

Cours: VISUALISATION DE DONNEES

Master MIV, 2022/2023

Prof. Slimane Larabi

VISUALISATION DE DONNEES

SOMMAIRE DU COURS:

Chapitre 1	Fondements de la Visualisation de Données
Chapitre 2	Techniques de Visualisation des Données Spatiales
Chapitre 3	Techniques de Visualisation des Données Géo Spatiales
Chapitre 4	Techniques de Visualisation d'Arbres et Graphes
Chapitre 5	Outils, algorithmes et techniques de visualisation de séries temporelles

VISUALISATION DE DONNEES

REFERENCES:

1. Visualization Analysis & Design, Tamara Munzner, U. British Columbia
2. The Visualization Handbook, Chuck Hansen, U. Utah, Chris Johnson, U. Utah
3. Information Visualization: Perception for Design, Colin Ware, U. New Hampshire
4. Readings in Information Visualization, Stuart Card, Xerox PARC, Jock Mackinlay, Xerox PARC, Ben Schneiderman, U. Maryland
5. The Visual Display of Quantitative Information, Edward Tufte, Yale
6. Visual Explanations, Edward Tufte, Yale
7. Envisioning Information, Edward Tufte, Yale
8. Interactive Data Visualization for the Web, Scott Murray, O'Reilly, March 2013 (ebook free)

VISUALISATION DE DONNEES

Objectif du cours:

Etudier les techniques et algorithmes pour la création de visualisations en se basant sur les principes du graphique, de l'art visuel, psychologie de perception et la science cognitive.

Transformer une grande quantité de données brutes en représentation graphique en exploitant la supériorité du traitement visuel de l'humain à détecter les motifs de sa capacité d'inférer. L'objectif est d'améliorer l'interprétation et la communication de données.

Alexandru C. Telea , **“Data Visualization -- Principles and Practice”**,
A K Peters, 2008

Population:

- Développer des applications pour la visualisation des big data,
- Développer des outils et systèmes de visualization des big data

CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

1.1 Définition de la visualisation

“Transformation of the symbolic into the geometric”

[McCormick et al. 1987]

“Finding the artificial memory that best supports our natural means of perception.”

[Bertin 1967]

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

• [John Tukey, 1977]

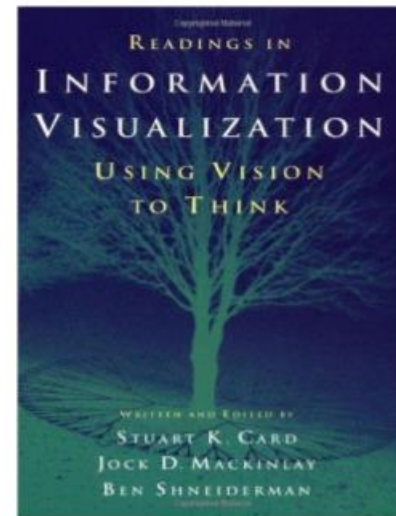
“The use of computer-generated, interactive, visual representations of data to amplify cognition.”

[Card, Mackinlay, & Shneiderman 1999]

CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

Visualisation d'information

*“L'utilisation de
représentation visuelles,
interactives et informatique
de données abstraites
pour amplifier la cognition.”*
Card, Mackinlay, & Shneiderman, 1999



1.2 Besoin de la visualisation de données

- **Croissance de production de données**

le volume de **données** générées dans le monde devrait dépasser 180 zettaoctets 10^{21}) à l'horizon 2025, soit une **croissance** annuelle moyenne de près de 40 % sur cinq ans.

<https://fr.statista.com/>

- La capacité de prendre des données et être capable de :
 - les comprendre,
 - les traiter,
 - extraire de l'information utile,
 - les visualiser,
 - les communiquer
- Ce sera une importante compétence à acquérir pour le futur en raison de la gratuité et l'omniprésence de données.

Pourquoi créer des visualisations?

- Répondre à des questions ou les découvrir
- Prendre décision
- Voir les données dans un contexte spécifique
- Etendre la mémoire
- Trouver des motifs
- Présenter des arguments ou raconter une histoire
- S'inspirer
- Communiquer l'information (partager, collaborer)

CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

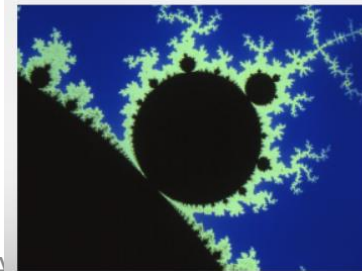
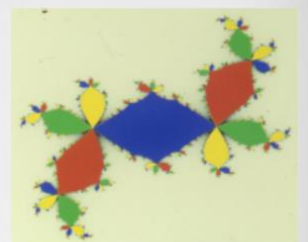
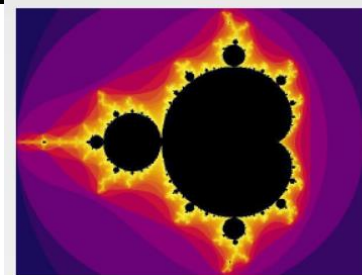
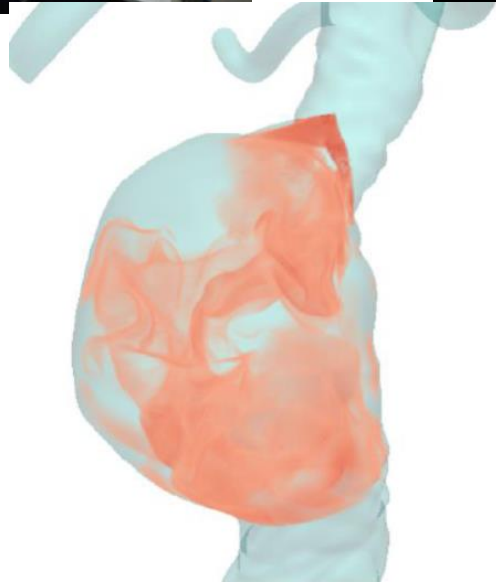
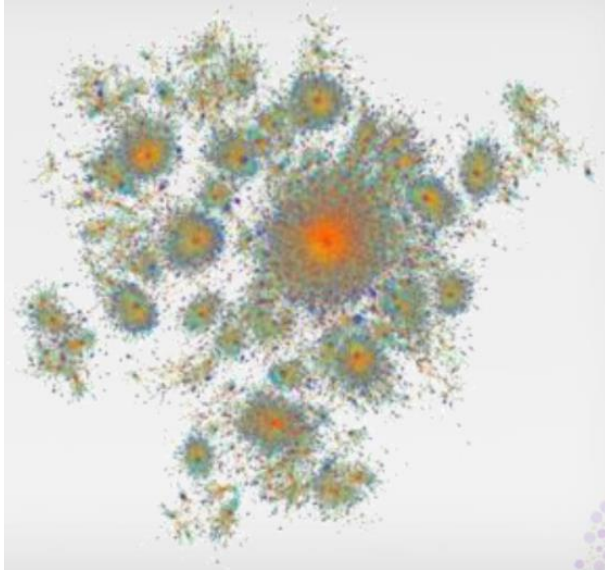
Domaines de la Visualisation

- Santé
- Politique
- Scientifique (Ingénierie, Mathématiques...)
- Finance (Suivi des performances des investissements)
- Logistique (Déterminer les meilleurs routes.)
- Systèmes d'Informations Géographiques

CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES



<http://www.war-tomms.blogspot.fr/>



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- **Quelques challenges:**
 - Données sur Réseaux en croissance
 - Texte et documents
 - Données images
 - Données géo-spatiales
 - Données hétérogènes

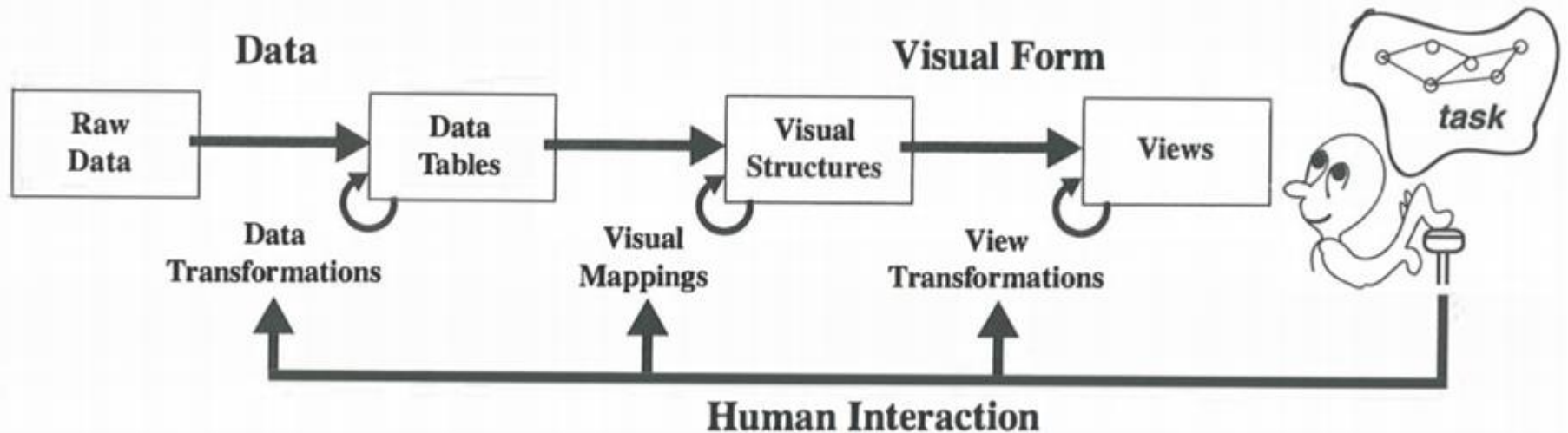
Techniques en relation avec la Visualisation

- Analyse de données
- Data mining
- Bases de données
- HCI (Human-Computer Interaction)
- Infographie (synthèse d'image)

Modes de la Visualisation

- Visualisation Interactive (web)
- Visualisation Présentation
- Interactive (Storytelling= Présentations via des pages web interactives)

1.3 Processus de la Visualisation



Visual Structures: the space in which we're going to create our visualization, marks, graphical properties

Views: Graphical parameters

Card, Mackinaly, Sheiderman, *Reading in Information Visualization: Using Vision to Think*, 199

1.4 Exemples de Visualisation



John Snow est un médecin britannique, pionnier dans les domaines de l'anesthésie, de l'hygiène et la santé publique. Ses travaux sur la propagation du choléra lui ont donné une place importante dans l'histoire de l'épidémiologie.



