

---

---

# Introduction à la visualisation de données

EIOM

---

# 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

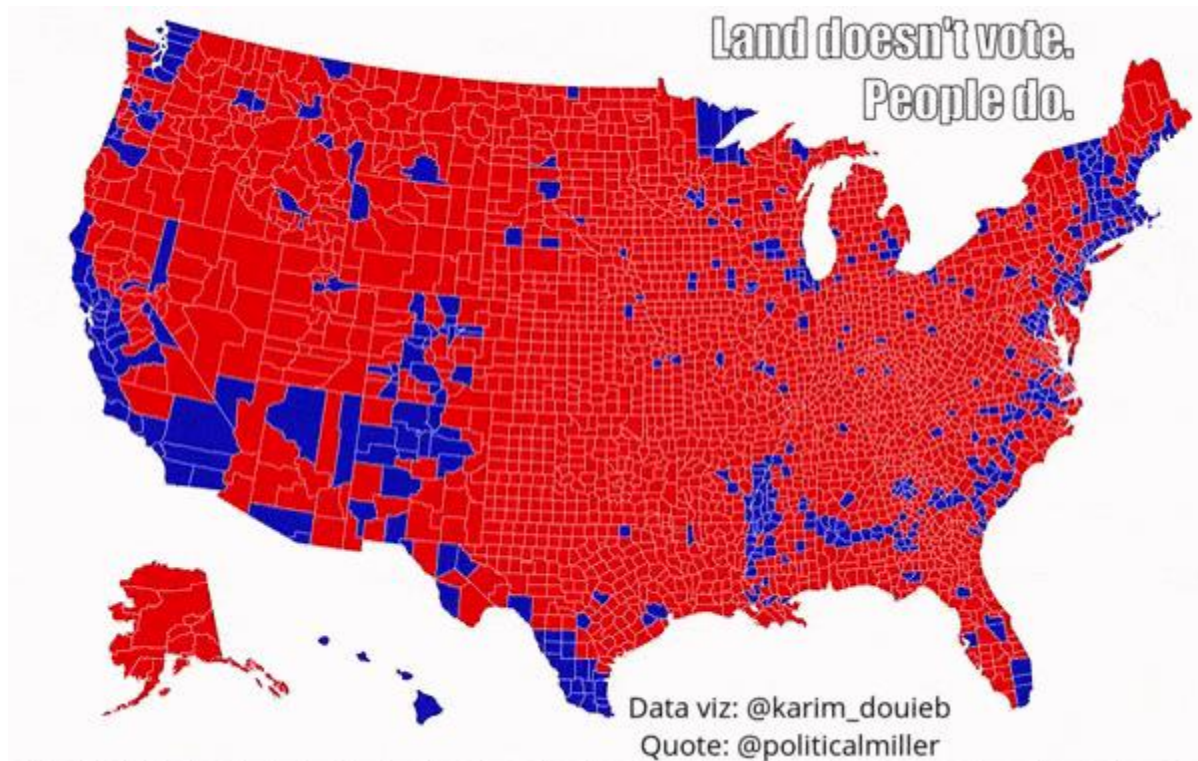
5

questions

\_\_\_\_\_

- \_\_\_\_\_





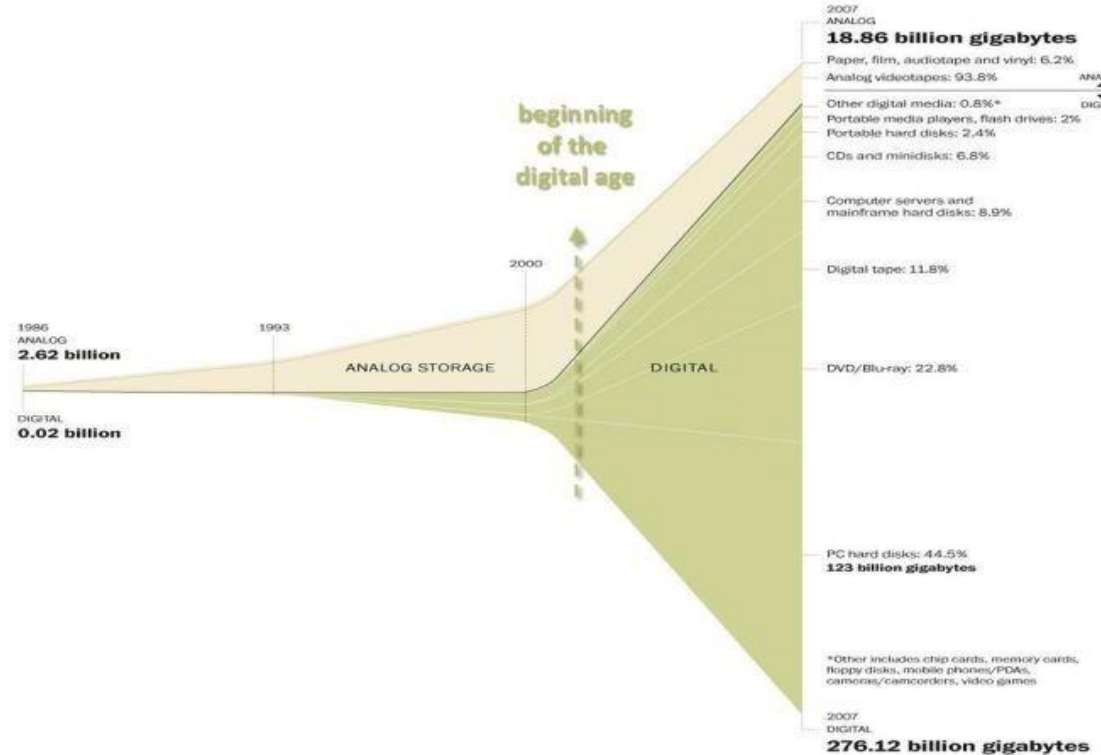
# Plan

- 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

