
Introduction à la visualisation de données

<https://lyondataviz.github.io/teaching/lyon1-m2/2017/>

2 Novembre 2017

Plan

- Présentation du cours
- Critique
- Pourquoi visualiser ?
- Qu'est ce que la visualisation
- Type de données
- Variables graphiques
- Mapping + visualisation pipeline
- Un classique



Nicolas Bonneel
nicolas.bonneel
@univ-lyon1.fr



Aurélien Tabard
aurelien.tabard
@univ-lyon1.fr



Romain Vuillemot
romain.vuillemot
@ec-lyon.fr

Déroulé

<https://lyondataviz.github.io/teaching/lyon1-m2/2017/>

02/11	Introduction à la visualisation de données Critique + Cours + TP	(4h)
09/11	Introduction à D3 Critique + Cours + TP	(4h)
16/11	Design de visualisation Présentation articles + Cours + Projet	(3h)
24/11	Visualisation de données temporelles Présentation articles + Cours + TP	(4h)
01/12	Visualisation de données spatiales Présentation articles + Cours + TP	(3h)
07/12	Séance revue de projet Présentation articles + Projet	(3h)
14/12	Visualisation de graphes Présentation articles + Cours + TP	(4h)
21/12	TP projet banalisé	(3h)
11/01	Soutenance projet	(4h)

Évaluation : Contrôle continu intégral

- Présentation d'un article scientifique (20%)
- Un TP noté (Visualisation de Graphes) (20%) - rendu le 17/12
- Une note de projet (60%)

Projet – thème 2017 : transports

Travail : en groupe (Trinôme)

Rendu : article + une visualisation Web interactive avec D3.js

- process book
- rédaction d'un article ou d'un site (33%)
 - 1e rendu le 29/11,
 - rendu final le 12/01
- peer review (5%) - le 07/12
- réalisation technique (45%) - rendu le 10/01
- présentation et démo (17%) - le 11/01

Lecture et présentation d'articles

Travail : en groupe (binôme)

Rendu : présentation + supports pdf.

- Vote sur les articles d'ici mercredi prochain.
- 6 minutes de présentations (répéter en amont)
- 3 minutes de questions de la classe
- Supports de présentation soumis la veille du cours

Détails à venir sur la page :

<https://lyondataviz.github.io/teaching/lyon1-m2/2017/articles.html>

Plan

- Présentation du cours
 - Critique
 - Pourquoi visualiser ?
 - Qu'est ce que la visualisation
 - Type de données
 - Variables graphiques
 - Mapping + visualisation pipeline
 - Un classique
-

Critique

Exercice

Analyse critique d'une visualisation

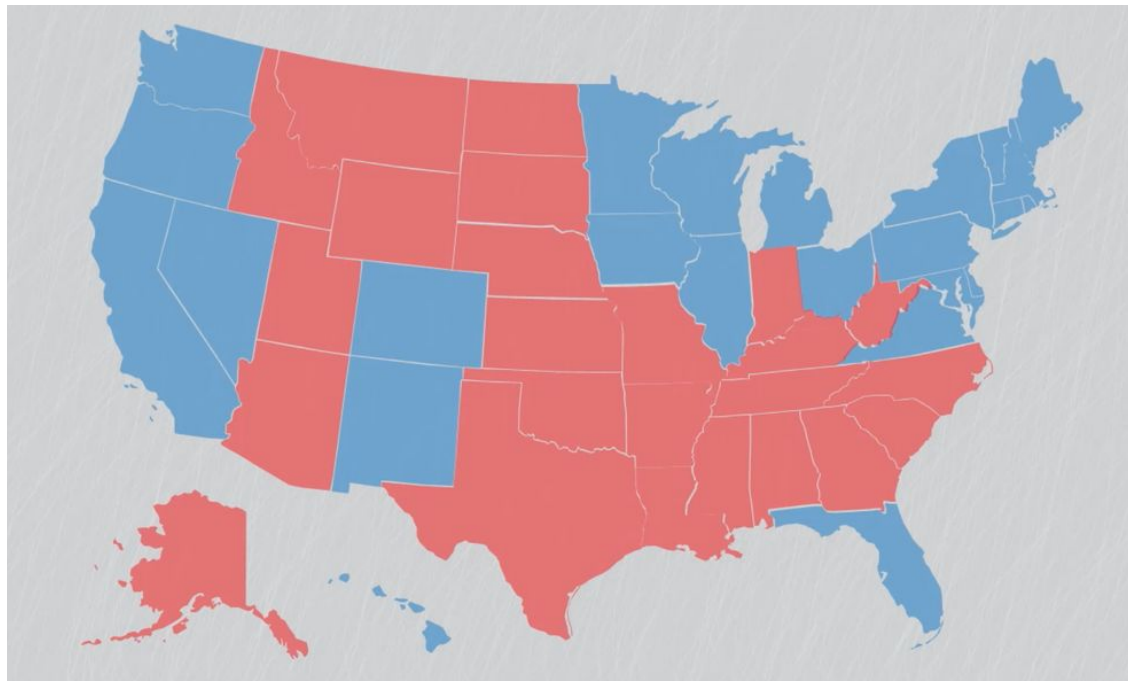
binome

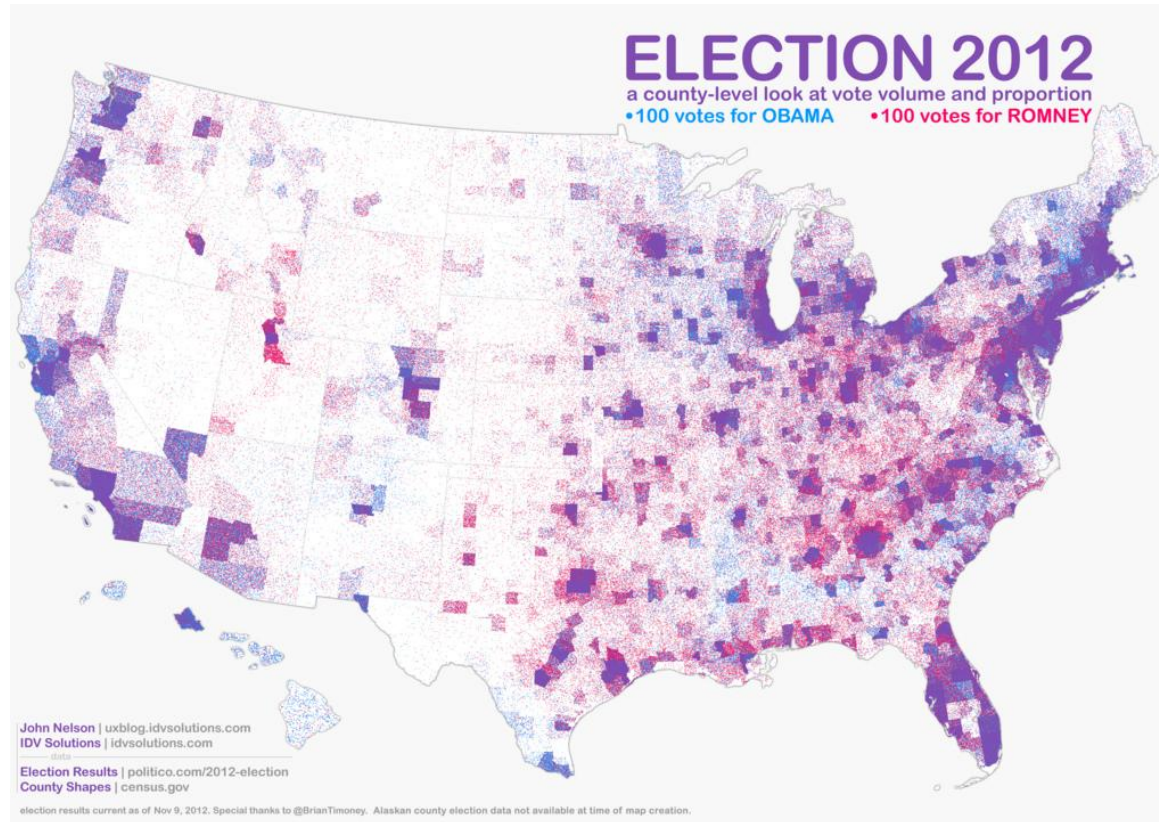
10 minutes

5 questions

Critique

- À qui s'adresse la visualisation ?
-> 1 proposition
- À quelle question la visualisation permet elle de répondre ?
-> 1 proposition
- Pourquoi (n')aimez vous (pas) cette visualisation ?
-> 2 raisons
- Quelles améliorations apporter ?
-> 3 propositions





<https://www.flickr.com/photos/idvsolutions/8182119174/sizes/k/in/photostream/>

The Electoral Map: Building a Path to Victory

◀ Prev Next ▶ Map 1 2 3 4 5 6 7 8 Make Your Own Scenarios

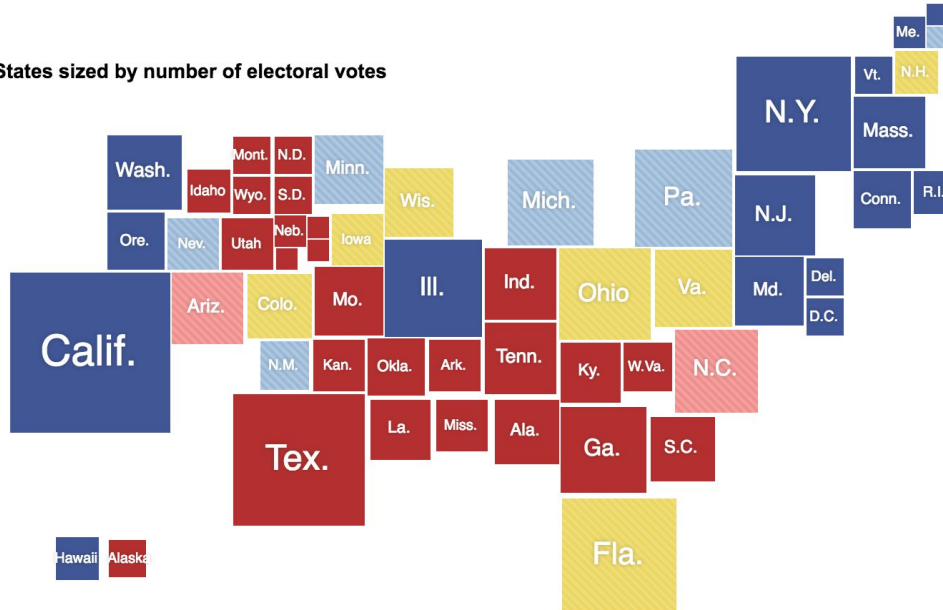
A New York Times assessment of how states may vote, based on polling, previous election results and the political geography in each state.

