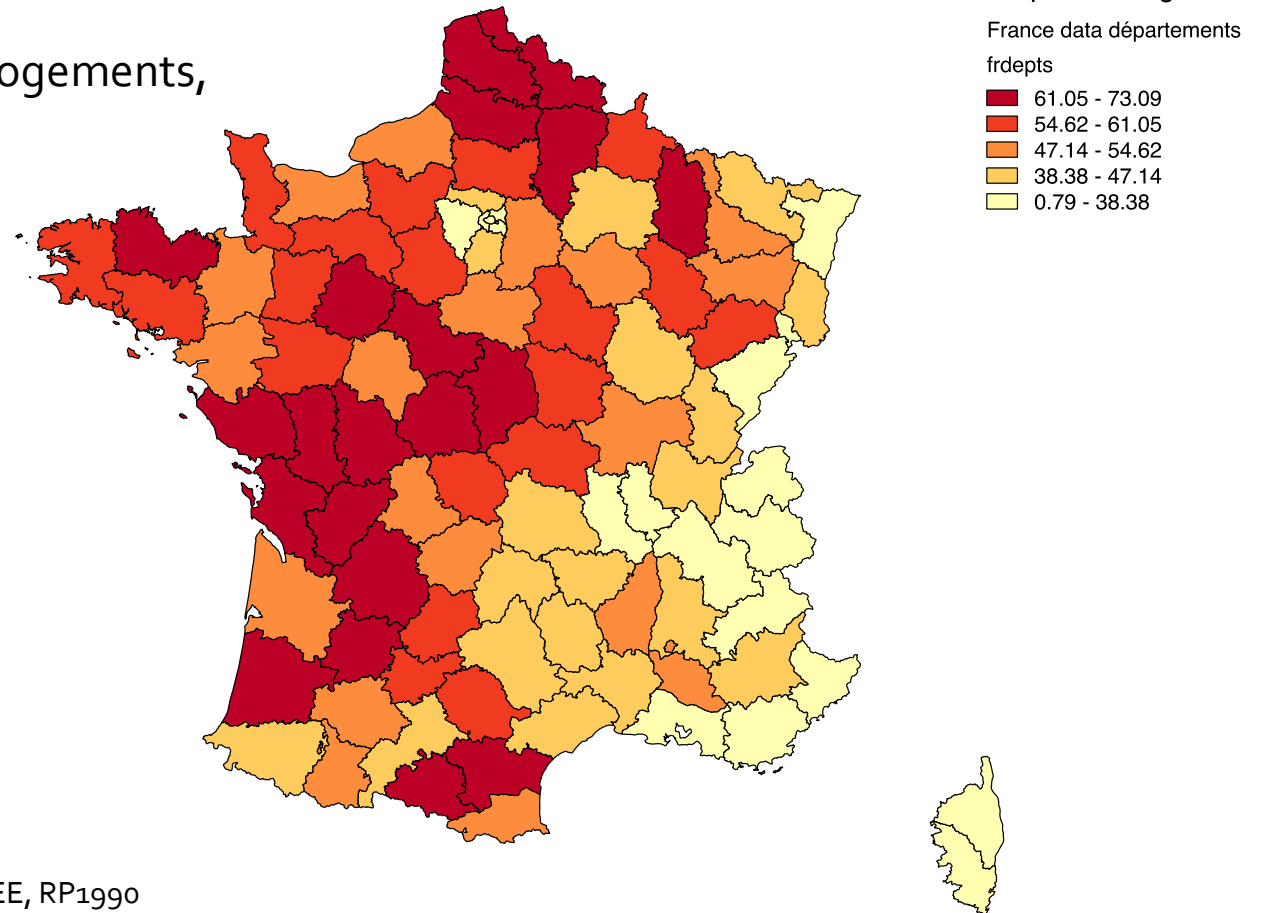


Découpage en classes exemple

En cartographie

La situation la plus fréquente, en cartographie, est celle de la représentation d'une variable dont l'effectif est trop grand pour être représenté de manière exhaustive.