---



---

# 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](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

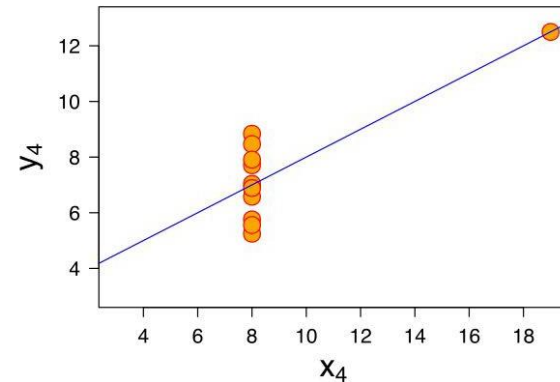
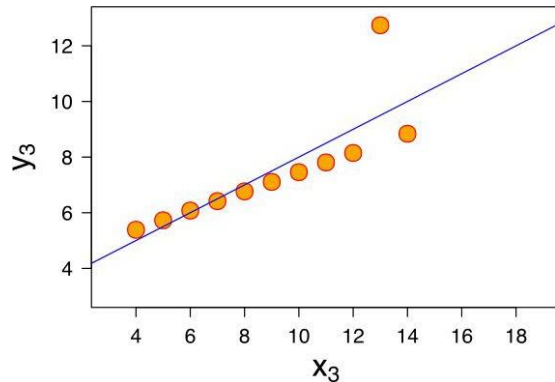
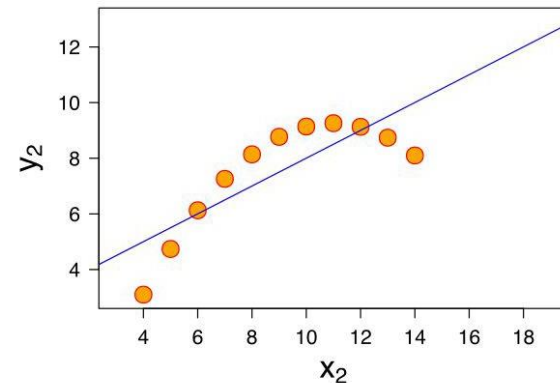
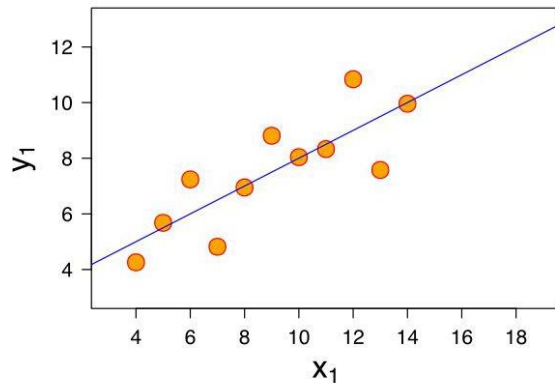
Régression linéaire:

$y = 3.00 + 0.500x$

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

Le Quartet d'Anscombe

[https://en.wikipedia.org/wiki/Anscombe%27s\\_quartet](https://en.wikipedia.org/wiki/Anscombe%27s_quartet)

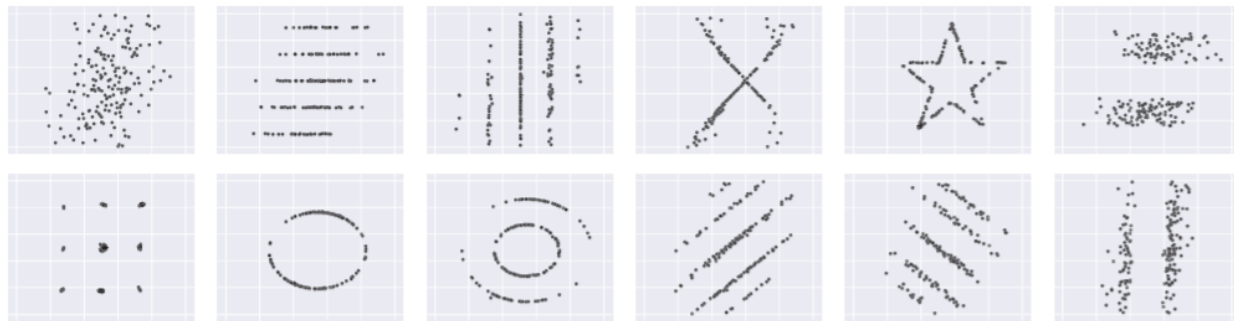


---

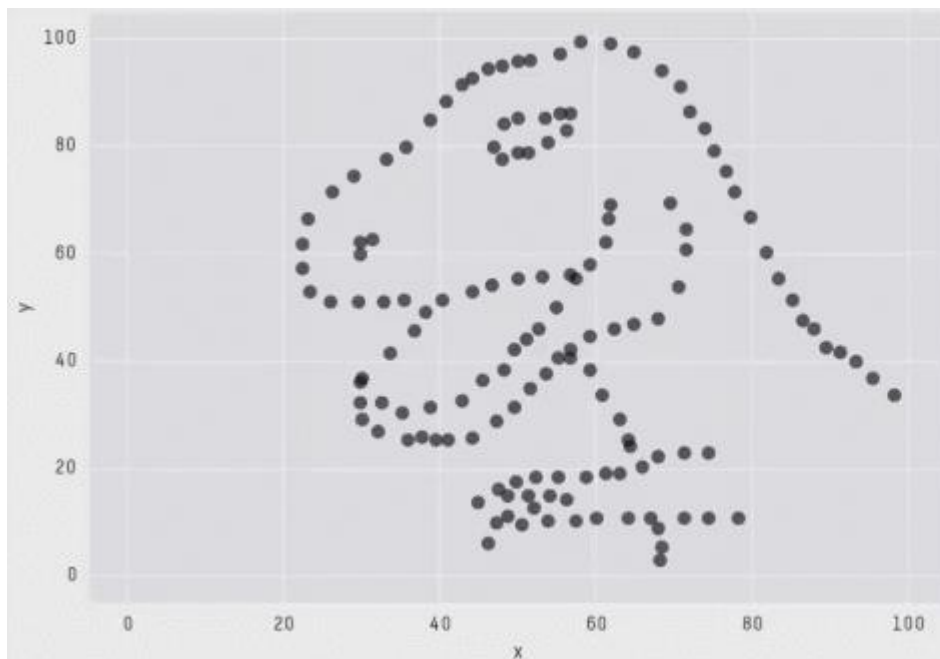
# Même stats mais different!