Obama
ELECTORAL VOTES **243**
Needs 27 to win

Needs 64 to win **206** Romney
ELECTORAL VOTES

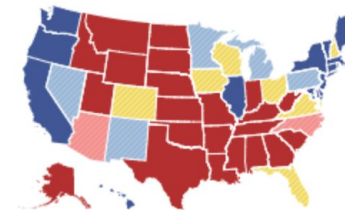


States sized by number of electoral votes



Maine and Nebraska give two electoral votes to the statewide winner and allocate the rest by congressional district.

Geographic View

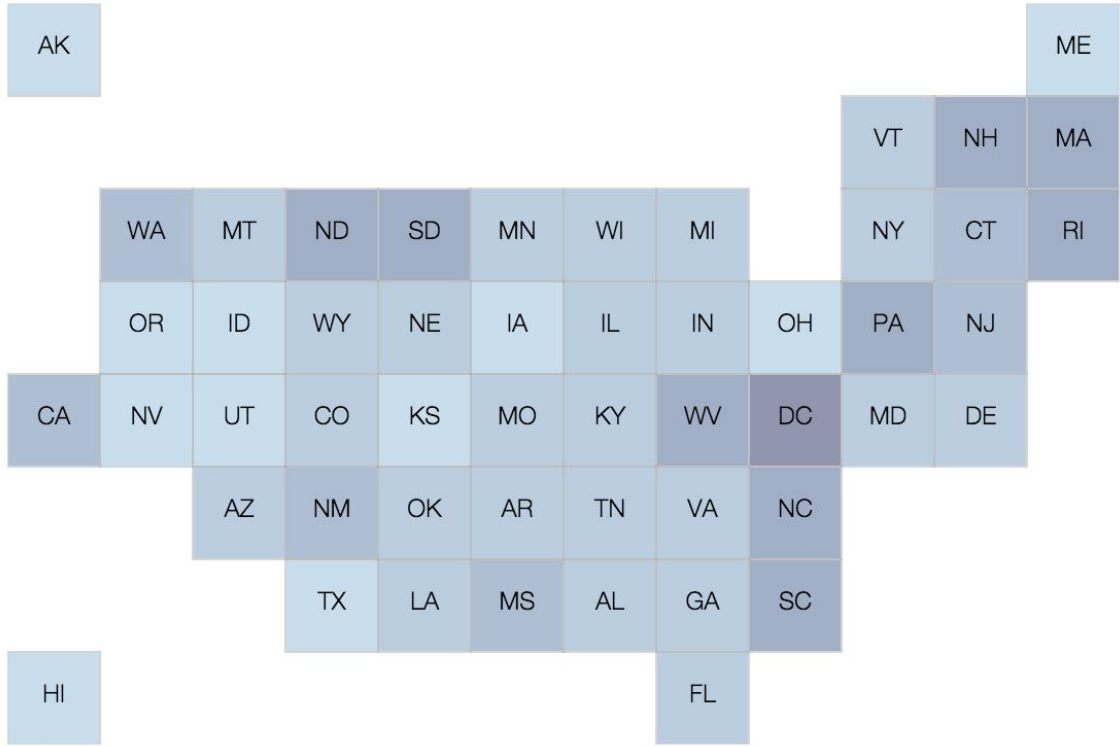


■

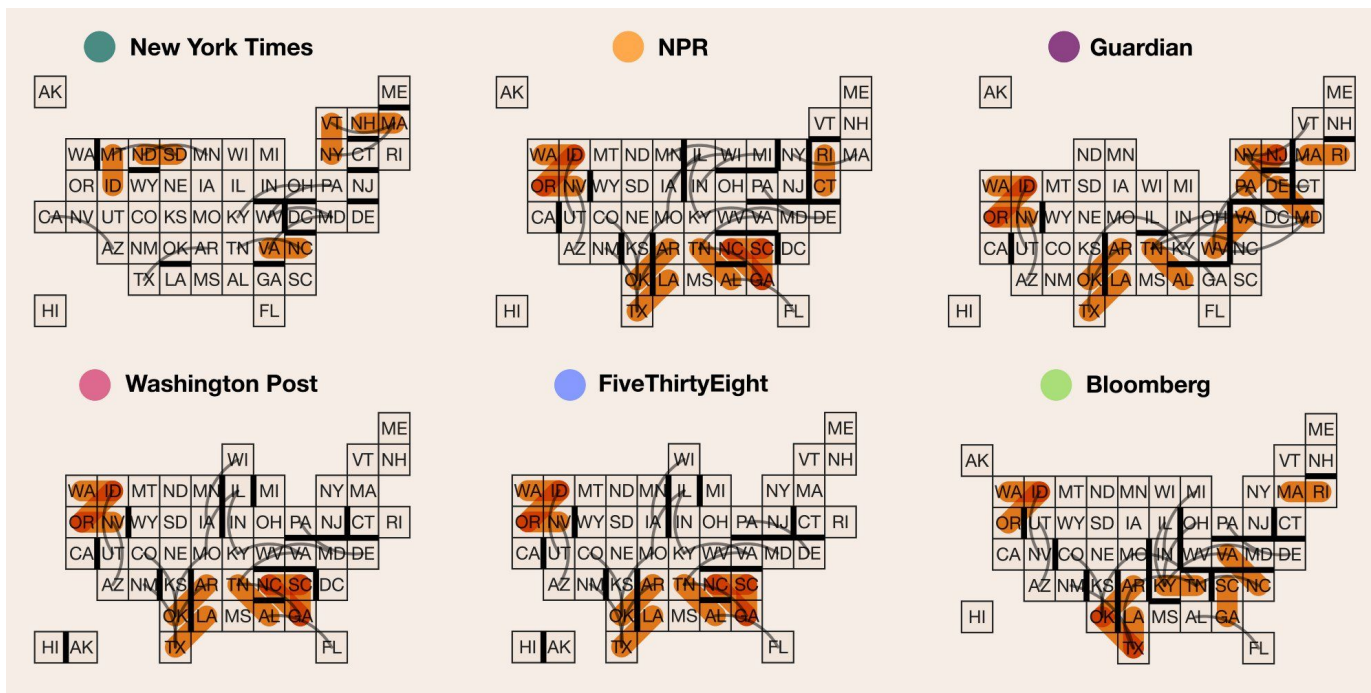
—

—

332	206
270 Needed to win	538 total votes



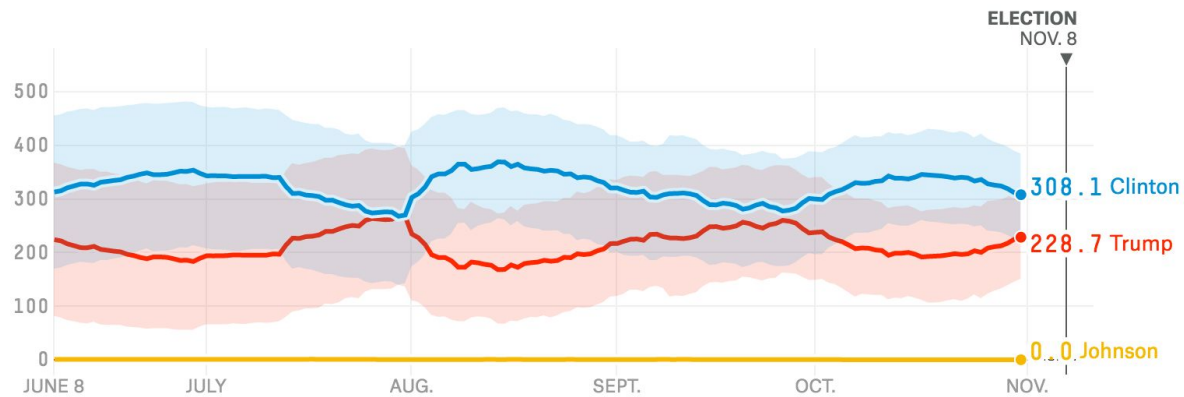
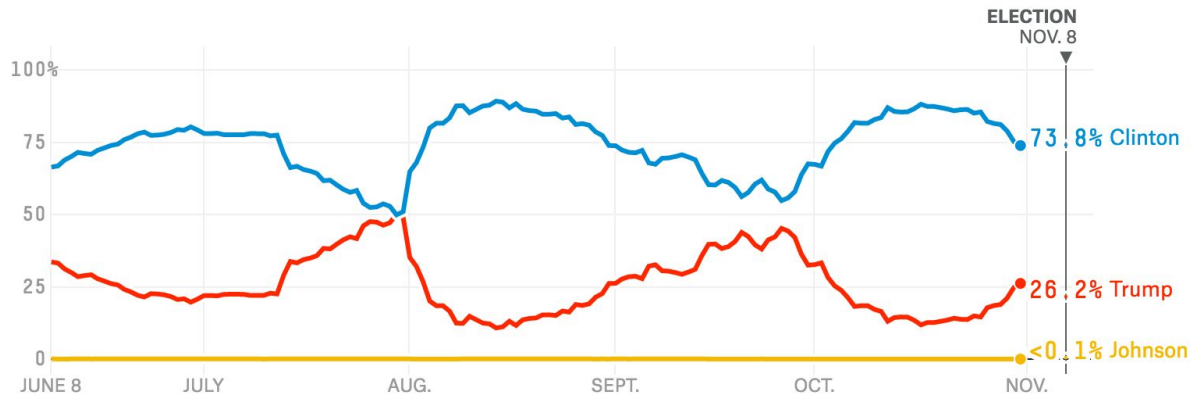
<https://github.com/kristw/gridmap-layout-usa>



Different US map layouts from six publishers.

Black border = invalid neighbors, Thick orange line = misdirection, Curve line = missing neighbors.