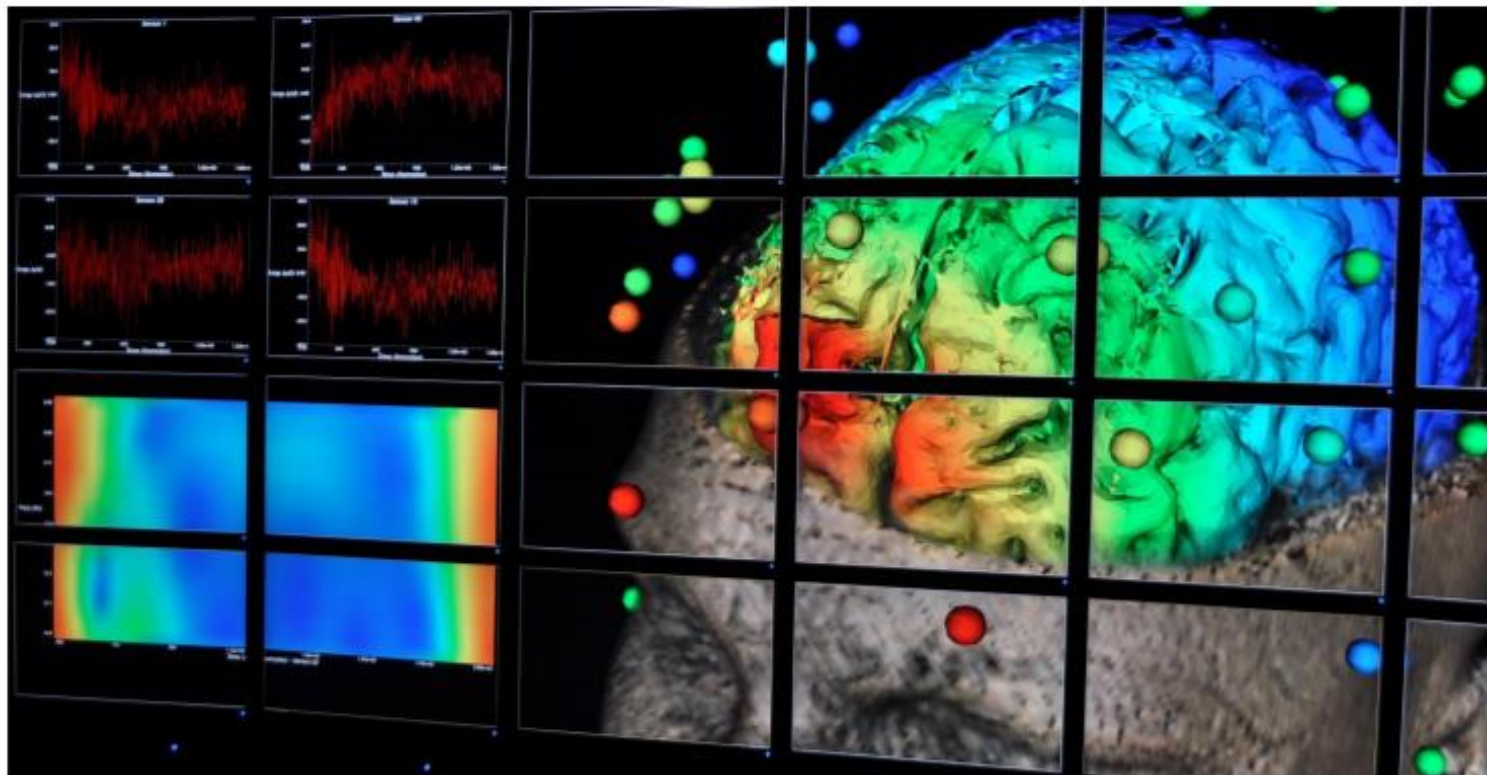
Mémorial (sous forme de pompe sans poignée), à Londres, aujourd'hui rue Broad street (aujourd'hui Broadwick)





[John Snow,](#)

1854. London.



VisTrails https://www.nsf.gov/discoveries/disc_images.jsp?cntn_id=114322&org=NSF

Who imported Coffee, not roasted in 2014?