Part des maisons individuelles dans le total des logements, en 1990



Source : INSEE, RP1990

Méthodes de choix du nombre de classes

- Méthode de Huntsbeerger :

$$H = 1 + 3,3 * \log_{10}(n)$$

- Méthode de Brooks-Carruthers :

$$BC \leq 5 * \log_{10}(n)$$

Avec $n = \text{nb. individus}$

Méthode	Classes pour 100 Individus	Classes pour 22 individus
Huntsbeerger	7,6	5,4
Brooks- Carruthers	10	6,7

Le choix du nombre de classes

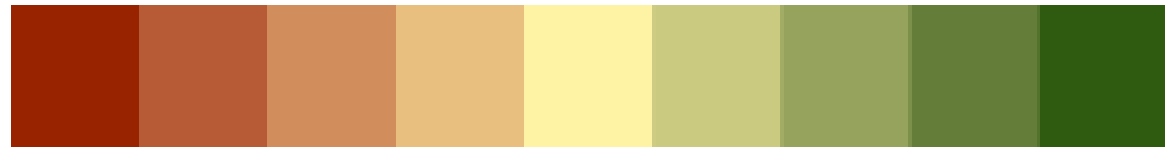
Le nombre *maximal* de classes que l'on peut réaliser en préservant la lisibilité de la carte ou du graphique en couleurs peut varier entre 7 et 9.

Attention à ce que la qualité d'impression soit adaptée.

7 classes si le dégradé de couleurs se réalise sur une seule teinte :



9 classes s'il fait jouer deux teintes :



Des outils pour créer des gradients de couleurs

(L. Jegou, 2014):

- Sur une seule variation :

<http://geotests.net/couleurs/gradients.html>

- Avec inflexion (changement de teinte) :

http://geotests.net/couleurs/gradients_inflex.html

Comment choisir les bornes de ces classes ?

- La **méthode de Jenks (« seuils naturels »)** -> respecte les ruptures dans la distribution
- Méthodes statistiques :
 - La **méthode des quantiles** ou discrétisation par équifréquence (ou équipopulation)
 - La **méthode des moyennes emboîtées**
 - La **méthode « standardisée »** ou l'écart à la moyenne basé sur l'écart-type
- Méthodes mathématiques :
 - La **méthode par équidistance** ou discrétisation en classe d'amplitude égale
 - La **méthode de la progression arithmétique**
 - La **méthode de la progression géométrique**
- La **méthode manuelle (découpage externe)**

Géovisualisation des discrétisations (L.Jegou, 2017)

<http://mappemonde.mgm.fr/geovisu/119/>

La sémiologie graphique

C'est « l'ensemble des règles qui permettent l'usage d'un système de signes graphiques pour transmettre l'information ».

(J.Bertin, *Sémiologie graphique*, 1967)

L'usage de la sémiologie graphique permet une représentation visuelle plus efficace et améliore la lisibilité d'un contenu graphique.

Une utilisation correcte de la sémiologie graphique permet :

- ✓ *La transcription, dans le système graphique d'un signe, d'un ensemble de données.*
- ✓ *La traitement de ces données afin de faire apparaitre l'information d'ensemble recherchée*
- ✓ *La construction d'images les mieux adaptées pour communiquer/ comprendre.*

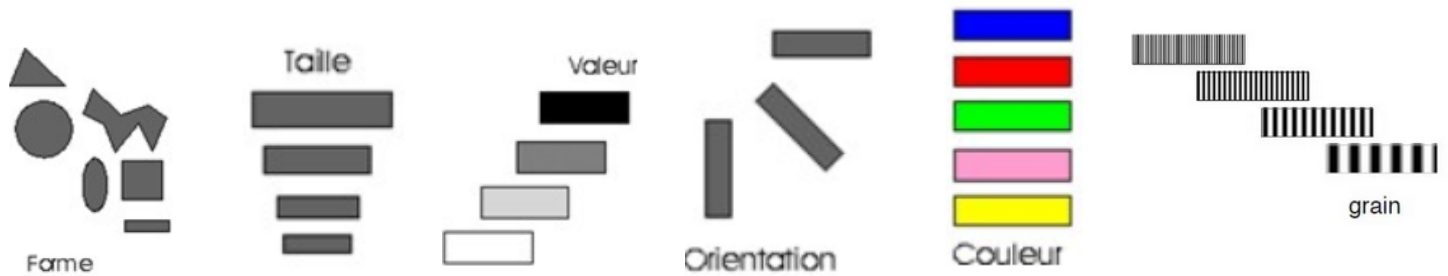
La sémiologie graphique : un langage de communication