<https://medium.com/@kristw/whose-grid-map-is-better-quality-metrics-for-grid-map-layouts-e3d6075d9e80>



KEY AVERAGE ——— 80% CHANCE OF FALLING IN RANGE

<http://www.ericson.net/content/2011/10/when-maps-shouldnt-be-maps/>

Plan

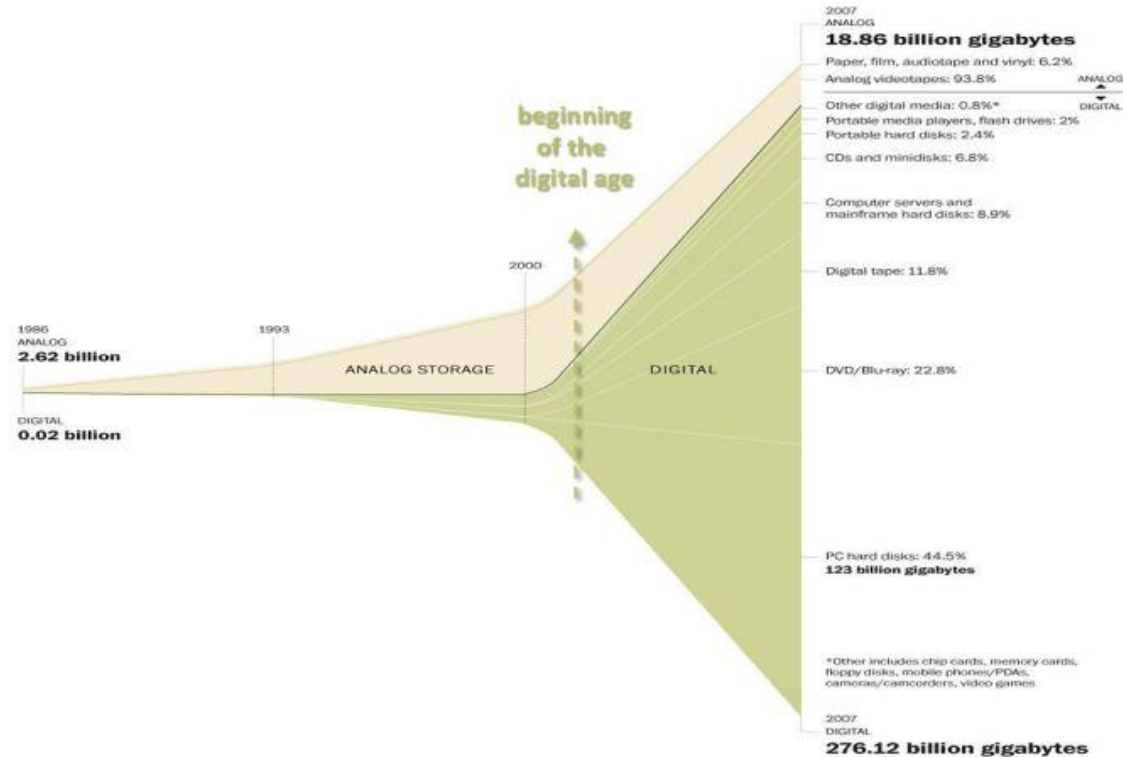
- Présentation du cours
- Critique
- **Pourquoi visualiser ?**
- Qu'est ce que la visualisation
- Type de données
- Variables graphiques
- Mapping + visualisation pipeline
- Un classique

Explosion des données

Neuman, Park et Panek, 2012.

Tracking the Flow of Information into the Home: An Empirical Assessment of the Digital Revolution in the U.S. from 1960–2005.

<http://ijoc.org/index.php/ijoc/article/view/1369/745>



<http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data>

Explosion de la quantité de données

- Comment faire sens des données ?
- Comment utiliser ces données dans les processus de décision ?
- Comment ne pas être surchargé ?

Défi: transformer les données en connaissance (découverte, compréhension) pour qu'elles deviennent utiles

“What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.”

Herb Simon
as quoted by Hal Varian
Scientific American
September 1995

Traiter les données : où l'ordinateur est plus efficace ?

Question bien définie, sur des données connues

- Quel est le taux de chômage ?
- Quel gène mute fréquemment sur tel ensemble de patients ?

Décisions doivent être faites en un minimum de temps

- High-frequency trading
- Détection de défaut sur une chaîne d'assemblage

Traiter les données : où l'humain est il plus performant ?

Quand les questions ne sont pas bien définies (exploration)

- Quelle combinaison de gènes peut être associée à un cancer ?

Quand les résultats peuvent donner lieu
à plusieurs interprétations

- Quelle est la relation entre l'emploi et la politique industrielle d'un pays?

Pourquoi ne pas s'appuyer sur l'analyse de données ?

Le Quartet d'Anscombe

https://en.wikipedia.org/wiki/Anscombe%27s_quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.10	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.10	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Statistiques

Moyenne

x: 9 y: 7.50

Variance

x: 11 y: 4.122

Corrélation

x – y: 0.816

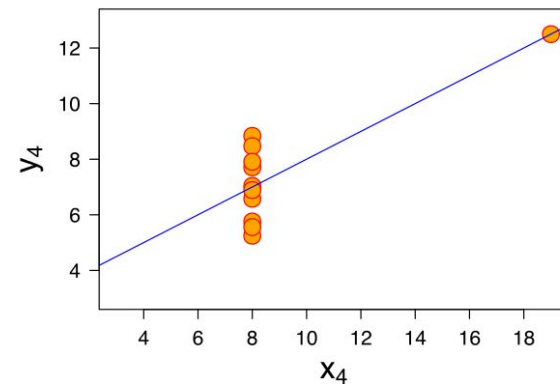
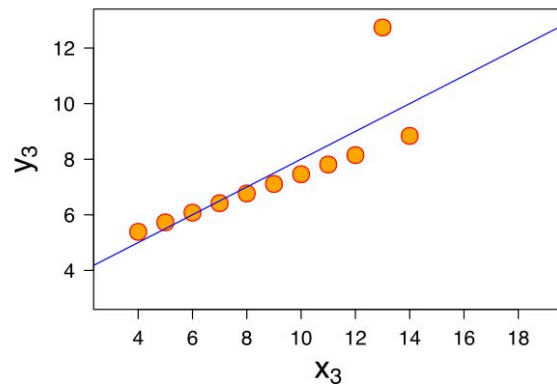
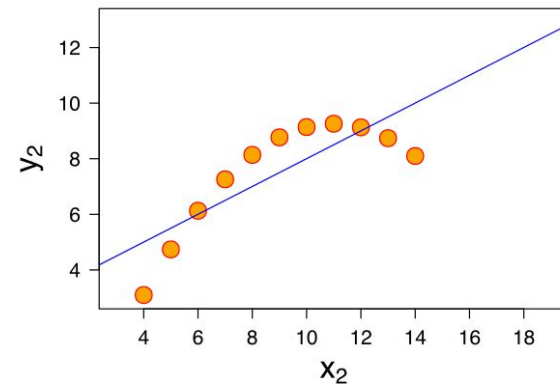
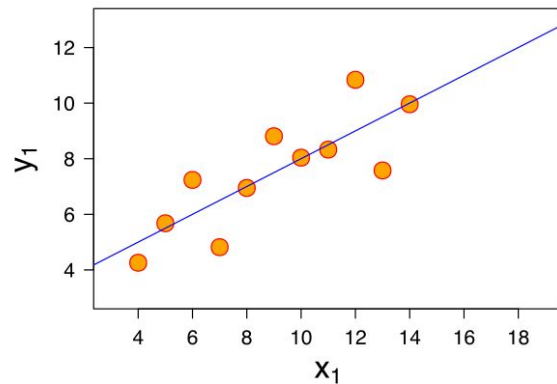
Régression linéaire:

$y = 3.00 + 0.500x$

Pourquoi ne pas s'appuyer sur l'analyse de données ?

Le Quartet d'Anscombe

https://en.wikipedia.org/wiki/Anscombe%27s_quartet



Pourquoi visualiser ?

Exercice

Brainstorming sur l'utilité
de la visualisation de données.

binome
3 minutes
5 raisons

Les trois raisons de la visualisation

Enregistrer de l'information

- Plan, photo

Faciliter le raisonnement sur de l'information (analyser)

- Analyser et calculer
- Reasonner sur les données
- Feedback et interaction

Transmettre de l'information (présenter)

- Partager et persuader
- Collaborer et itérer
- Mettre en avant un aspect des données

Enregistrer de l'information



© Mike Kelley – Photoviz <http://shop.gestalten.com/photoviz.html>

Faciliter le raisonnement

Épidémie de Choléra à Londres (1854)

Analyse de données visuelle pour comprendre le problème

[https://fr.wikipedia.org/wiki/%C3%89pid%C3%A9mie_de_chol%C3%A9ra_de_Broad_Street_\(1854\)](https://fr.wikipedia.org/wiki/%C3%89pid%C3%A9mie_de_chol%C3%A9ra_de_Broad_Street_(1854))



John Snow, 1854

Transmettre de l'information

<http://www.oecdbetterlifeindex.org/>



How's life?

Pourquoi la visualisation est difficile ?

Exercice

Visualiser les quantités suivantes :

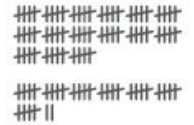
75 et 37

75, 37

a



b



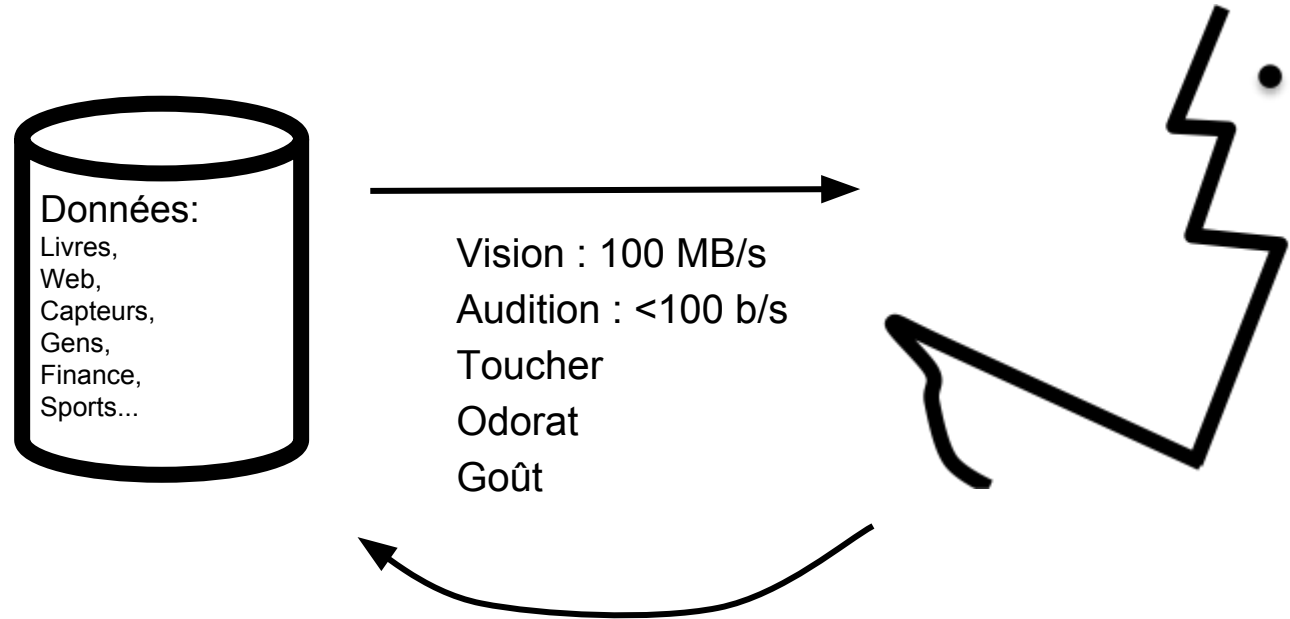
c

Pourquoi la visualisation est difficile ?