\$28,68 USD

Select a Country

All Countries

Exports

↔ imports

Product



Coffee, not roasted

Trade Data Visualization



Complexity Visualization



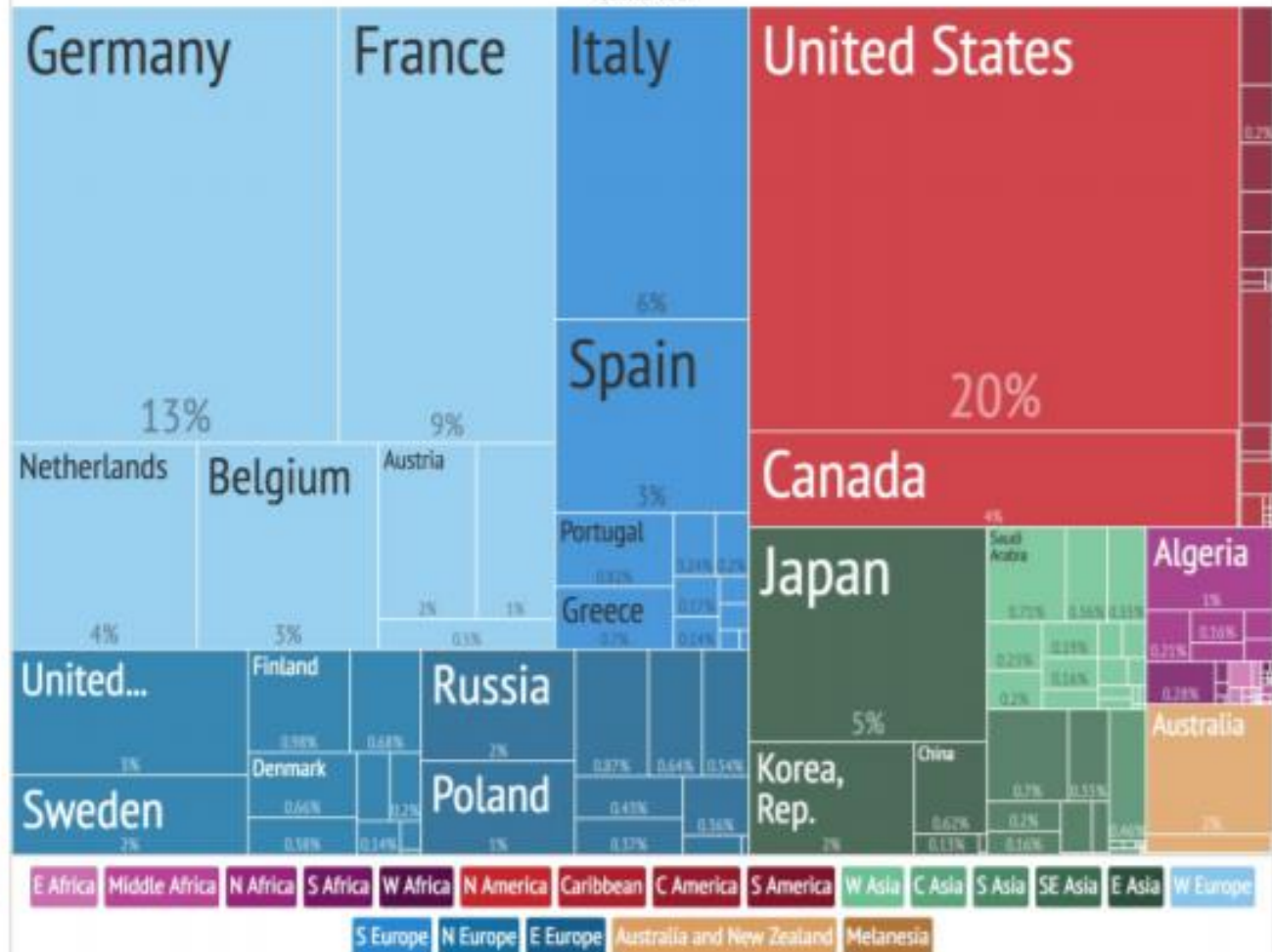
Time Period

2014

End

Legend

Welcome to the Atlas of Economic
Inequalities. Please note in advance to



1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

Product Class

SITC4

H54

Detail Level

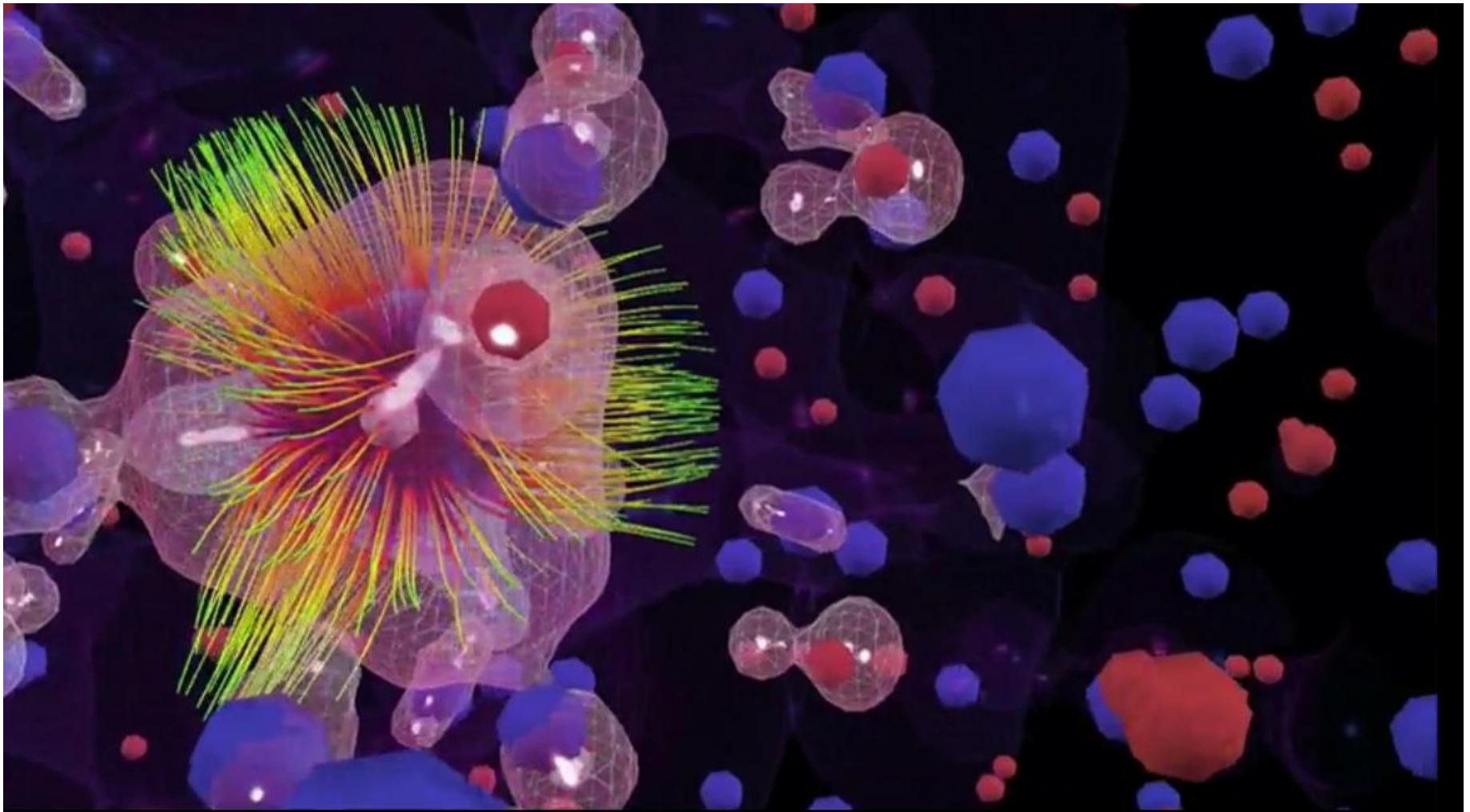
Low

Hi

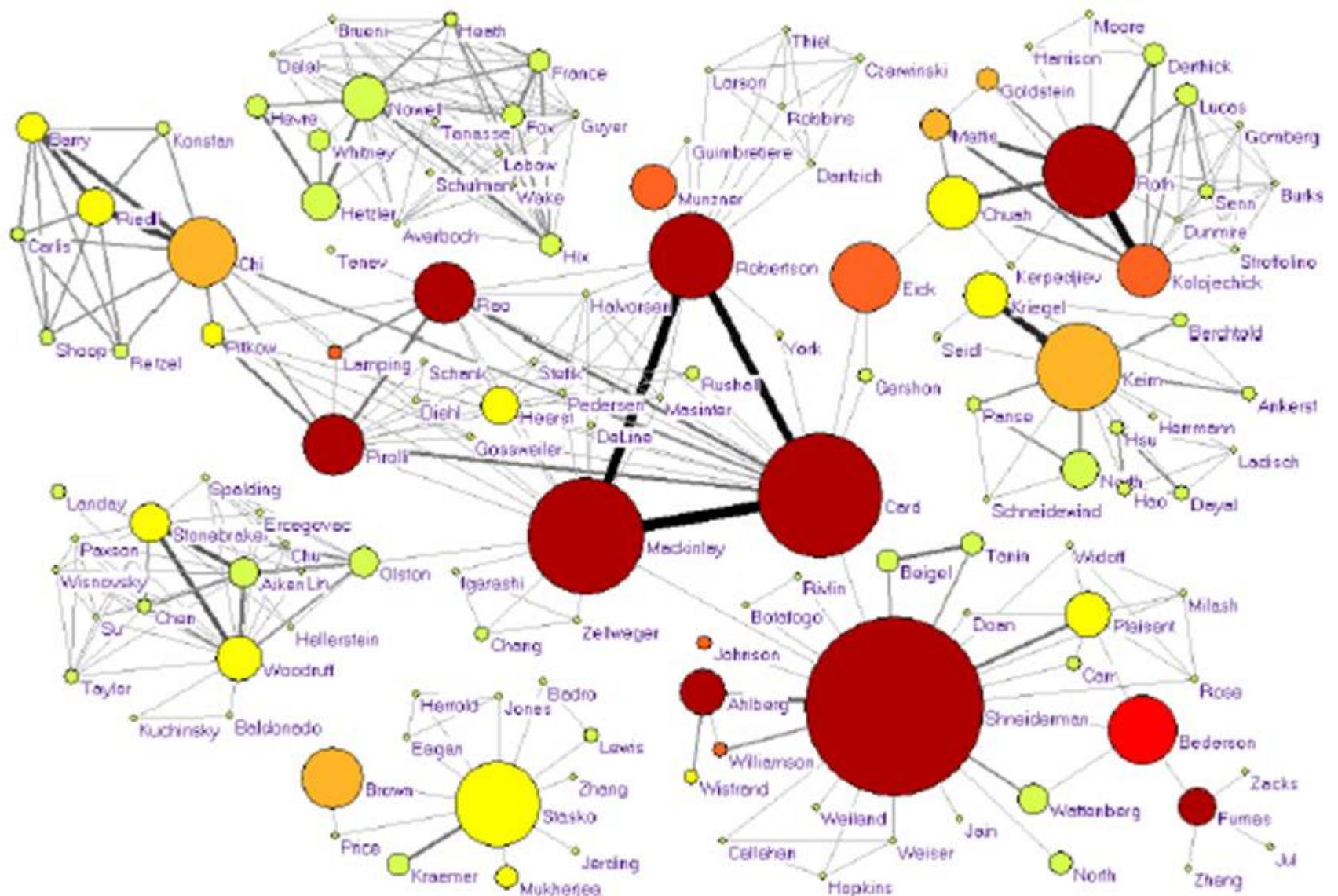
Trade Flow

Gross

Net



Visualisation scientifique



Color Code:



node size - number of papers

node color - number of citations

line width - number of times co-authored

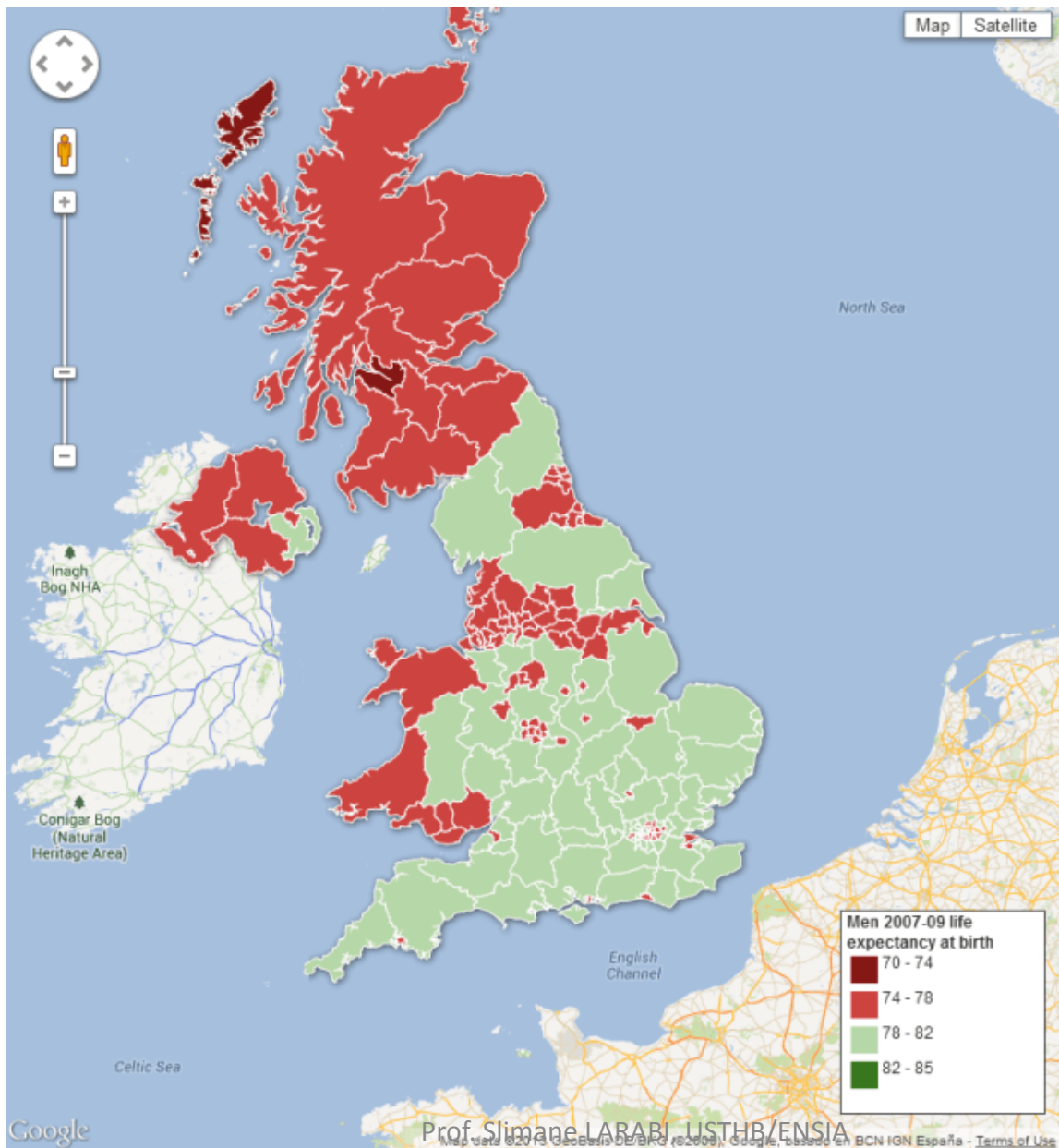
Prof. Slimane LABABRI/USTH/ENSIA



US Air Traffic, Aaron Koblin

*THE YELLOW LINES REPRESENT BUSES IN REAL TIME AND THE RED
CORRESPONDS TO DENSITY OF PEOPLE. SEP 2006.*





CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

[Hans Rosling](#)

2007



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

1.5 Principes de base du codage visuel

Attributs à utiliser dans la production de visualisation

Quels sont les éléments visuels qui attirent l'attention d'un observateur pour une visualisation de données (attributs du traitement pré-attentif) :

Référence : Colin Ware. Information Visualization : Perception for Design. Morgan Kaufmann, San Francisco, CA, 2nd edition, 2004

CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- La couleur : Combien de chiffres 9 sont écrits sur la liste suivante ?

1 2 4 6 5 3 4 7 8 3 4 5 6 2 3 5 6 9 0 7 6 4 3 1 2 4 5 6 3 4 5
3 5 6 7 3 9 5 4 6 1 3 2 9 0 7 5 4 6 3 2 1 2 6 9 0 5 4 2 3 1 4
2 4 3 5 4 3 1 6 9 6 4 7 8 6 9 0 3 2 1 4 6 7 3 4 1 2 5 4 6 2 6
6 5 3 4 5 2 9 8 7 5 4 6 5 6 2 3 4 5 8 9 0 8 9 0 3 2 4 5 6 1 2
3 4 5 2 3 1 6 7 8 3 4 6 5 1 2 7 9 8 6 5 6 4 3 5 7 8 1 0 2 1 4
1 3 2 5 4 6 4 5 3 2 7 0 5 6 7 8 3 2 4 5 6 7 5 4 2 3 6 7 8 4 3
5 6 3 4 5 6 7 2 3 4 1 2 6 7 8 3 4 6 7 2 3 4 1 5 6 4 2 3 4 1 4

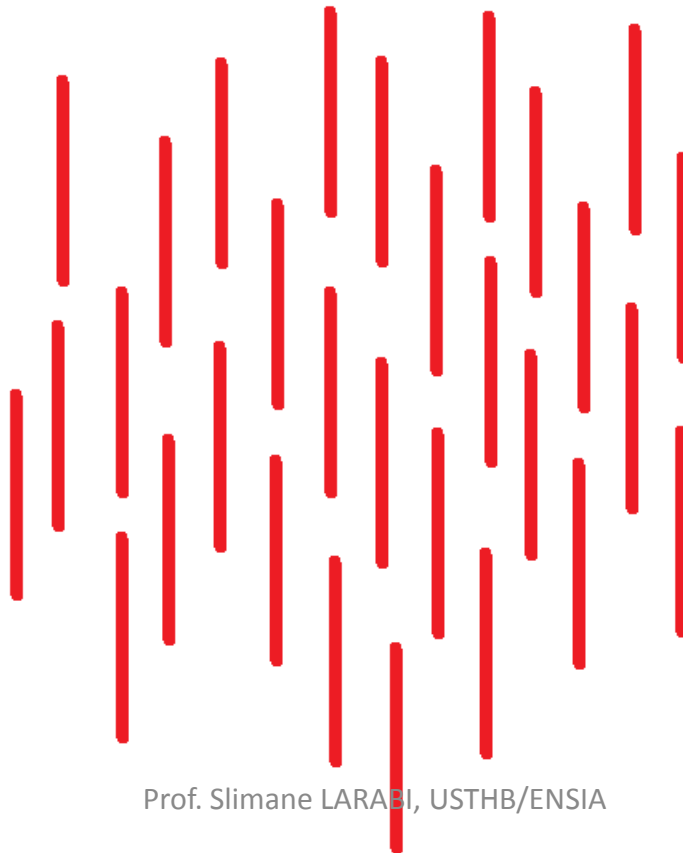
CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- La couleur : Combien de chiffres 9 sont écrits sur la liste suivante ?

1 2 4 6 5 3 4 7 8 3 4 5 6 2 3 5 6 9 0 7 6 4 3 1 2 4 5 6 3 4 5
3 5 6 7 3 9 5 4 6 1 3 2 9 0 7 5 4 6 3 2 1 2 6 9 0 5 4 2 3 1 4
2 4 3 5 4 3 1 6 9 6 4 7 8 6 9 0 3 2 1 4 6 7 3 4 1 2 5 4 6 2 6
6 5 3 4 5 2 9 8 7 5 4 6 5 6 2 3 4 5 8 9 0 8 9 0 3 2 4 5 6 1 2
3 4 5 2 3 1 6 7 8 3 4 6 5 1 2 7 9 8 6 5 6 4 3 5 7 8 1 0 2 1 4
1 3 2 5 4 6 4 5 3 2 7 0 5 6 7 8 3 2 4 5 6 7 5 4 2 3 6 7 8 4 3
5 6 3 4 5 6 7 2 3 4 1 2 6 7 8 3 4 6 7 2 3 4 1 5 6 4 2 3 4 1 4

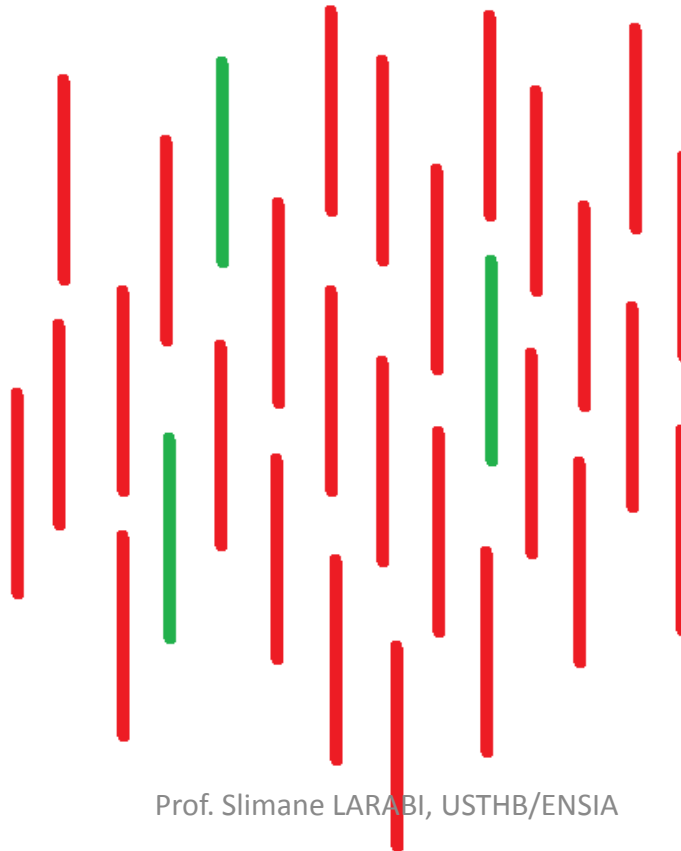
CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

Teinte (hue)