Les données (x,y) ont les mêmes statistiques descriptives!

( $\bar{x}=54.02$ ,  $\bar{y}=48.09$ ,  $sdx = 14.52$ ,  $sd_y = 24.79$ , *Pearson's*  $r = +0.32$ )



# En image



X Mean: 54.2659224  
Y Mean: 47.8313999  
X SD : 16.7649829  
Y SD : 26.9342120  
Corr. : -0.0642526

# Pourquoi visualiser ?

## **Exercice**

Brainstorming sur l'utilité  
de la visualisation de  
données.

binome

5 raisons

---

# Les trois raisons de la visualisation

---

Enregistrer de l'information

- Plan, photo, image

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](http://shop.gestalten.com/photoviz.html) – Photoviz <http://shop.gestalten.com/photoviz.html>

# Faciliter le raisonnement

Épidémie de Choléra à Londres (1854)

Analyse de données visuelle pour comprendre le problème

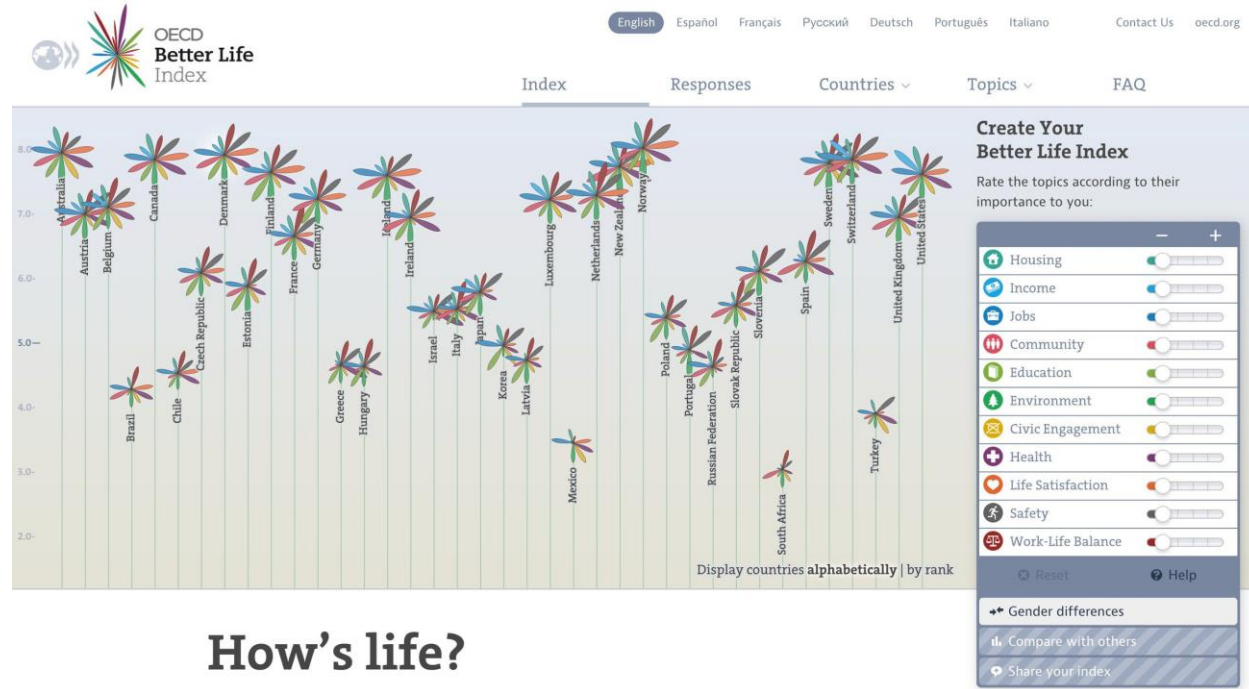
[https://fr.wikipedia.org/wiki/%C3%89pid%C3%A9mie\\_de\\_chol%C3%A9ra\\_de\\_Broad\\_Street\\_\(1854\)](https://fr.wikipedia.org/wiki/%C3%89pid%C3%A9mie_de_chol%C3%A9ra_de_Broad_Street_(1854))