Exercice

Visualiser la distribution suivante :

Le défi



Plan

- Présentation du cours
- Critique
- Pourquoi visualiser ?
- **Qu'est ce que la visualisation**
- Type de données
- Variables graphiques
- Mapping + visualisation pipeline
- Un classique

Les différents types de visualisation : Infographics

WHEN THE WORLD WASHES

WASHING HABITS ARE DIFFERENT ACROSS THE GLOBE



SHOWER vs. BATH

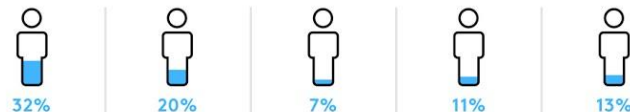
SHOWERING IS THE MOST COMMON METHOD OF WASHING

% Of People Showering Per Week



BUT BATHING IS STILL POPULAR IN EUROPE

% Of People Bathing Per Week



BRAZIL WATER CRISIS

ENVIRONMENTAL CHANGES CAN FORCE BEHAVIOUR TO CHANGE

Weekly showers and showering duration declined



Despite water shortage, they are still taking longer showers than most other countries



PEOPLE IN BRAZIL STILL TAKE MORE SHOWERS

Average Number Of Showers Per Week



Les différents types de visualisation : Storytelling

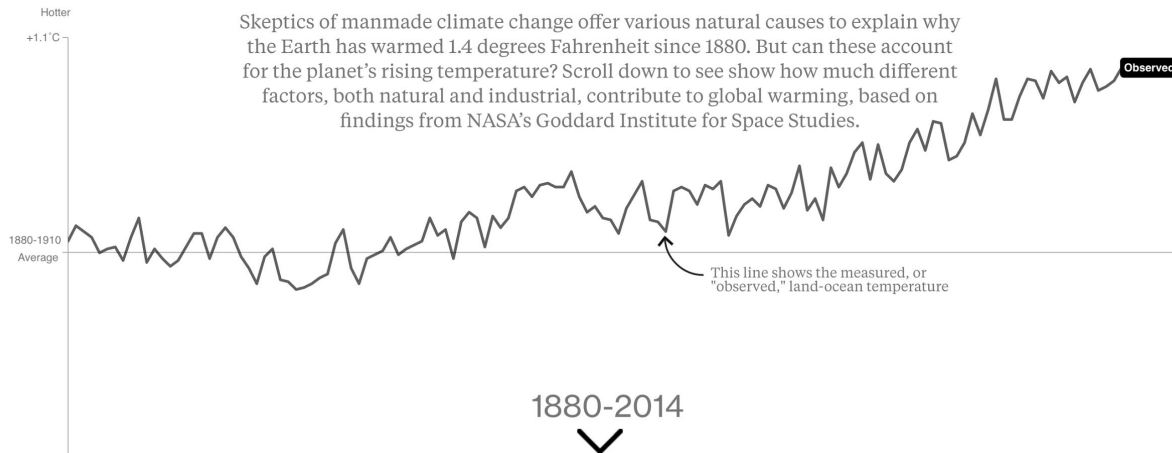
Bloomberg



What's Really Warming the World?

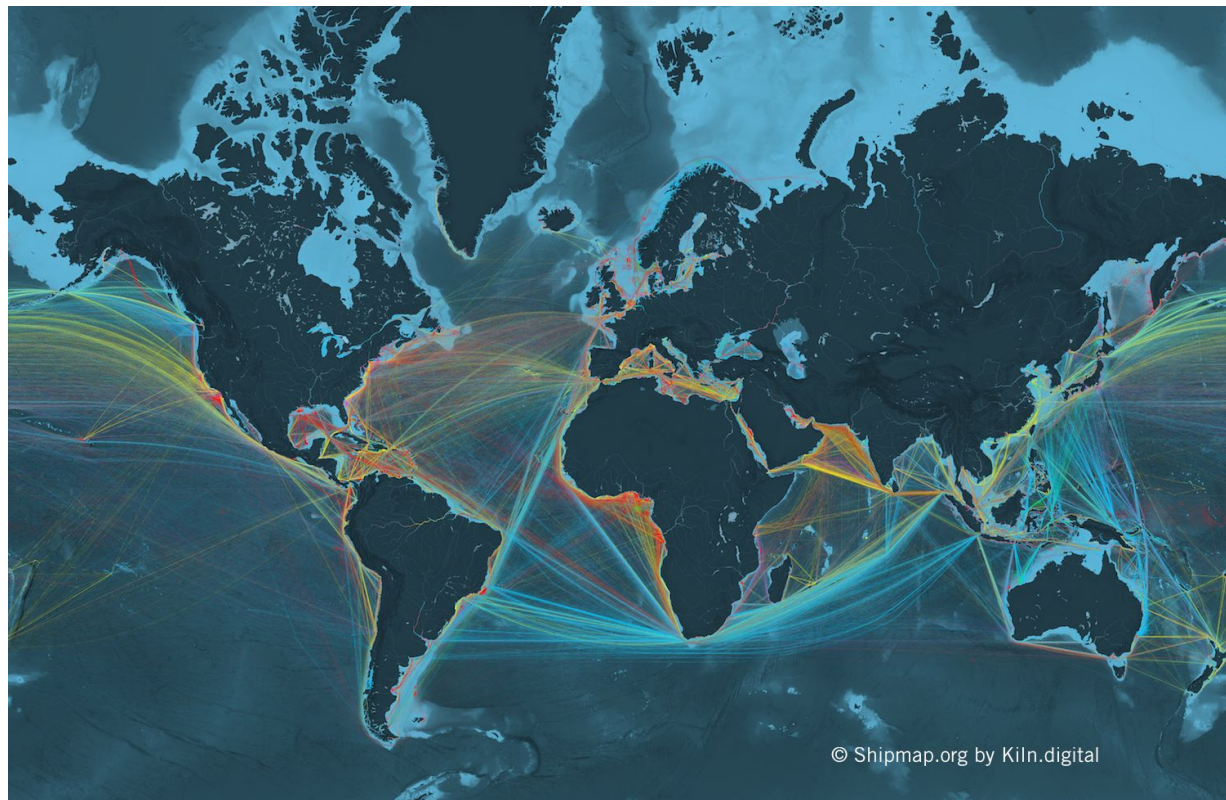
By Eric Roston and Blacki Miglozzi | June 24, 2015

Skeptics of manmade climate change offer various natural causes to explain why the Earth has warmed 1.4 degrees Fahrenheit since 1880. But can these account for the planet's rising temperature? Scroll down to see how much different factors, both natural and industrial, contribute to global warming, based on findings from NASA's Goddard Institute for Space Studies.



<http://www.bloomberg.com/graphics/2015-whats-warming-the-world/>

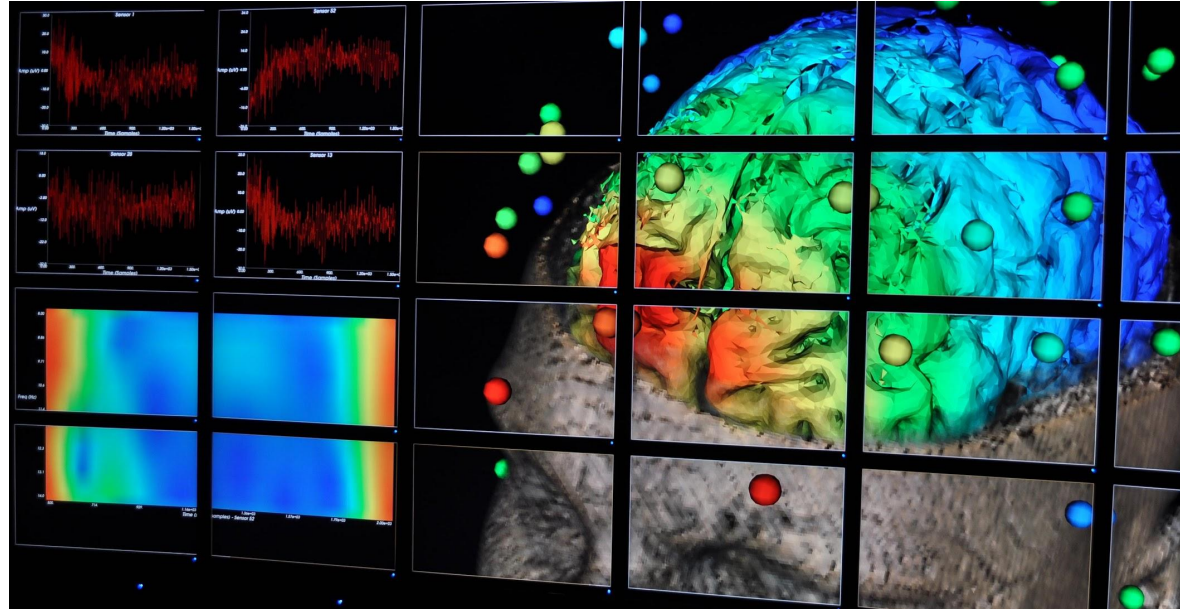
Les différents types de visualisation : Cartographie



—

Les différents types de visualisation :