- Composantes du langage graphique :

- **Alphabet** = signes graphiques élémentaires :
point, trait, surface



- **Vocabulaire** = variables visuelles (façon de faire varier les figurés)

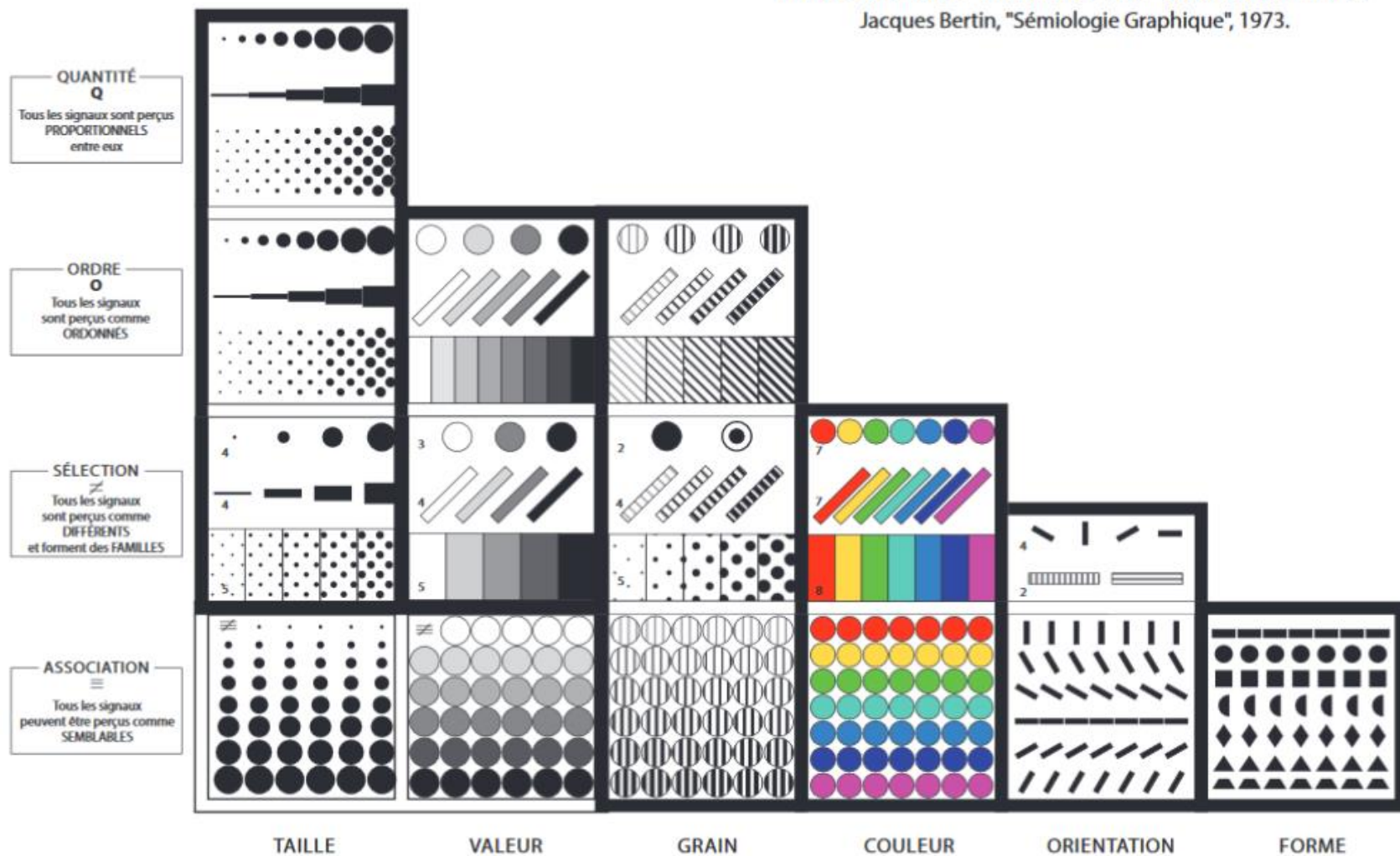


- ... les règles de la perception visuelle définissant la syntaxe (le style)