John Snow, 1854

# Transmettre de l'information

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



# Pourquoi la visualisation est difficile ?

## Exercice

Visualiser les quantités suivantes :

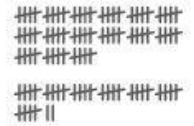
75 et  
37

75, 37

*a*

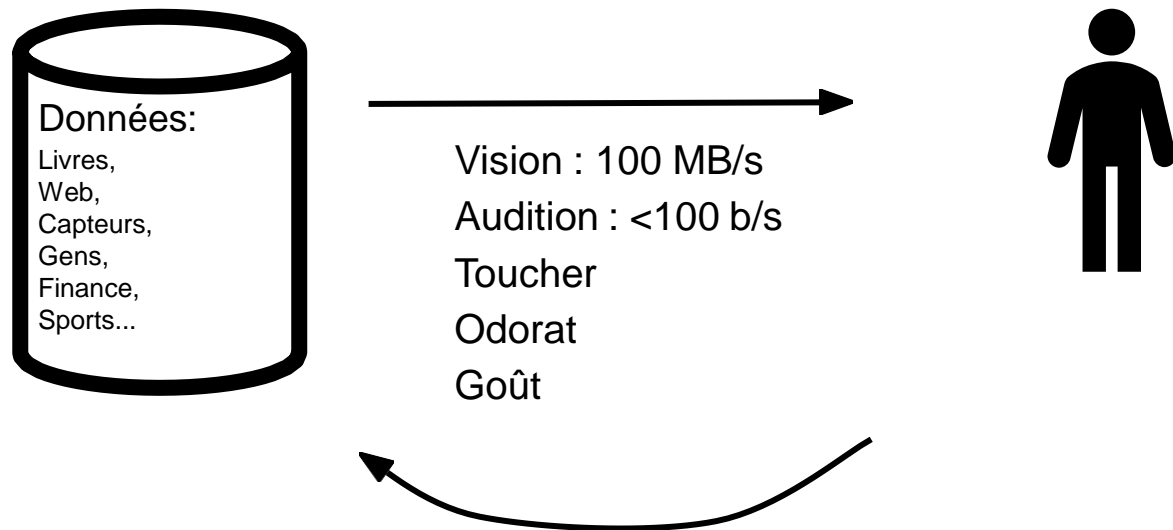


*b*



*c*

# 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

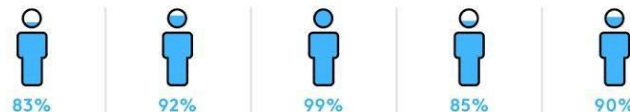