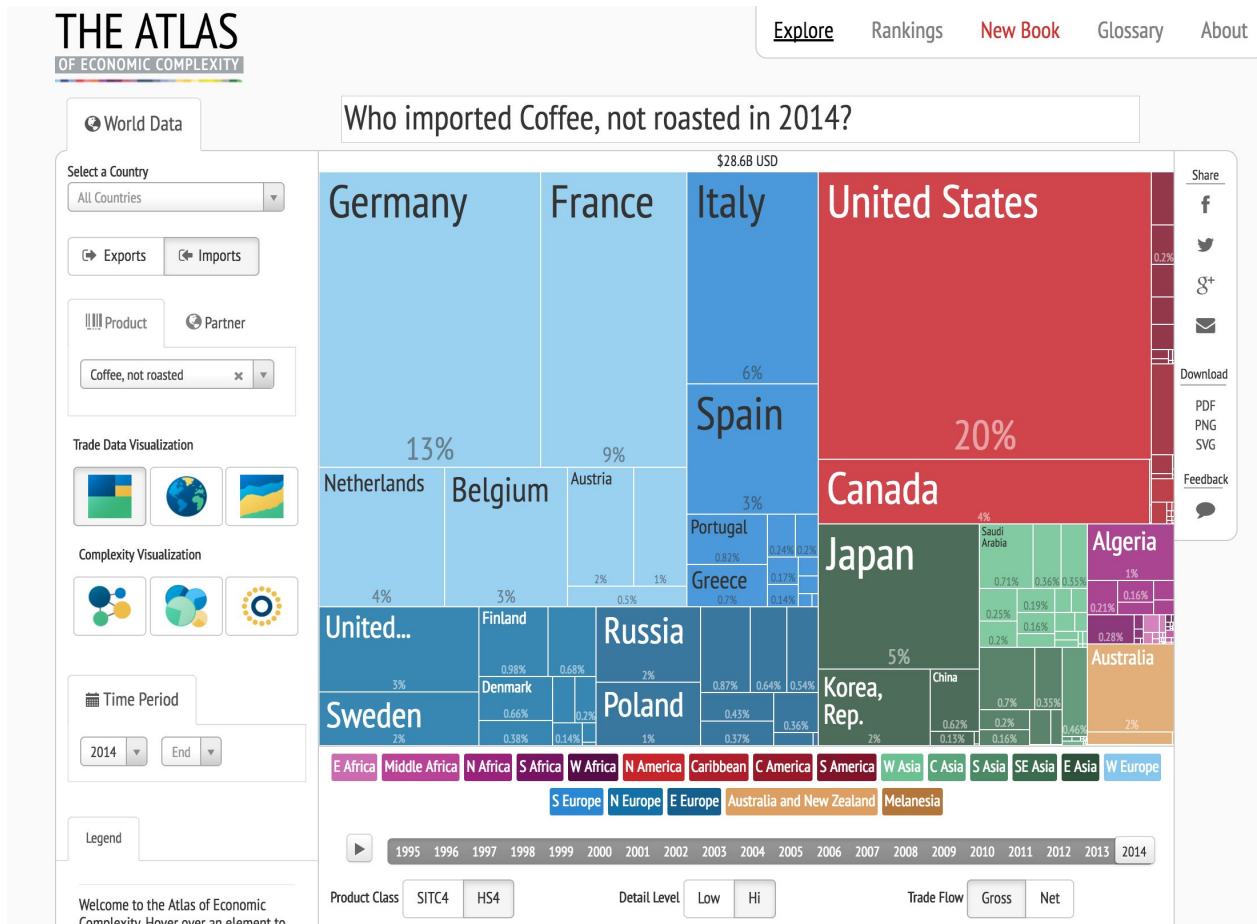
Visualisation scientifique



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

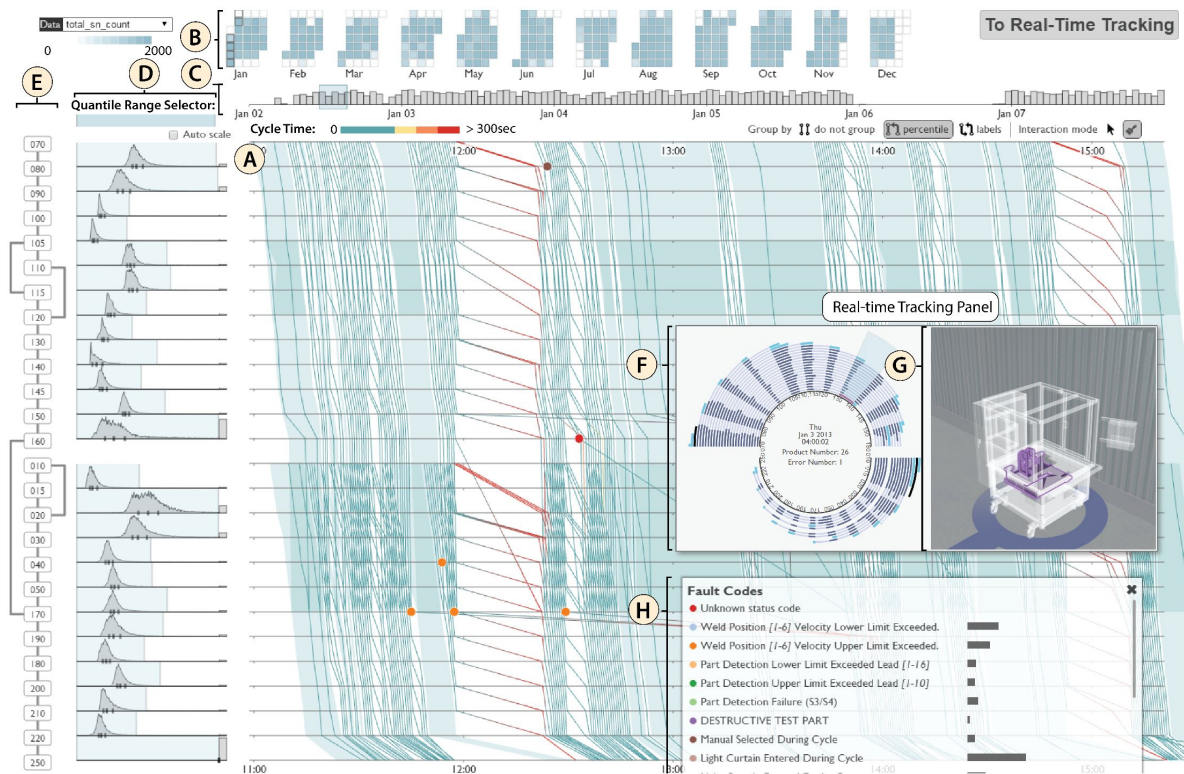
Les différents types de visualisation :

Visualisation d'information



Les différents
types de
visualisation :

Visual
Analytics

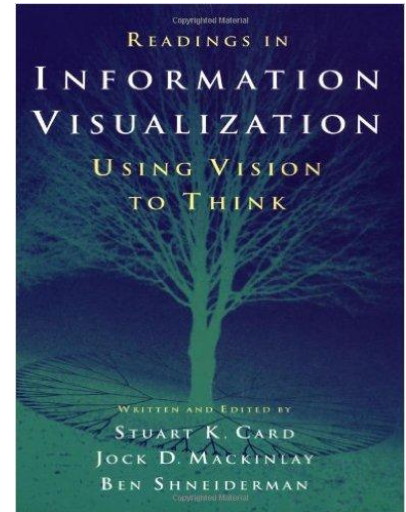


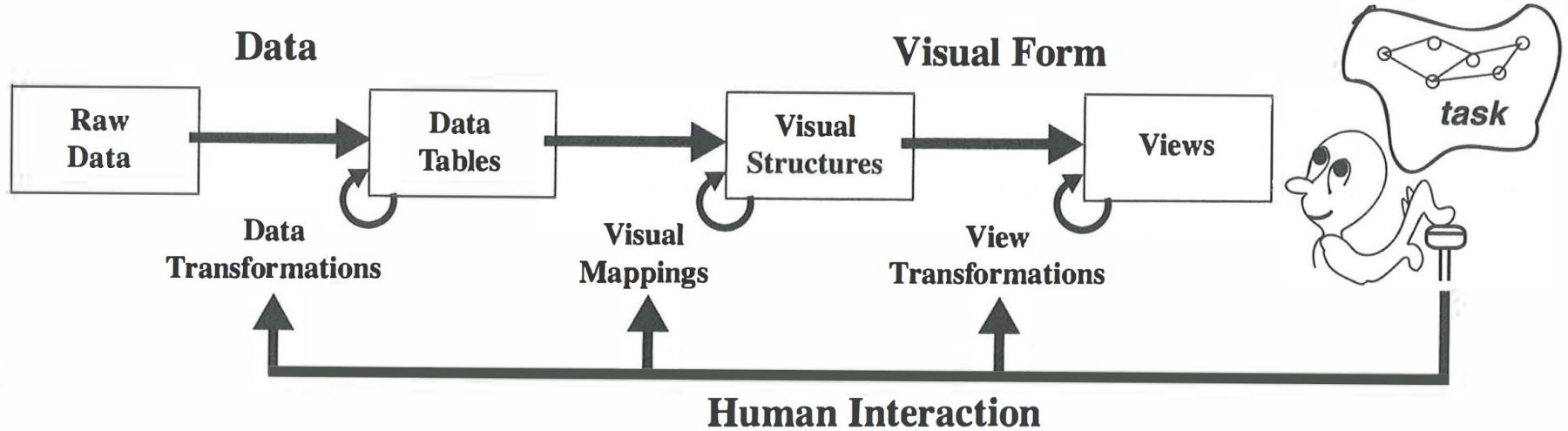
Définition

Visualisation d'information

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

Card, Mackinlay, & Shneiderman, 1999





Raw Data: idiosyncratic formats

Data Tables: relations (cases by variables) + metadata

Visual Structures: spatial substrates + marks + graphical properties

Views: graphical parameters (position, scaling, clipping, ...)