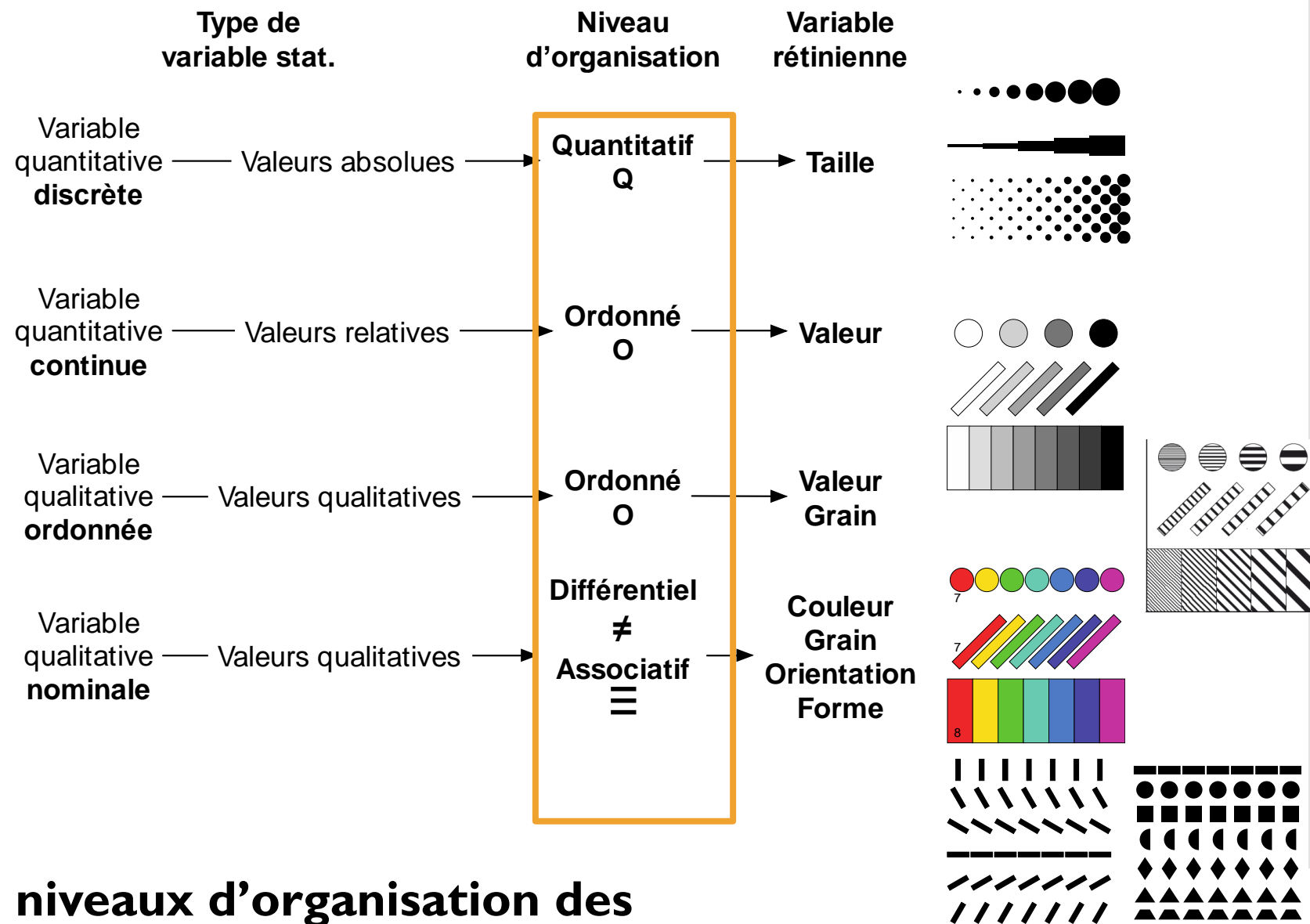
Choix sémiologique en fonction de la variable statistique

NIVEAU DES VARIABLES RÉTINIENNES

Jacques Bertin, "Sémiologie Graphique", 1973.