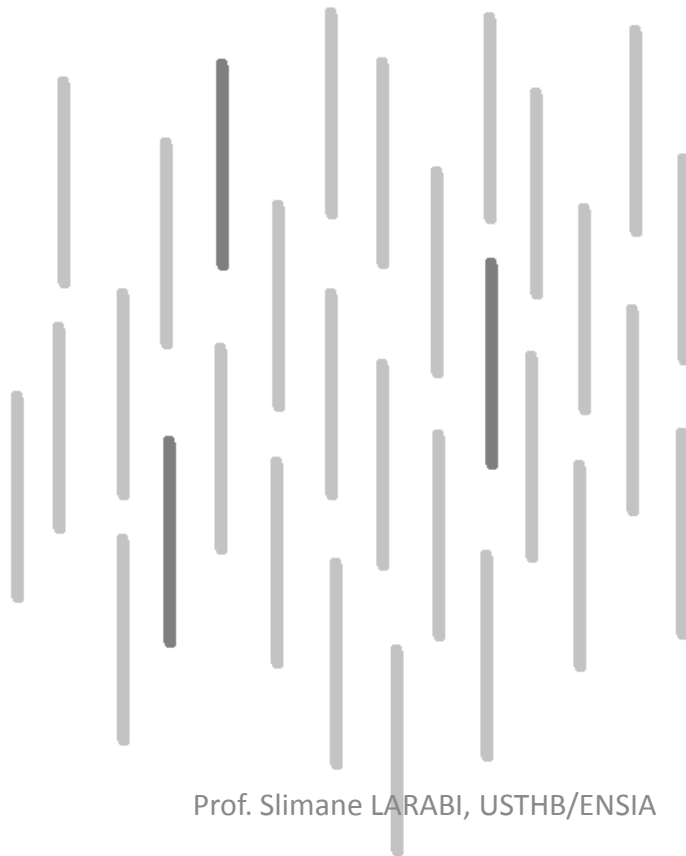
CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

Teinte (hue)



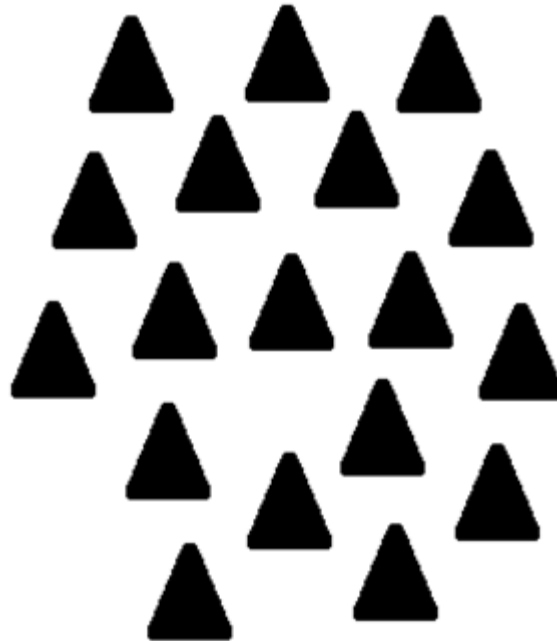
CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- Intensité



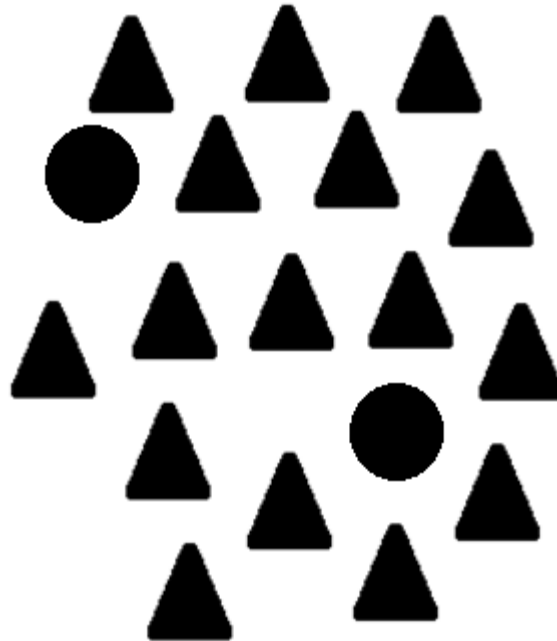
CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- Forme



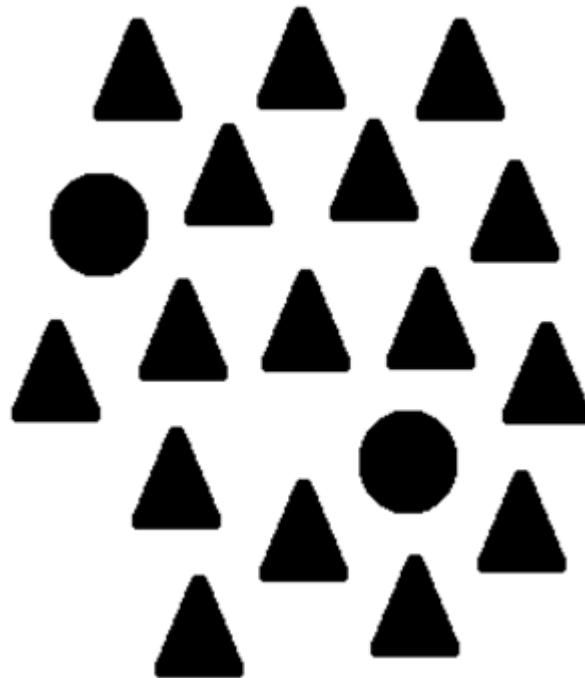
CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- Forme



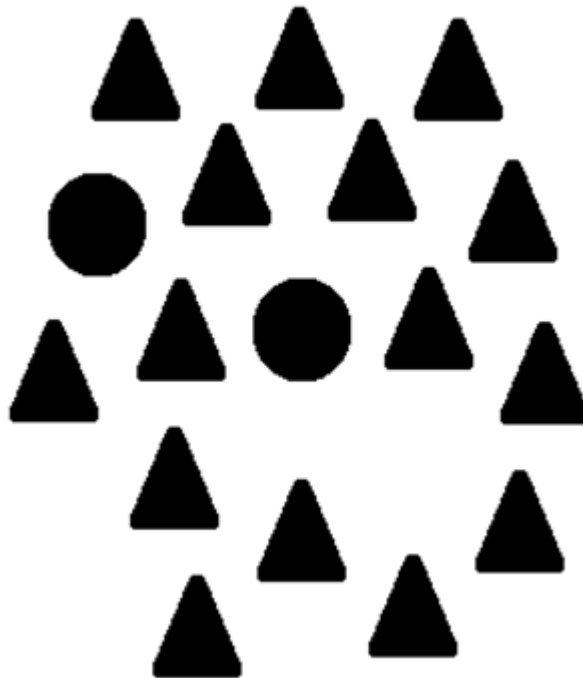
CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- Mouvement



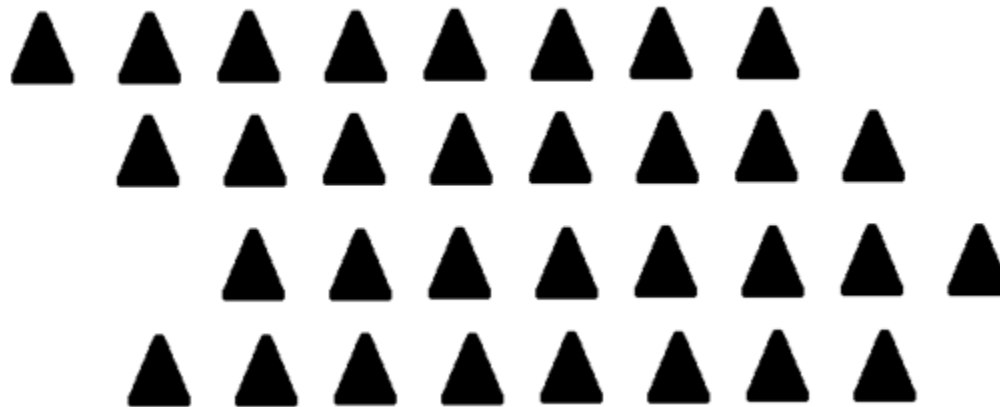
CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- Mouvement



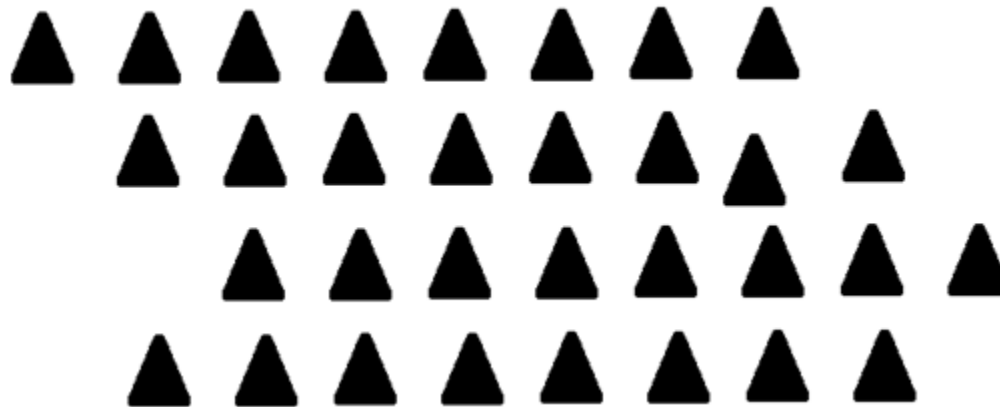
CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- Position spatiale