[Card, Mackinlay, Shneiderman, Readings in Information Visualization: Using Vision to Think, 1999]

Plan

- Présentation du cours
- Critique
- Pourquoi visualiser ?
- Qu'est ce que la visualisation
- Type de données
- Variables graphiques
- Mapping + visualisation pipeline
- Un classique

Les données

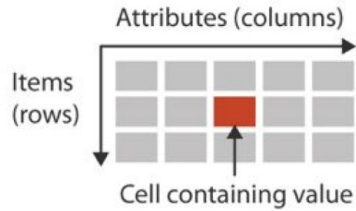
À la base de toute visualisation

Un bon designer de visualisation doit connaître :

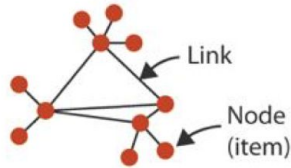
- Les propriétés des données
- Les méta-données associées
- Ce que les gens veulent tirer des données

Types de jeux de données

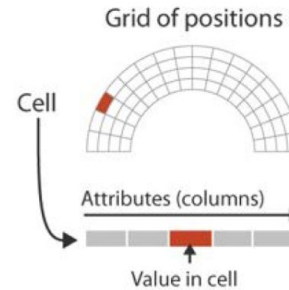
→ Tables



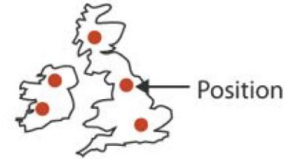
→ Networks



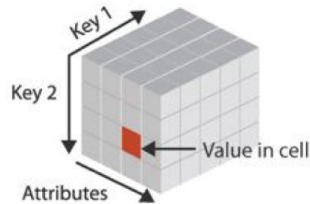
→ Fields (Continuous)



→ Geometry (Spatial)



→ *Multidimensional Table*



→ Trees



-> *Ce qu'on veut visualiser*

Type de données de base

Unités fondamentales

Constituent les jeux de données

- Item / élément
- Lien
- Attribut
- Position
- Grille

Exemple item (élément)/attribut

A	B	C	S	T	U
Order ID	Order Date	Order Priority	Product Container	Product Base Margin	Ship Date
3	10/14/06	5-Low	Large Box	0.8	10/21/06
6	2/21/08	4-Not Specified	Small Pack	0.55	2/22/08
32	7/16/07	2-High	Small Pack	0.79	7/17/07
32	7/16/07	2-High	Jumbo Box		7/17/07
32	7/16/07	2-High	Medium Box		7/18/07
32	7/16/07	2-High	Medium Box	0.65	7/18/07
35	10/23/07	4-Not Specified	Wrap Bag	0.52	10/24/07
35	10/23/07	4-Not Specified	Small Box	0.58	10/25/07
36	11/3/07	1-Urgent	Small Box	0.55	11/3/07
65	3/18/07	1-Urgent	Small Pack	0.49	3/19/07
66	1/20/05	5-Low	Wrap Bag	0.56	1/20/05
69		5 4-Not Specified	Small Pack	0.44	6/6/05
69		5 4-Not Specified	Wrap Bag	0.6	6/6/05
70	12/18/06	5-Low	Small Box	0.59	12/23/06
70	12/18/06	5-Low	Wrap Bag	0.82	12/23/06
96	4/17/05	2-High	Small Box	0.55	4/19/05
97	1/29/06	3-Medium	Small Box	0.38	1/30/06
129	11/19/08	5-Low	Small Box	0.37	11/28/08
130	5/8/08	2-High	Small Box	0.37	5/9/08
130	5/8/08	2-High	Medium Box	0.38	5/10/08
130	5/8/08	2-High	Small Box	0.6	5/11/08
132	6/11/06	3-Medium	Medium Box	0.6	6/12/06
132	6/11/06	3-Medium	Jumbo Box	0.69	6/14/06
134	5/1/08	4-Not Specified	Large Box	0.82	5/3/08
135	10/21/07	4-Not Specified	Small Pack	0.64	10/23/07
166	9/12/07	2-High	Small Box	0.55	9/14/07
193	8/8/06	1-Urgent	Medium Box	0.57	8/10/06
194	4/5/08	3-Medium	Wrap Bag	0.42	4/7/08

Élément et attribut

Élément :

- Entité individuelle, discrète.
- Ex: un patient, une voiture

Attribut :

- Propriété mesurée ou observée
- Ex: taille, pression sanguine (patient), vitesse (voiture)

Lien, Position et Grille

Lien

- Relation entre deux éléments
- Ex : “amitié sur Facebook”

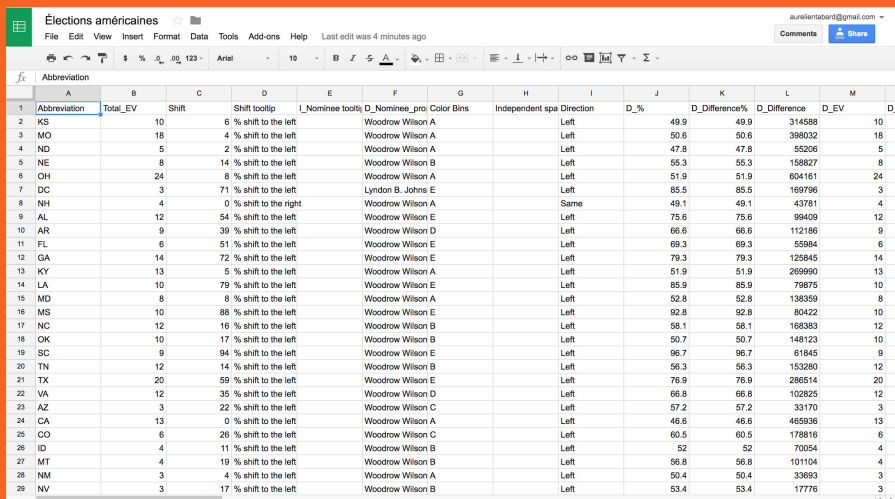
Position

- Données spatiales (en 2D ou 3D)
- Ex : latitude/longitude

Grille

- Stratégie d'échantillonnage pour données continues
- Ex: positions de stations météo

Données



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Abbreviation	Total_EV	Shift	Shift tooltip	I_Nominee tooltip	D_Nominee_pro	Color Bins	Independent spa	Direction	D_%	D_Difference%	D_Difference	D_EV	D_
2	KS		10	6 % shift to the left	Woodrow Wilson A				Left	49.9	49.9	314588	10	
3	MO		18	4 % shift to the left	Woodrow Wilson A				Left	50.6	50.6	398032	18	
4	ND		5	2 % shift to the left	Woodrow Wilson A				Left	47.8	47.8	55206	5	
5	NE		8	14 % shift to the left	Woodrow Wilson B				Left	55.3	55.3	158627	8	
6	OH		24	8 % shift to the left	Woodrow Wilson A				Left	51.9	51.9	604161	24	
7	DC		3	71 % shift to the left	Lyndon B. Johns E				Left	85.5	85.5	169796	3	
8	NH		4	0 % shift to the right	Woodrow Wilson A				Same	49.1	49.1	43781	4	
9	AL		12	54 % shift to the left	Woodrow Wilson E				Left	75.6	75.6	99409	12	
10	AR		9	39 % shift to the left	Woodrow Wilson D				Left	66.6	66.6	112186	9	
11	FL		6	51 % shift to the left	Woodrow Wilson E				Left	69.3	69.3	55984	6	
12	GA		14	72 % shift to the left	Woodrow Wilson E				Left	79.3	79.3	125845	14	
13	KY		13	5 % shift to the left	Woodrow Wilson A				Left	51.9	51.9	269990	13	
14	LA		10	79 % shift to the left	Woodrow Wilson E				Left	85.9	85.9	79875	10	
15	MD		8	8 % shift to the left	Woodrow Wilson A				Left	52.8	52.8	138359	8	
16	MS		10	88 % shift to the left	Woodrow Wilson E				Left	92.8	92.8	80422	10	
17	NC		12	16 % shift to the left	Woodrow Wilson B				Left	58.1	58.1	168383	12	
18	OK		10	17 % shift to the left	Woodrow Wilson B				Left	50.7	50.7	148123	10	
19	SC		9	94 % shift to the left	Woodrow Wilson E				Left	96.7	96.7	61845	9	
20	TN		12	14 % shift to the left	Woodrow Wilson B				Left	55.3	55.3	153280	12	
21	TX		20	59 % shift to the left	Woodrow Wilson E				Left	78.9	78.9	286514	20	
22	VA		12	35 % shift to the left	Woodrow Wilson D				Left	66.8	66.8	102825	12	
23	AZ		3	22 % shift to the left	Woodrow Wilson C				Left	57.2	57.2	33170	3	
24	CA		13	0 % shift to the left	Woodrow Wilson A				Left	46.6	46.6	465936	13	
25	CO		6	26 % shift to the left	Woodrow Wilson C				Left	60.5	60.5	178816	6	
26	ID		4	11 % shift to the left	Woodrow Wilson B				Left	52	52	70054	4	
27	MT		4	19 % shift to the left	Woodrow Wilson B				Left	56.8	56.8	101104	4	
28	NM		3	4 % shift to the left	Woodrow Wilson A				Left	50.4	50.4	33693	3	
29	NV		3	17 % shift to the left	Woodrow Wilson B				Left	53.4	53.4	17776	3	

Exercice : <https://goo.gl/5bPs9s>

Trouver à quoi correspond :

- Un item / un élément / une variable (indépendante)
- Un attribut / une dimension / une variable (dépendante) / une feature
- Les clés

Où est définie la sémantique de la table ?

Type d'échelles

Nominale (catégoriel)

- Fruits: pommes, oranges, ...

Ordinale (ordonné)

- Qualité d'un frigo: A+, A++, A+++ ...
- Peut être compté et ordonné mais pas mesuré

Intervalle (zéro arbitraire)

- Dates, longitude, latitude

Ratio (zéro fixé)

- Le zéro a un sens (rien)
- Mesure physique : poids, longueur, ...