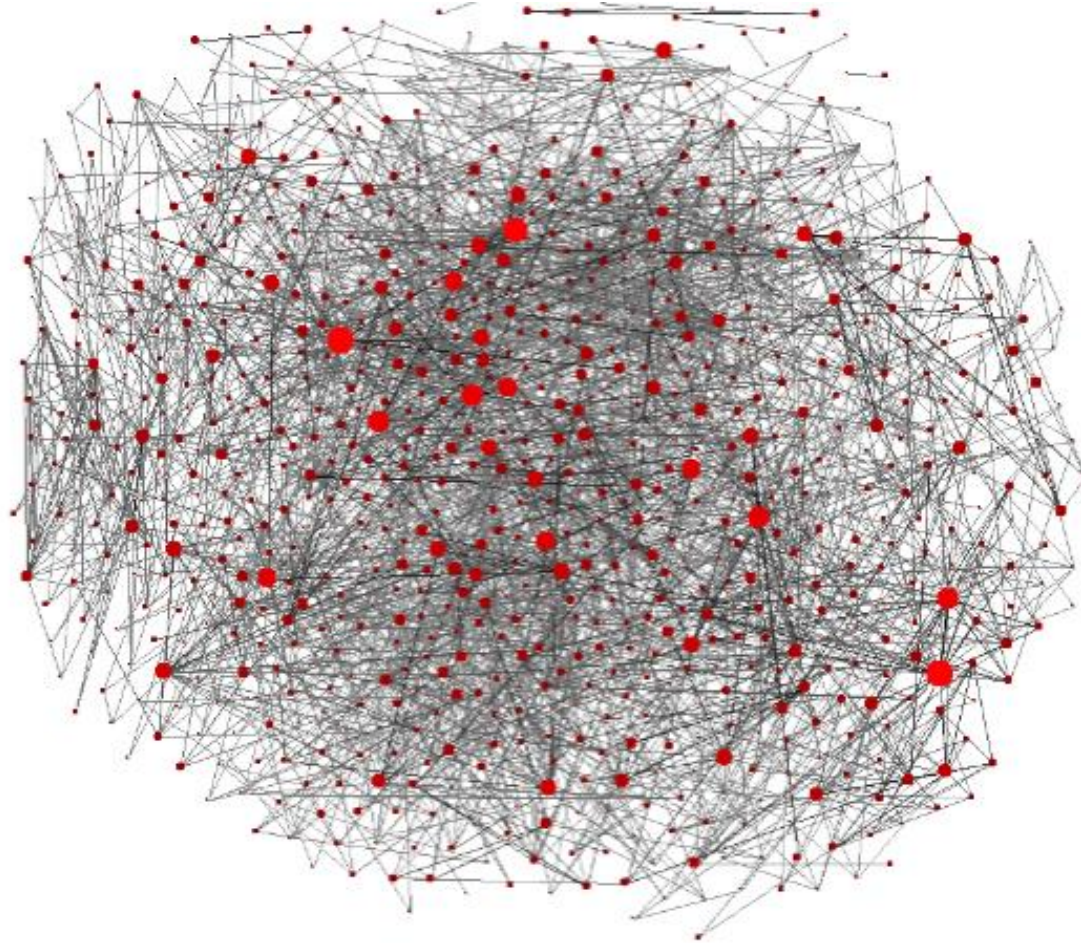


Choix sémiologique en fonction de la variable statistique