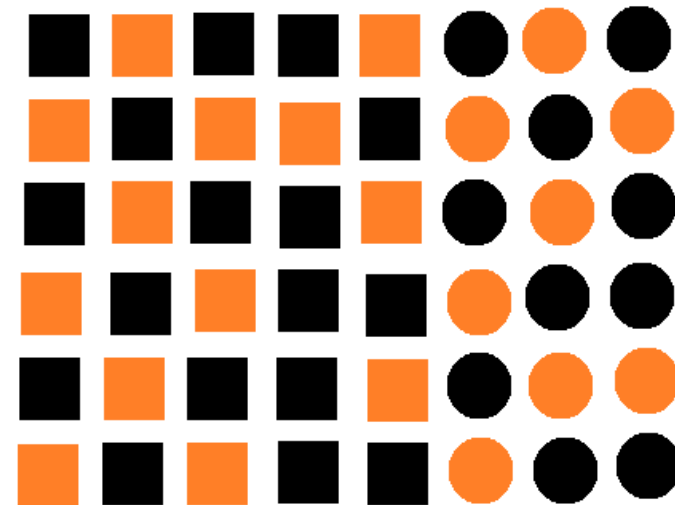
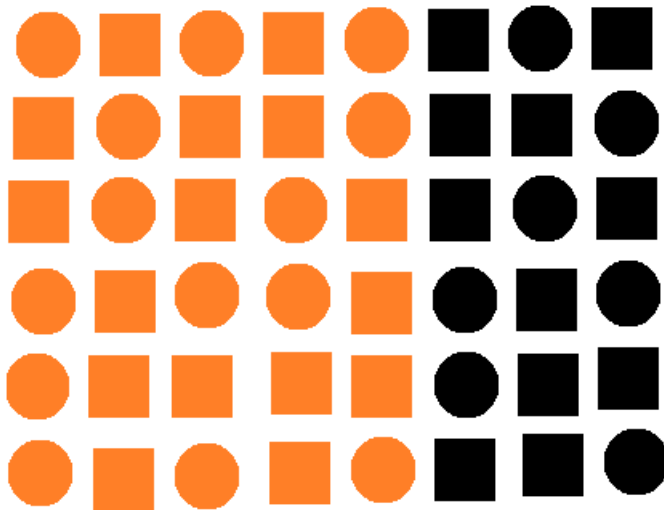
CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- Position spatiale



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

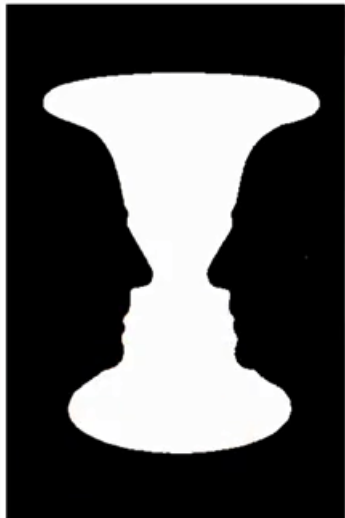
La forme ou la couleur sont des attributs pré-attentifs qui permettent par exemple de délimiter la frontière présente dans la présentation.



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

Autres attributs

- L'espace négatif



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

Précautions à prendre en compte pour l'utilisation de la couleur :

- Noir et blanc (largement suffisant)

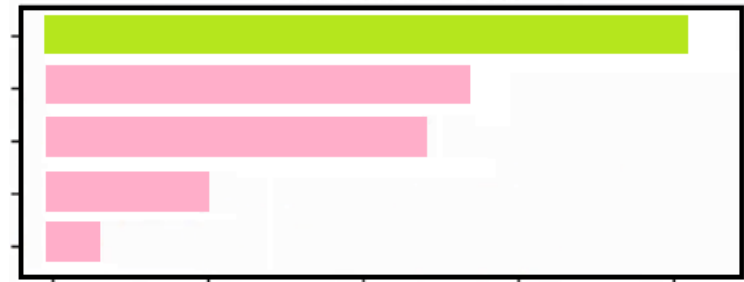
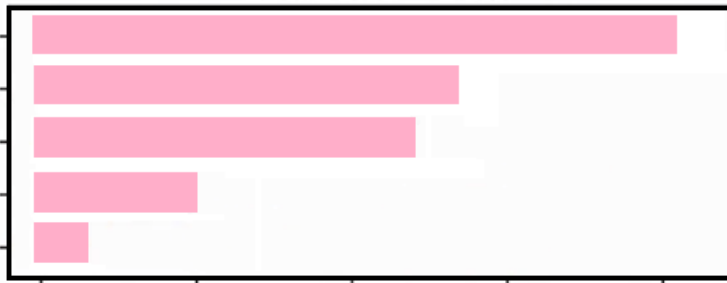


CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

Précautions à prendre en compte pour l'utilisation de la couleur :

Utiliser des teintes moyennes ou couleurs pastels

- Utiliser la couleur pour rehausser l'information



1.6 Le Langage Visuel

Les images sont perçues comme un ensemble de signes.

L'émetteur code l'information comme signes.

Le récepteur reçoit et décode l'information à partir de signes

« J. Bertin, Sémiologie Graphique. Les diagrammes, les réseaux, les cartes, Paris, La Haye, Mouton, Gauthier-Villars, 1967 »

- **Définition de la sémiologie graphique:**
- « L'ensemble de règles qui permettent l'usage d'un système de signes graphiques pour transmettre l'information ».

CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

Les variables de Jacques Bertin

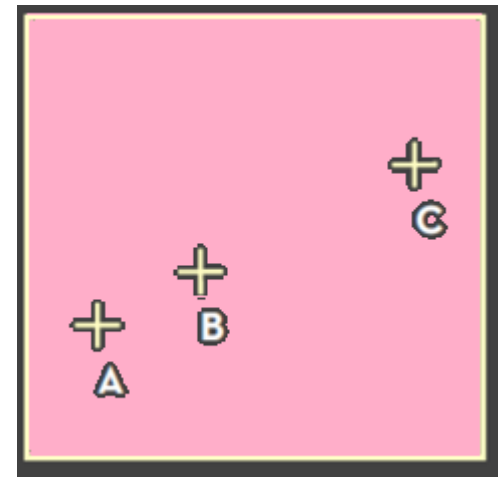
Cartographe français,
auteur de la sémiologie graphique



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

Sémiologie est la science des signes

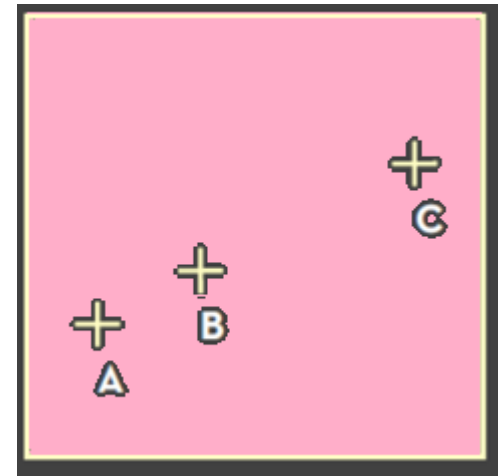
Exemple de sémiologie de graphique selon Jacques Bertin, 1967



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

Sémiologie est la science des signes
sémiologie de graphique selon Jacques Bertin, 1967

A, B, C sont distincts
B est entre A et C
BC est deux fois plus long que AB



- **Types de variables:**
 - **Nominale,**
 - **Ordinale,**
 - **Quantitative**

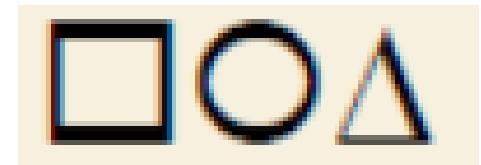
- **Types de variables:**

- **N – Nominale**

- Fruits: Pommes, oranges, ...

- Forme: carré, triangle, ...

- Moyen de transport: camion, navire,...



 Automobile

 Camion

 Navire

 Moto

- **N – Nominale**

- Catégorie: sexe (M/F)
- Qualité (qualité de chocolat: b, n,...)
- Couleur: r, v, b, j (peut importe la permutation)

- **Types de variable: Nominale, Ordinale et Quantitative**
- **O – Ordinale:**
 - Qualité de viande: A, AA, AAA
 - Mention: P, AB, B, TB
 - Taille: S, M, L, XL
 - Rang,
 - Satisfaction,
 - Imagination (Fanciness)

- **Types de variable:** Nominale, Ordinale et Quantitative
- **Q – Quantitative:**
 - Intervalle: âge
 - Ratio: mesure: (discrete ou continue)

- **Types de variable: Nominale, Ordinale et Quantitative**
- N –Operations: =, ≠
- O –Operations: =, ≠, <, >
- Q - Operations: =, ≠, <, > - %
-

- **Exemple: N, O, Q?**
- **Population**
- **Année**
- **Age**
- **Sexe (M/F)**
- **Situation familiale**

- **Exemple: N, O, Q?**
- **Population**
- **Année**
- **Age**
- **Sexe (M/F)**
- **Situation familiale**
- Q-Ratio
- Q-Intervalle (pour les statistiques)
- Q-Ratio
- N
- N

CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

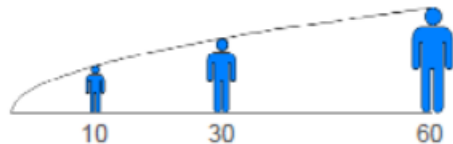
- **Exemple: N, O, Q?**

	Point	Region
Space	(Latitude, Longitude) e.g. "My current location"	Geographic Identifier e.g. Countries, States, Counties
	Categorical	Categorical
Time	Instant e.g. "Right now"	Interval e.g. "This year"
	Quantitative	Ordinal
Quantity	Value e.g. 5.2	Interval e.g. "5 - 10 years old"
	Quantitative	Ordinal

- **Les variables visuelles :**
 - **Taille,**
 - **Forme,**
 - **Couleur,**
 - **Valeur,**
 - **Texture,**
 - **Orientation.**

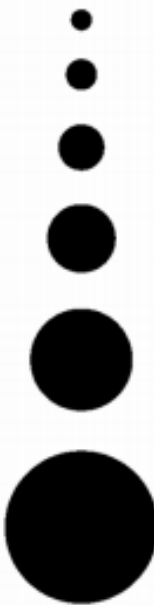

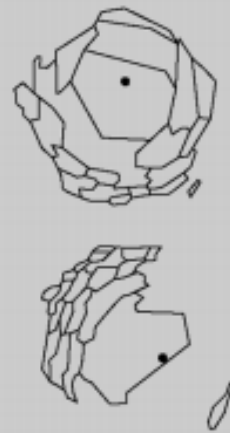
CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- **Langage Visuel comme système de signes**
- **La taille :**
 - La représentation par symbole d'une variable quantitative implique une modification proportionnelle de sa taille par rapport à la valeur à visualiser.
 - Le changement s'effectue alors sur la longueur, la largeur ou la superficie du symbole.