WASHING HABITS ARE DIFFERENT ACROSS THE GLOBE



#### SHOWER vs. BATH

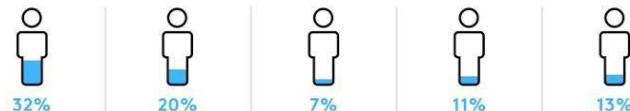
SHOWERING IS THE MOST COMMON METHOD OF WASHING

% 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

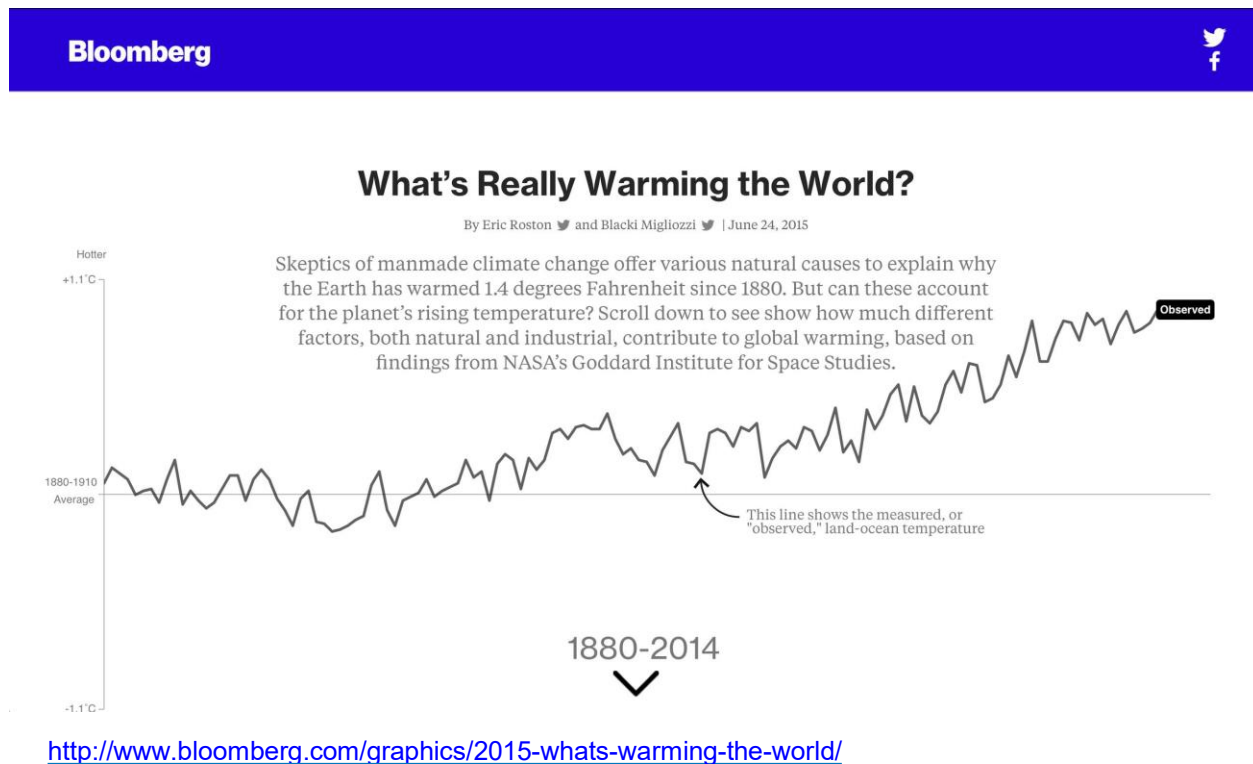


PEOPLE IN BRAZIL STILL TAKE MORE SHOWERS

Average Number Of Showers Per Week