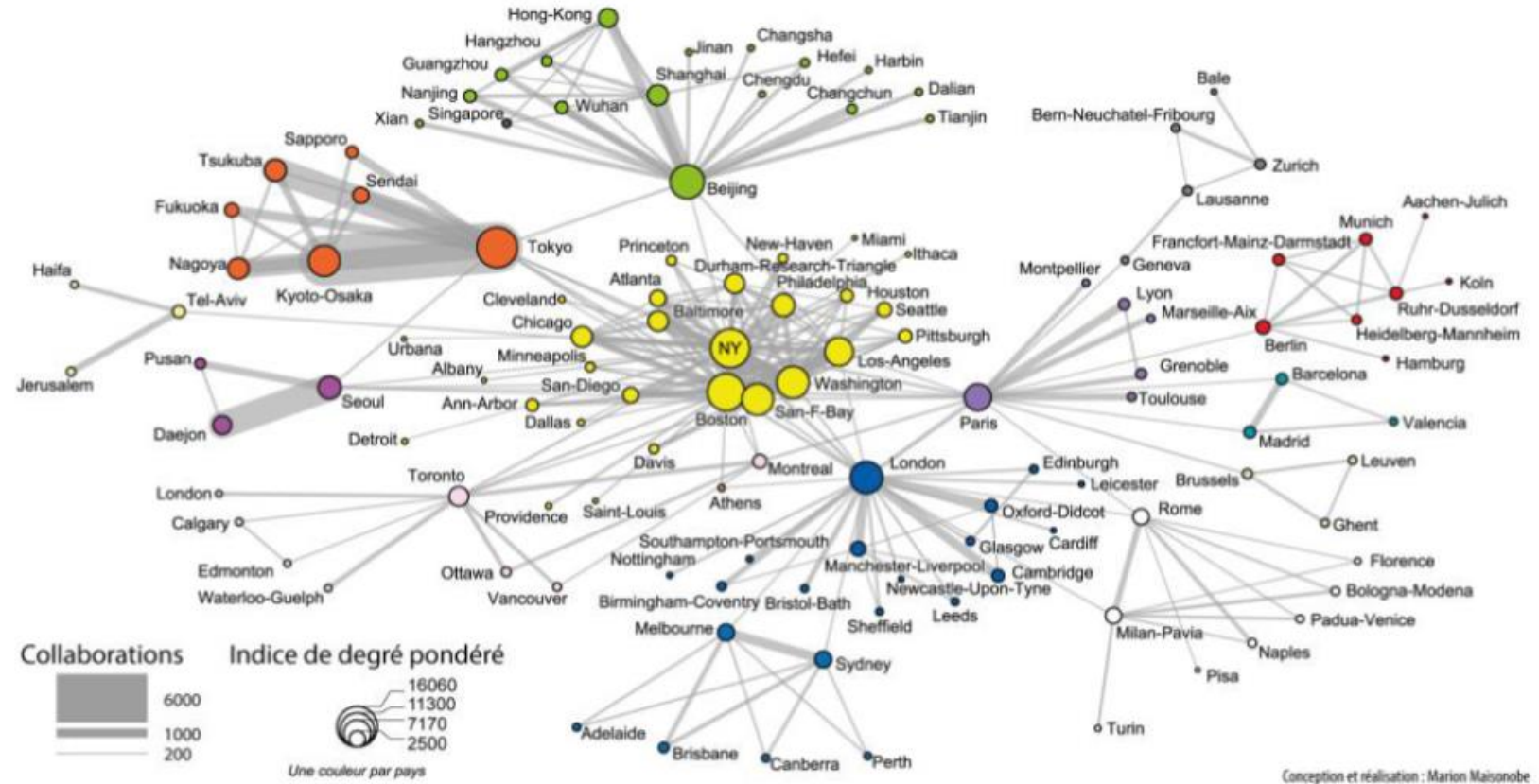
4 niveaux d'organisation des variables à représenter