Type d'échelles

Nominale (catégoriel)

- Opérations : $=$, $\neq$

Ordinale (ordonné)

- Opérations : $=$, $\neq$, $>$, $<$

Intervalle (zéro arbitraire)

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

ex : $[1989 - 1999] + [2002 - 2012]$

peut mesurer les distances

Ratio (zéro fixé)

- Opérations : $=$, $\neq$, $>$, $<$, $+$, $-$, $\times$, $\div$

ex : $10\text{kg} / 5\text{kg}$

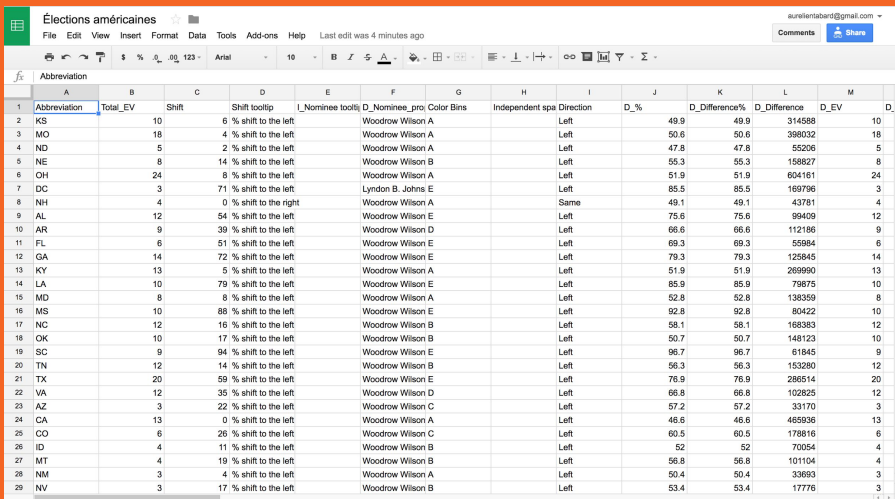
peut mesurer les proportions

Données

Exercice : <https://goo.gl/5bPs9s>

Trouver un type de données :

- Nominal / Catégoriel
- Ordinal / Ordonné
- Interval
- Ratio



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Abbreviation	Total_EV	Shift	Shift tooltip	I_Nominee tooltip	D_Nominee_pro	Color Bins	Independent spa	Direction	D_%	D_Difference%	D_Difference	D_EV	D_
2	KS		10	6 % shift to the left	Woodrow Wilson A				Left	49.9	49.9	314588	10	
3	MO		18	4 % shift to the left	Woodrow Wilson A				Left	50.6	50.6	398032	18	
4	ND		5	2 % shift to the left	Woodrow Wilson A				Left	47.8	47.8	55206	5	
5	NE		8	14 % shift to the left	Woodrow Wilson B				Left	55.3	55.3	158627	8	
6	OH		24	8 % shift to the left	Woodrow Wilson A				Left	51.9	51.9	604161	24	
7	DC		3	71 % shift to the left	Lyndon B. Johns E				Left	85.5	85.5	169796	3	
8	NH		4	0 % shift to the right	Woodrow Wilson A				Same	49.1	49.1	43781	4	
9	AL		12	54 % shift to the left	Woodrow Wilson E				Left	75.6	75.6	99409	12	
10	AR		9	39 % shift to the left	Woodrow Wilson D				Left	66.6	66.6	112186	9	
11	FL		6	51 % shift to the left	Woodrow Wilson E				Left	69.3	69.3	55984	6	
12	GA		14	72 % shift to the left	Woodrow Wilson E				Left	79.3	79.3	125845	14	
13	KY		13	5 % shift to the left	Woodrow Wilson A				Left	51.9	51.9	269990	13	
14	LA		10	79 % shift to the left	Woodrow Wilson E				Left	85.9	85.9	79875	10	
15	MD		8	8 % shift to the left	Woodrow Wilson A				Left	52.8	52.8	138359	8	
16	MS		10	88 % shift to the left	Woodrow Wilson E				Left	92.8	92.8	80422	10	
17	NC		12	16 % shift to the left	Woodrow Wilson B				Left	58.1	58.1	168383	12	
18	OK		10	17 % shift to the left	Woodrow Wilson B				Left	50.7	50.7	148123	10	
19	SC		9	94 % shift to the left	Woodrow Wilson E				Left	96.7	96.7	61845	9	
20	TN		12	14 % shift to the left	Woodrow Wilson D				Left	58.3	58.3	153280	12	
21	TX		20	59 % shift to the left	Woodrow Wilson E				Left	78.9	78.9	286514	20	
22	VA		12	35 % shift to the left	Woodrow Wilson D				Left	66.8	66.8	102825	12	
23	AZ		3	2 % shift to the left	Woodrow Wilson C				Left	57.2	57.2	33170	3	
24	CA		13	0 % shift to the left	Woodrow Wilson A				Left	46.6	46.6	465936	13	
25	CO		6	26 % shift to the left	Woodrow Wilson C				Left	60.5	60.5	178816	6	
26	ID		4	11 % shift to the left	Woodrow Wilson B				Left	52	52	70554	4	
27	MT		4	19 % shift to the left	Woodrow Wilson D				Left	56.8	56.8	101104	4	
28	NM		3	4 % shift to the left	Woodrow Wilson A				Left	50.4	50.4	33693	3	
29	NV		3	17 % shift to the left	Woodrow Wilson B				Left	53.4	53.4	17776	3	

Modèle de données vs. conceptuel

Modèle de données (description bas niveau)

- Flottants : 32.5, 54.0, -17.3

Modèle conceptuel (construction mentale)

- Température

Type de données

- Valeur continue avec 1 chiffre significatif (Quantitatif)
- Chaud - tiède - froid - glacé (Ordinal)
- Brulé / pas brulé (Nominal)

Plan

- Présentation du cours
- Critique
- Pourquoi visualiser ?
- Qu'est ce que la visualisation
- Type de données
- **Variables graphiques**
- Mapping + visualisation pipeline
- Un classique

Les variables de Jacques Bertin

Cartographe français,
auteur de la sémiologie graphique



LES VARIABLES DE L'IMAGE											
POINTS			LIGNES			ZONES					
XY 2 DIMENSIONS DU PLAN											
TAILLE											
VALEUR											
LES VARIABLES DE SÉPARATION DES IMAGES											
GRAIN											
COULEUR											
ORIENTATION											
FORME											

Marques simples

Munzner, 2014,
Visualization Analysis and Design.

➔ Points



➔ Lines



➔ Areas



Canaux visuals

Munzner, 2014,
Visualization Analysis and Design.

→ Position

→ Horizontal



→ Vertical



→ Both



→ Color



→ Shape



→ Tilt



→ Size

→ Length



→ Area



→ Volume



Plan

- Présentation du cours
- Critique
- Pourquoi visualiser ?
- Qu'est ce que la visualisation
- Type de données
- Variables graphiques
- Mapping + visualisation pipeline
- Un classique

Mapping

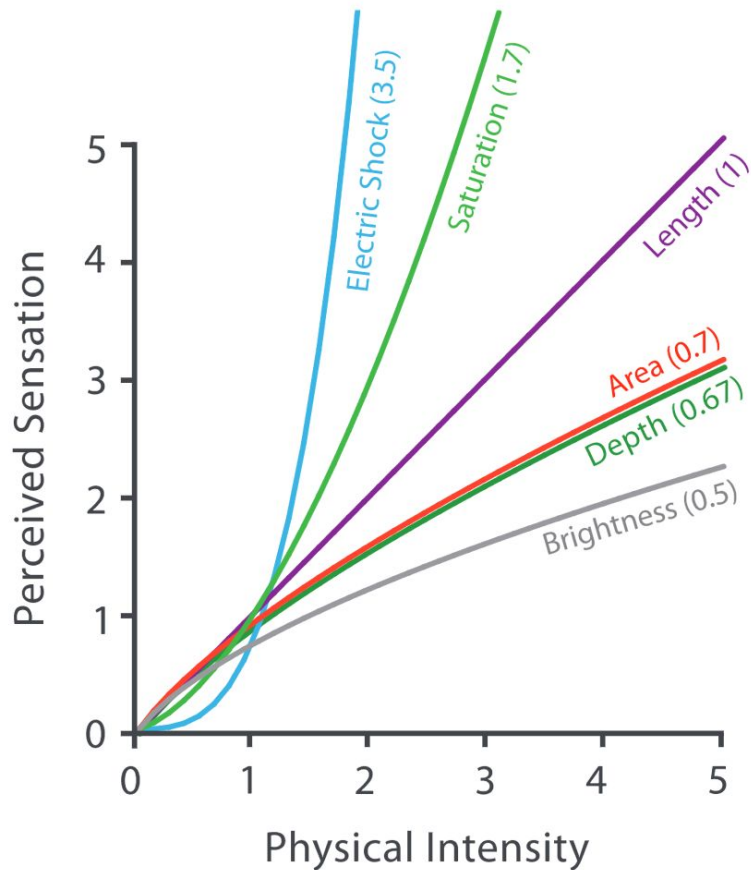
- Le travail de base consiste à mapper des données → des marques graphiques et propriétés
- Ensuite rajouter de l'interaction pour naviguer dans et manipuler les données

Question:

- Quels mapping choisir ?
- Quelles marques pour quelles données ?

— Efficacité de la perception humaine

Steven's Psychophysical Power Law: $S = I^N$



Efficacité des canaux

Munzner, 2014,
Visualization Analysis and Design.

➔ **Magnitude Channels: Ordered Attributes**



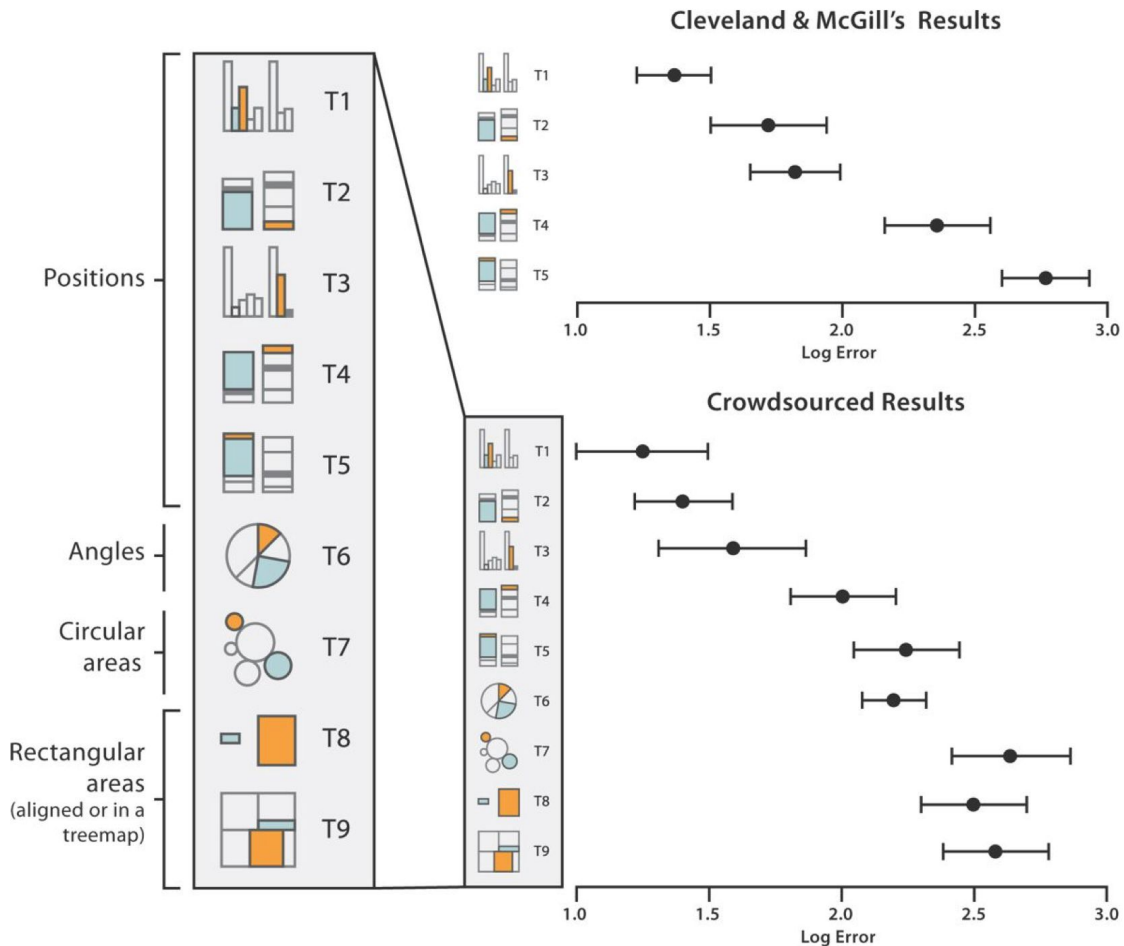
➔ **Identity Channels: Categorical Attributes**