# Les différents types de visualisation : Storytelling



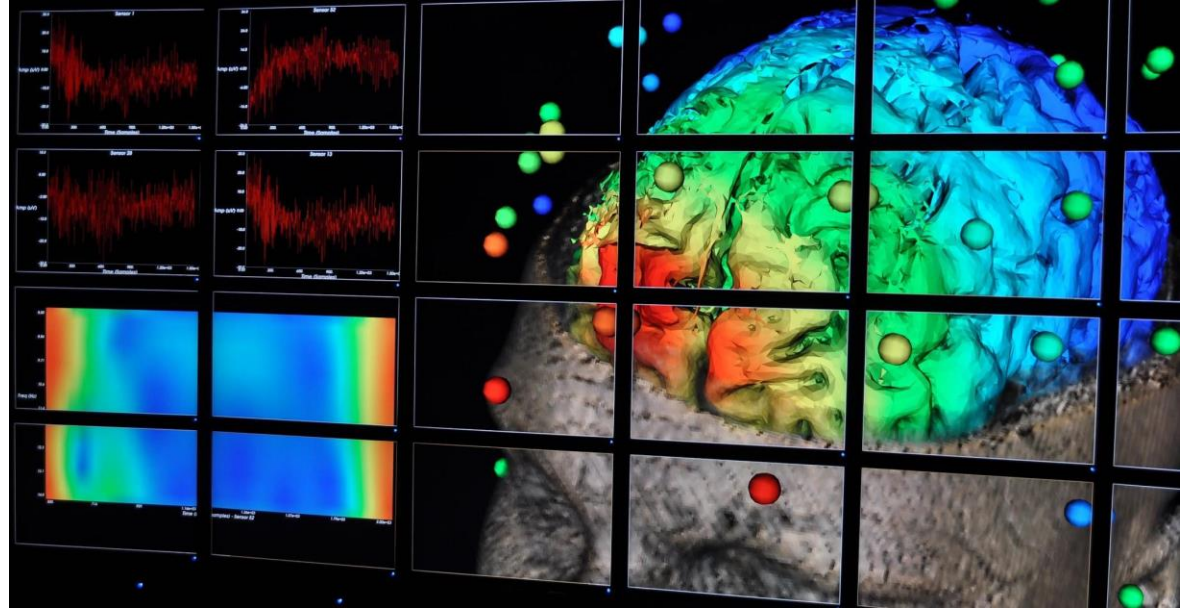


---

# Les différents types de visualisation : Cartographie



# Les différents types de visualisation Visualisation scientifique

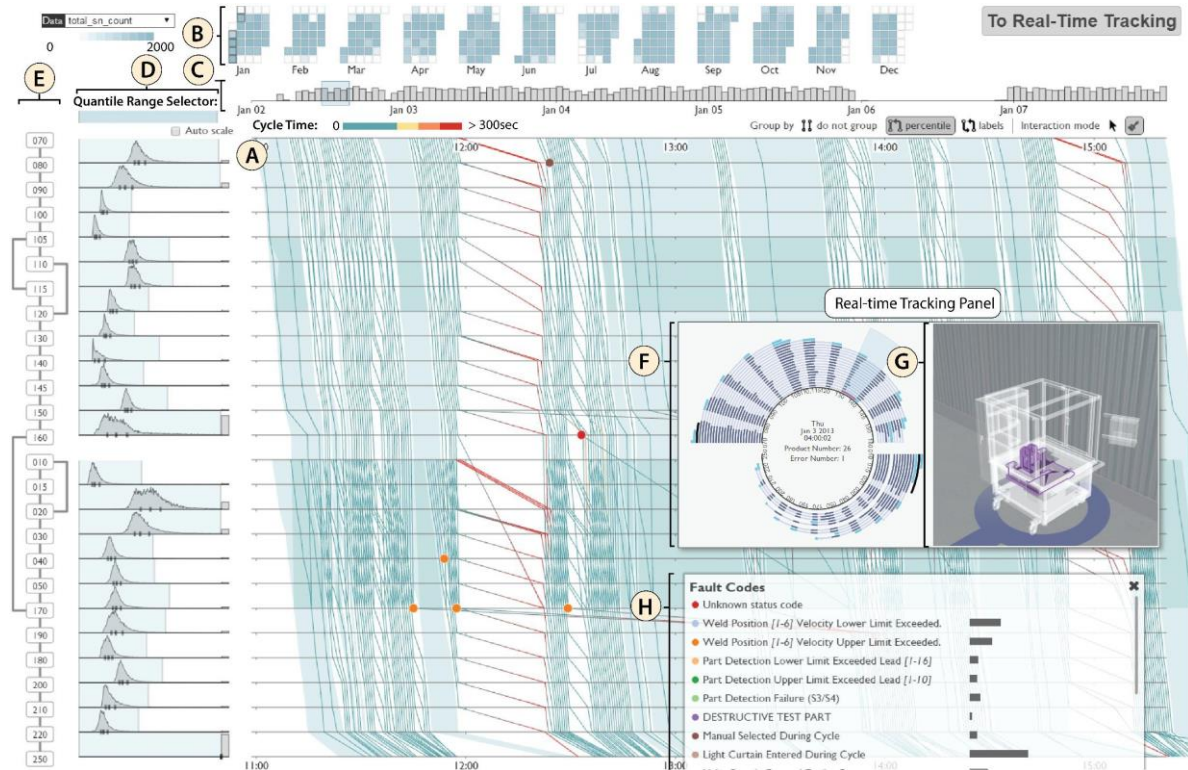


VisTrails [https://www.nsf.gov/discoveries/disc\\_images.jsp?cntn\\_id=114322&org=NSF](https://www.nsf.gov/discoveries/disc_images.jsp?cntn_id=114322&org=NSF)