L'enjeu d'une bonne visualisation



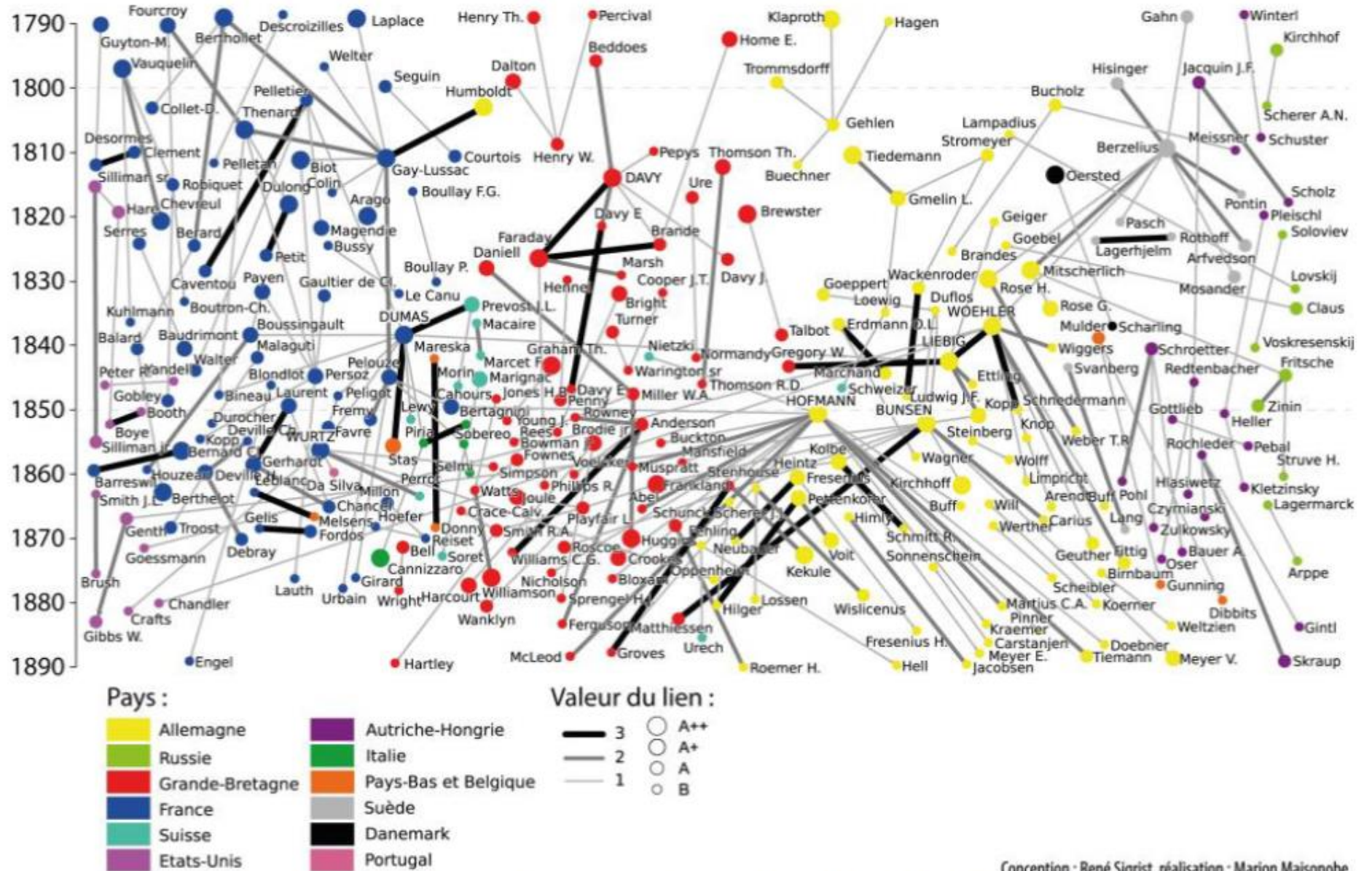
Graphe des réseaux de co-auteurs avec au moins 2 publications internationales et 2 co-publications entre les co-auteurs, 2009-2015

L'enjeu d'une bonne visualisation



Graphe n°4 : Composante principale du réseau
des collaborations scientifiques mondiales en 2007

L'enjeu d'une bonne visualisation



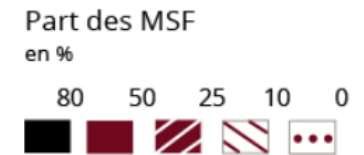
Conception : René Sigrist, réalisation : Marion Maisonobe

Graph n°3 : Liens de coopération entre chimistes (1810-1860)

L'enjeu d'une bonne visualisation

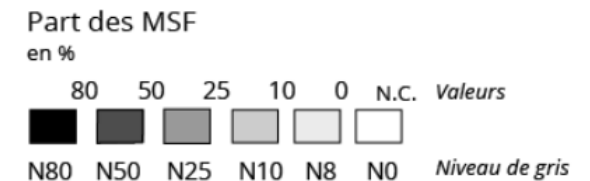


Légende d'origine



MSF : Mutilations sexuelles féminines.