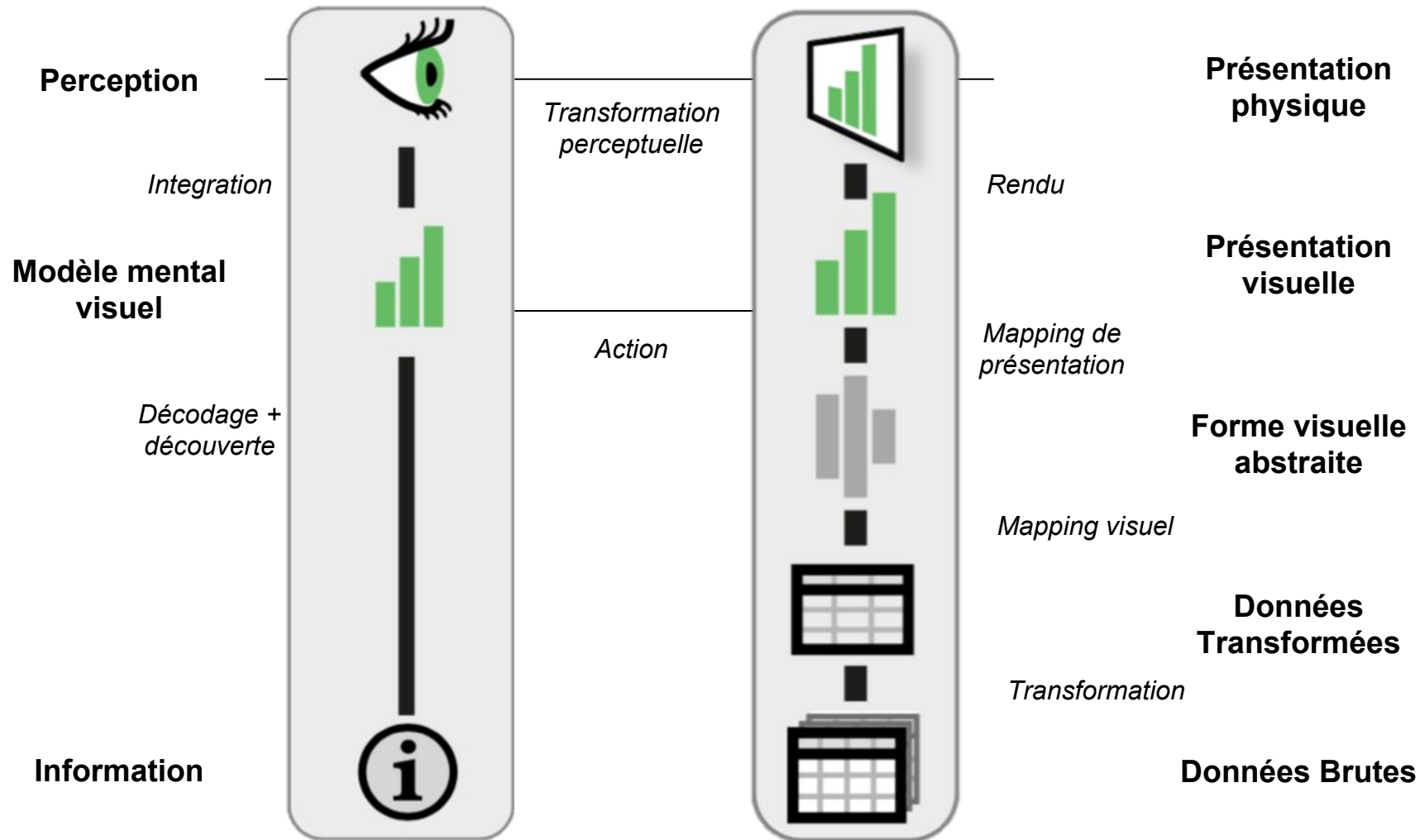


[VAD Fig 5.1]

Taux d'erreurs des canaux

Cleveland et McGill, 84
Heer et Bostock, 10





Plan

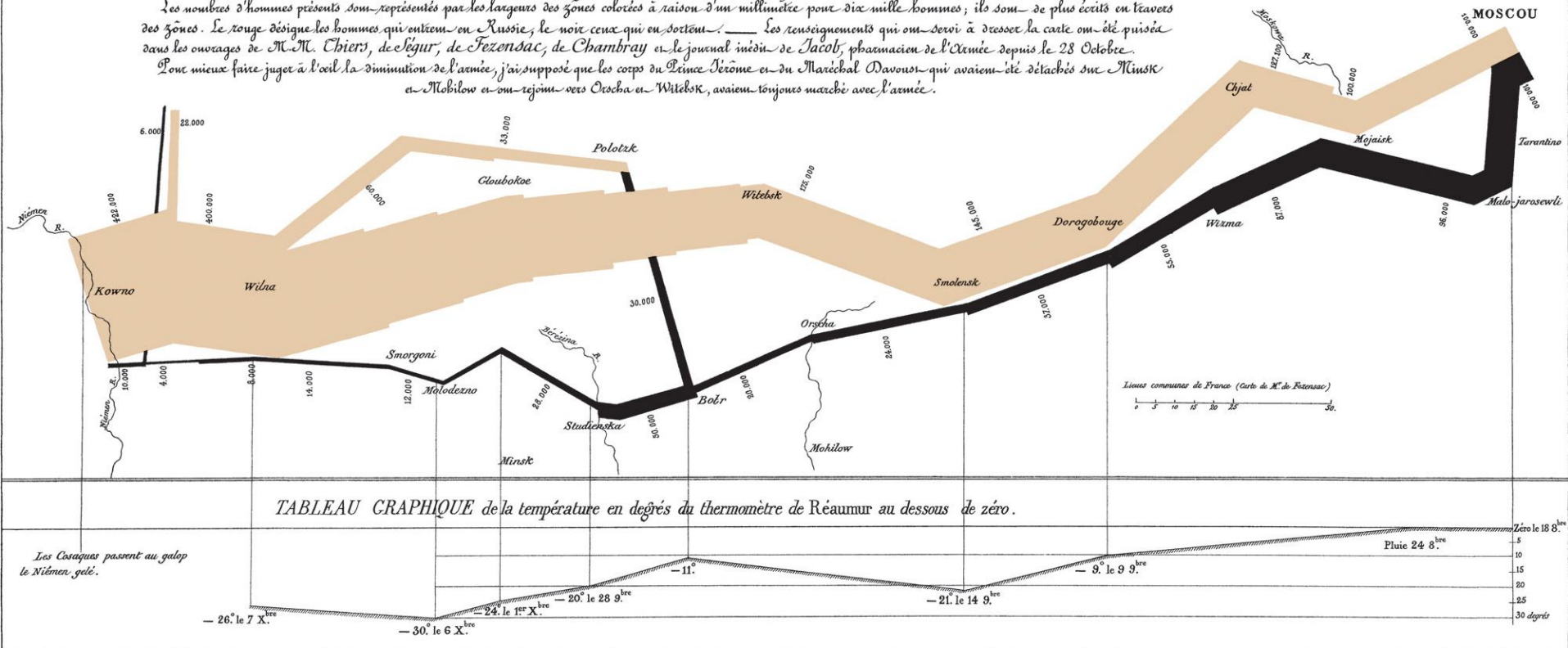
- Présentation du cours
- Critique
- Pourquoi visualiser ?
- Qu'est ce que la visualisation
- Type de données
- Variables graphiques
- Mapping + visualisation pipeline
- Un classique

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

Dressée 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 Fégur, de Fezensac, 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 Davoust qui avaient été détachés sur Minsk et Mohilow en ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.



Joseph Minard 1869 : Perte Napoléonienne de la campagne de Russie (diagramme de Sankey)

BILAN

Bilan

- Présentation du cours
- Critique
- Pourquoi visualiser ?
- Qu'est ce que la visualisation ?
- Type de données
- Variables graphiques
- Mapping + visualisation pipeline

Choix d'articles

<https://lyondataviz.github.io/teaching/lyon1-m2/2017/articles.html>

Projets

<https://lyondataviz.github.io/teaching/lyon1-m2/2017/projets.html>

PAUSE

Exercice Tableau

Visualiser avec Tableau les résultats
des élections présidentielles américaines

<https://lyondataviz.github.io/teaching/lyon1-m2/2017/tp1.html>

[http://www.dummies.com/programming/big-data/
big-data-visualization/tableau-for-dummies-cheat-sheet/](http://www.dummies.com/programming/big-data/big-data-visualization/tableau-for-dummies-cheat-sheet/)

Dimensions vs. measures

Dimensions:

- Discrete variables describing data
- Dates, categories of values (independent vars)

Measures:

- Data values that can be aggregated
- Numbers to be analyzed (dependent vars)
- Aggregate as sum, count, average, std. deviation