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- Langage Visuel
- La taille :

Points	Lignes	Surfaces
		

CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- **Langage Visuel comme système de signes**
- **La forme :**
- Une modification de la forme implique plus souvent un changement qualitatif de la variable. En fonction de la classe à laquelle appartient l'observation, on variera la forme visuelle du symbole.



Poste d'urgence



Clinique



CHR



CHU

Type d'établissements de soin



Automobile



Camion



Navire



Moto

Mode de transport attitré
Prof. Slimane LARABI, USTHB/ENSIA

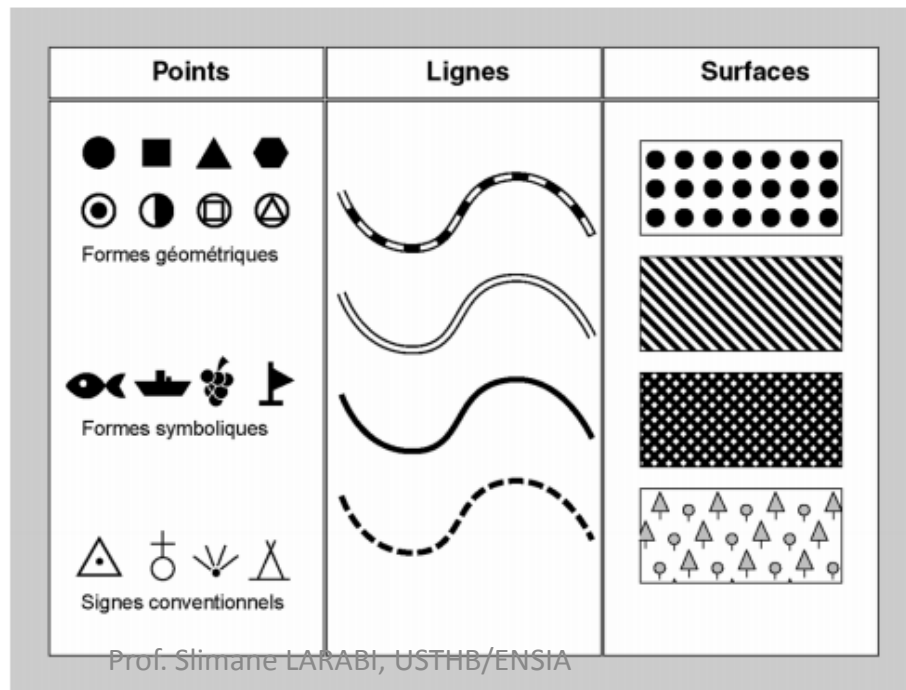
CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- Langage Visuel comme système de signes

Les variables visuelles :

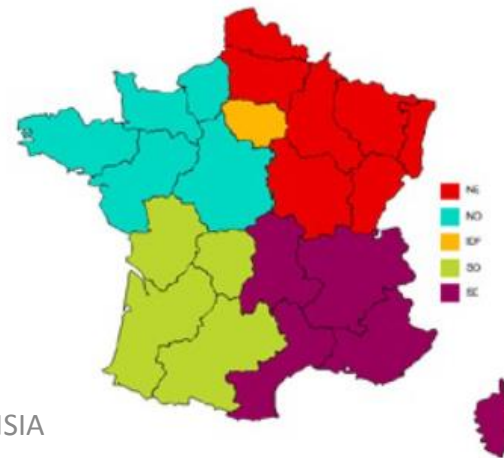
- La forme

La forme



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- **Langage Visuel comme système de signes**
- **La couleur :**
 - La variation de symbole par la couleur bénéficie d'une excellente perception par
 - l'oeil humain. La couleur peut varier selon trois nuances : le ton, la saturation et l'intensité.



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- **Langage Visuel**
- **La couleur : 12 couleurs recommandées (6+6)**



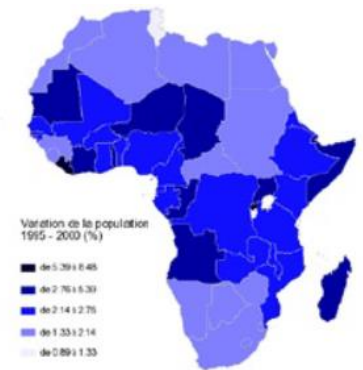
[Healy, "Choosing effective colors for data visualization" Proc. Visualization, 1996]

- **Langage Visuel comme système de signes**
- **La couleur :**
- **Le ton (hue):**
- La variation du ton (couleur dominante) est plus fréquemment utilisée
- pour une variable qualitative.



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- **Langage Visuel comme système de signes**
- **La couleur :**
- **La saturation :**
- La variation de l'intensité et/ou de la saturation permet de construire des palettes de dégradés. Elles illustrent des évolutions de données quantitatives

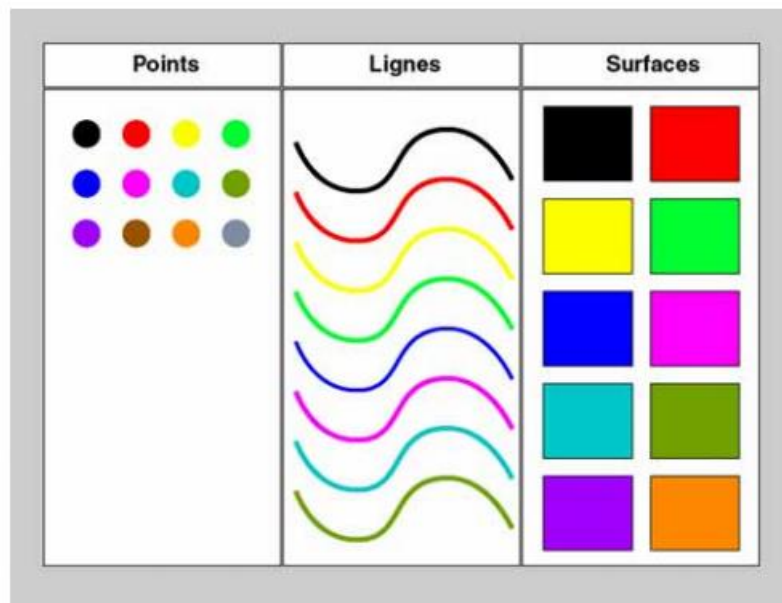


- **Langage Visuel comme système de signes**
- **La couleur :**
- **L'intensité :**
- *L'intensité correspond également à la variable visuelle de « valeur »*

CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

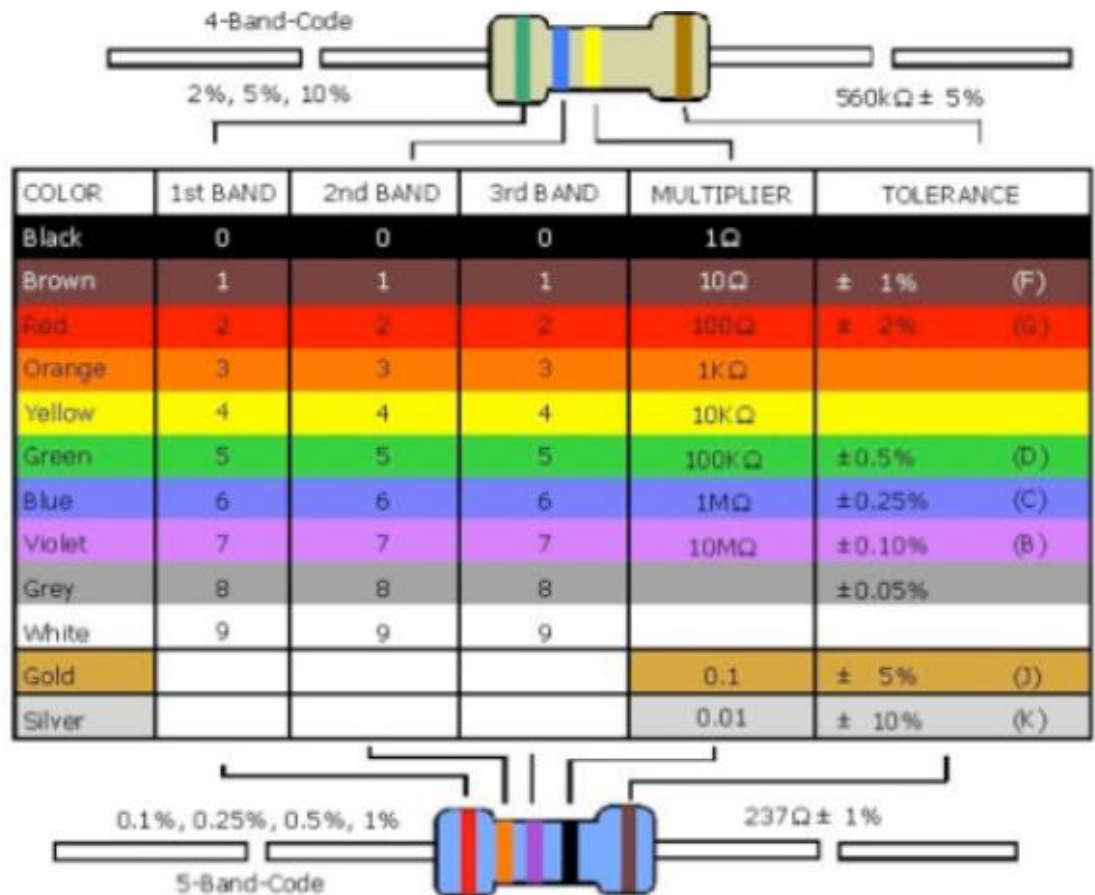
- Langage Visuel comme système de signes
- La couleur :