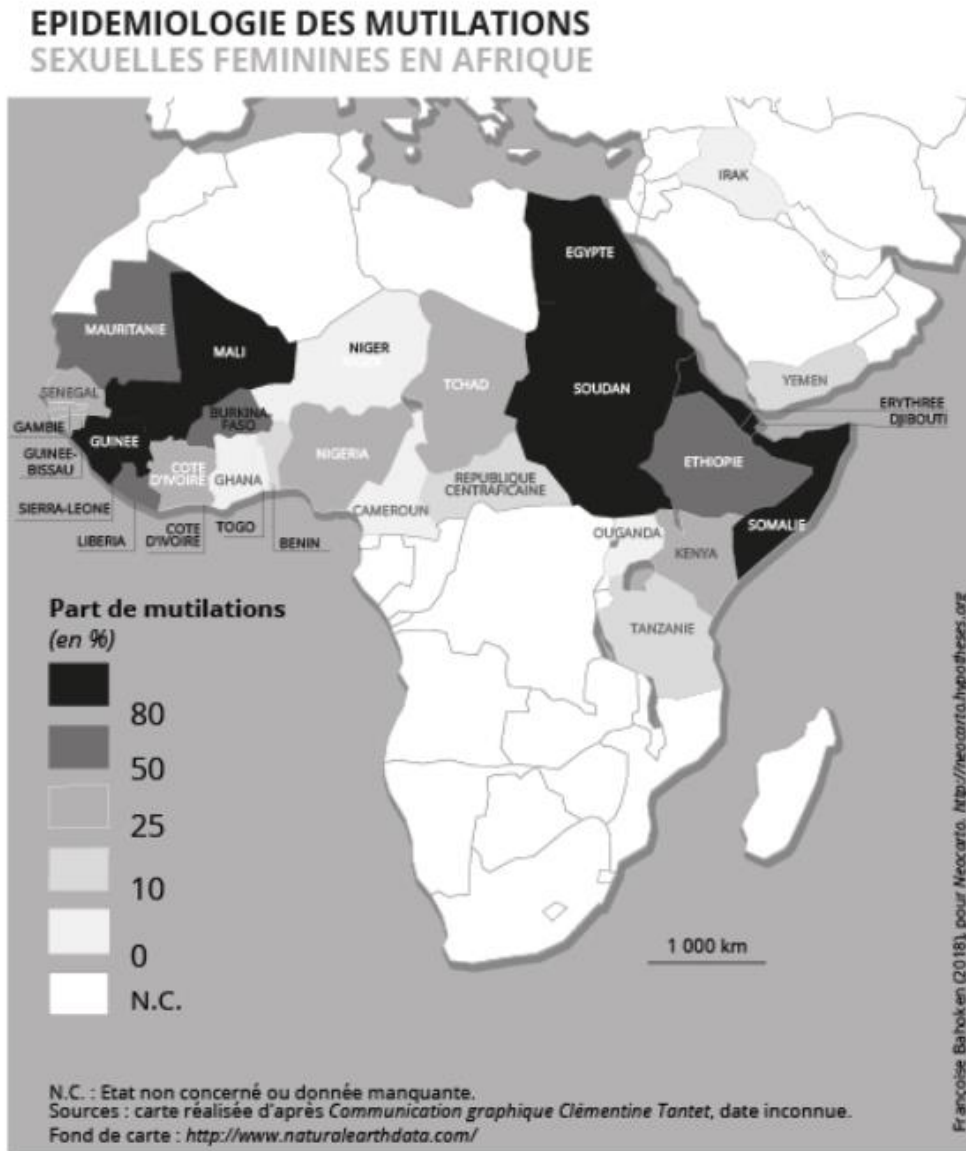
Modification de la légende



N. C. : Non Concerné
(information inexistante, manquante, ou non disponible)

Espace colorimétrique (CMJN)
(Cyan, Magenta, Jaune, Noir)
C,M,J = 0 % ; N=x %

L'enjeu d'une
bonne
visualisation



<https://neocarto.hypotheses.org/3820> (F. Bahoken, 2018)

Pour conclure :

Représenter
correctement
pour mieux
explorer

- En amont, il existe des possibilités de data-visualisation
-> au service de l'exploration/compréhension d'un corpus
- Devant les données, le mapping et la représentation graphique
-> l'occasion de simplifier et de composer l'information pour la rendre lisible
- Respecter la forme de la **distribution** des valeurs
(celle du *graphique de fréquences*)
- Créer des **classes** qui correspondent à des « groupes naturels » de valeurs (sous-ensembles).
- Utiliser les **variables rétinienne**s adaptées aux types de variables statistiques à représenter.

A retenir :

Plus une image est complexe à percevoir,
plus elle sera longue et difficile à lire