La couleur



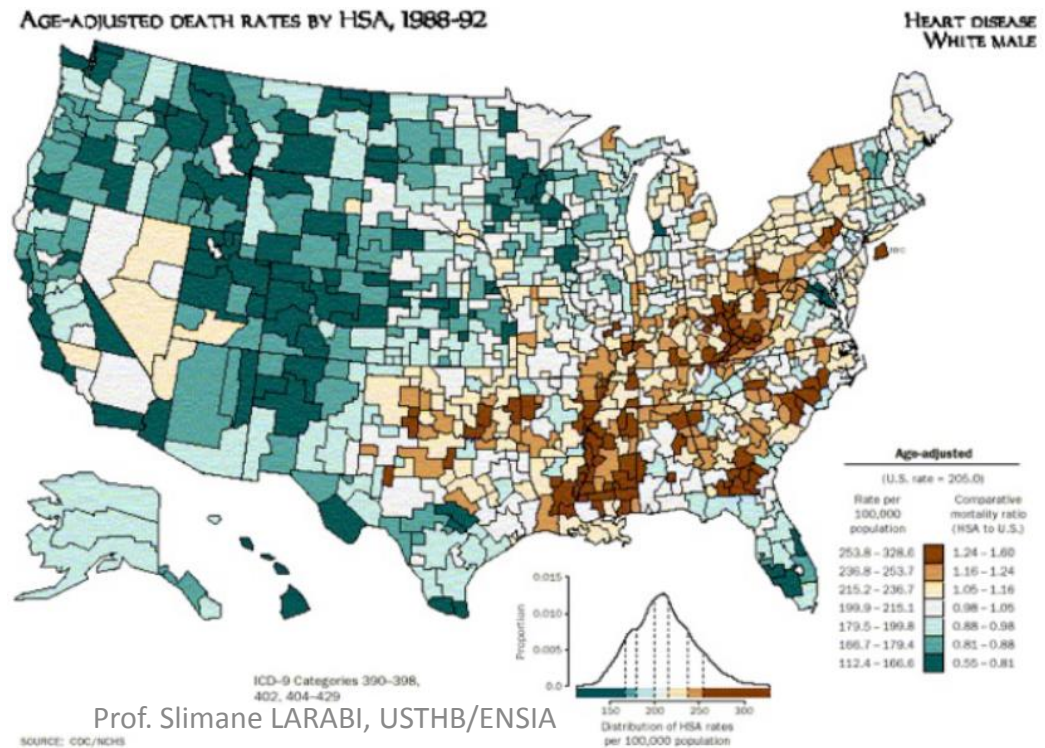
CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- Langage Visuel
- La couleur :



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- Langage Visuel
- La couleur :
- Couleur Quantitative



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES




- **Langage Visuel comme système de signes**
- **La valeur :**
- La valeur est « le rapport entre la quantité de noir et de blanc sur une surface donnée » (D. Poidevin, «*La carte, Moyen d'action*»).
- On considère que le noir a plus grande valeur que le blanc. Ainsi, une valeur élevée sera représentée avec une plus grande proportion de noir que de blanc. Les variations de valeur seront donc utilisées pour des données quantitatives.

Langage Visuel comme système de signes

- **La valeur :**

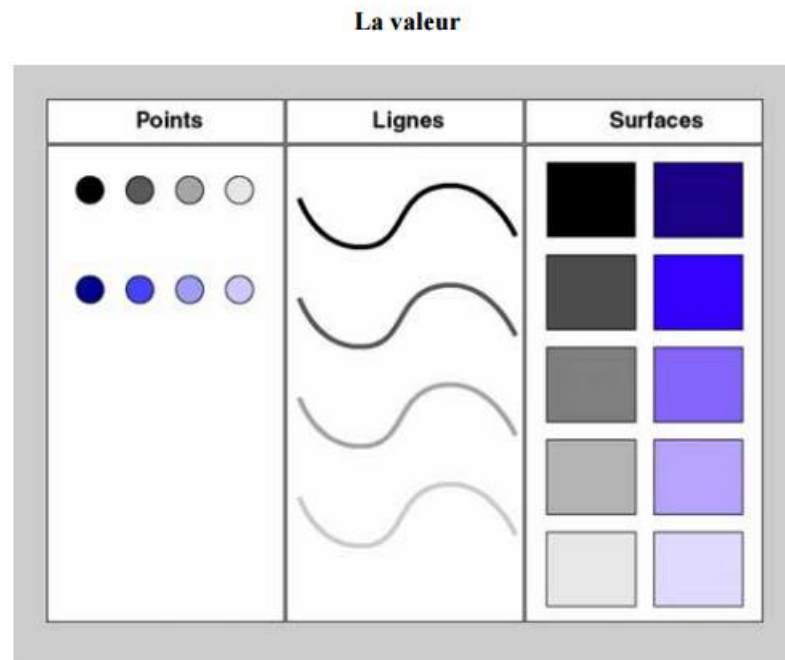
Dans une variation par couleur, cette représentation sera influencée par « l'intensité» de la couleur.

Mais la valeur peut également être mise en jeu dans d'autres types de représentation (hachures, pointillées, grains, etc.).

- Modification de la valeur par l'épaisseur du trait : 
- Modification de la valeur par l'écartement du trait : 
- Modification de la valeur par la teinte : 

CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- Langage Visuel
- La valeur



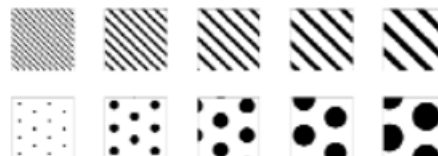
CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- **Langage Visuel**
- **Texture**

Contrairement à la variation par la valeur, qui implique une évolution blanc-noir, la variation par le grain vise à conserver cet équilibre. Ici, c'est le nombre d'éléments de la trame qui varie, et non la valeur.

Pour maintenir cet équilibre, une variation par le grain commence par une trame très resserrée de taches très petites, et se termine par un nombre faible de taches épaisses plus espacées.

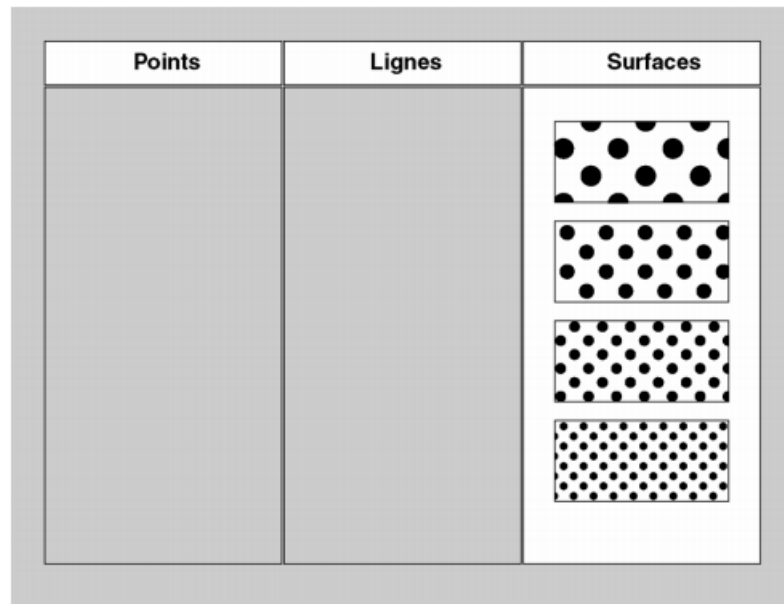
Cette variation peut s'identifier à un zoom sur une rame spécifique (les rapports taille/espacement des éléments est conservé).



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- Langage Visuel comme système de signes
- Texture

Le grain



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- **Langage Visuel comme système de signes**
- **Orientation :**

La sixième variable visuelle, l'orientation, consiste à changer l'angle des hachures dans la représentation. Cette évolution d'orientation permet de distinguer certaines classes d'observation (données qualitatives).

Elle concerne les objets ponctuels ou surfaciques.

En représentation surfacique, la distinction est parfois difficile à faire lorsqu'on enregistre un trop grand nombre de classes. Cette variation par l'orientation est donc à utiliser avec précaution.



CHAPITRE 1. FONDEMENTS DE LA VISUALISATION DE DONNEES

- Langage Visuel comme système de signes
- Orientation :