# Visualisation d'information



# Les différents types de visualisation : Visual Analytics





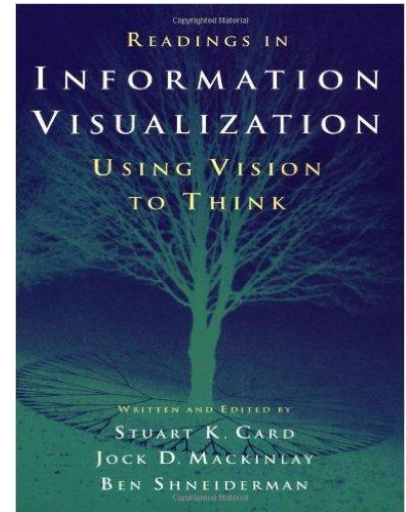
---

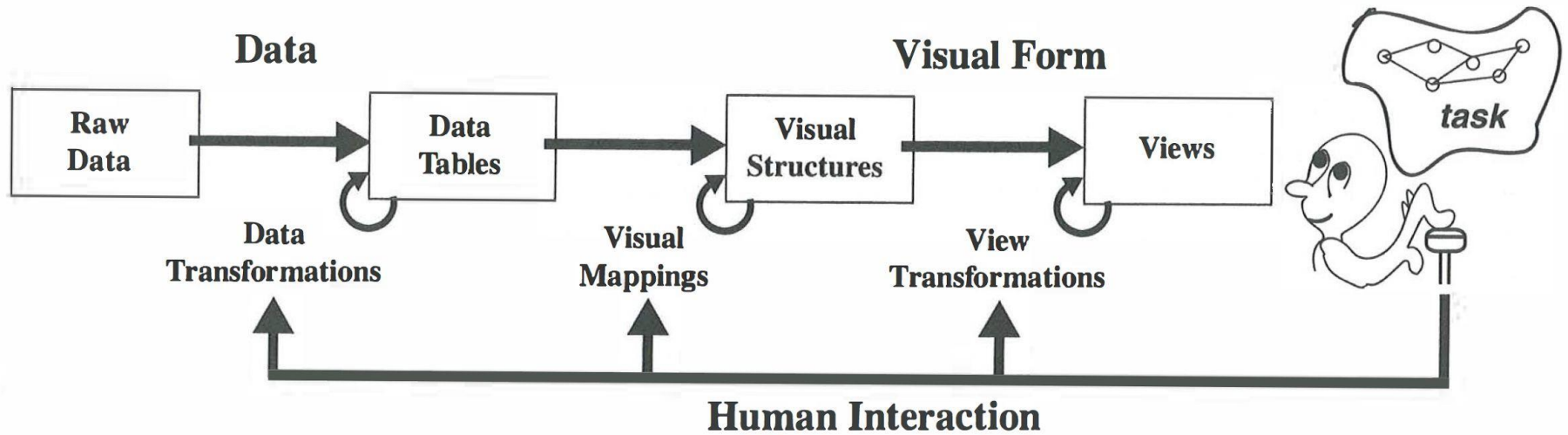
# Définition

## 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

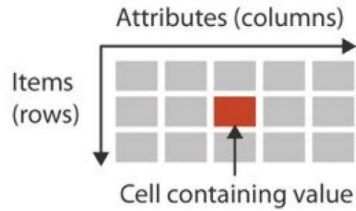
Un bon designer de visualisation doit connaître :

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

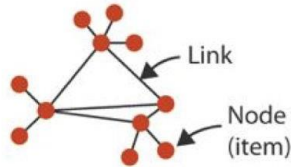


# Types de jeux de données

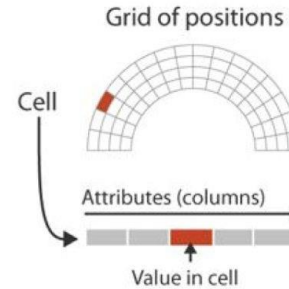
→ Tables



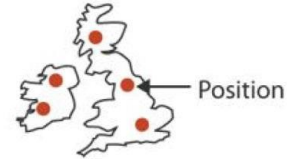
→ Networks



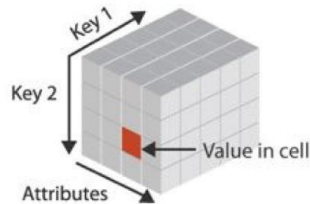
→ Fields (Continuous)



→ Geometry (Spatial)



→ Multidimensional Table



→ Trees



-> *Ce qu'on veut visualiser*

---

# Type de données de base

Unités fondamentales

Constituent les jeux de donnés

- 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)

---

# 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 : poid, longueur, ...

# Plan

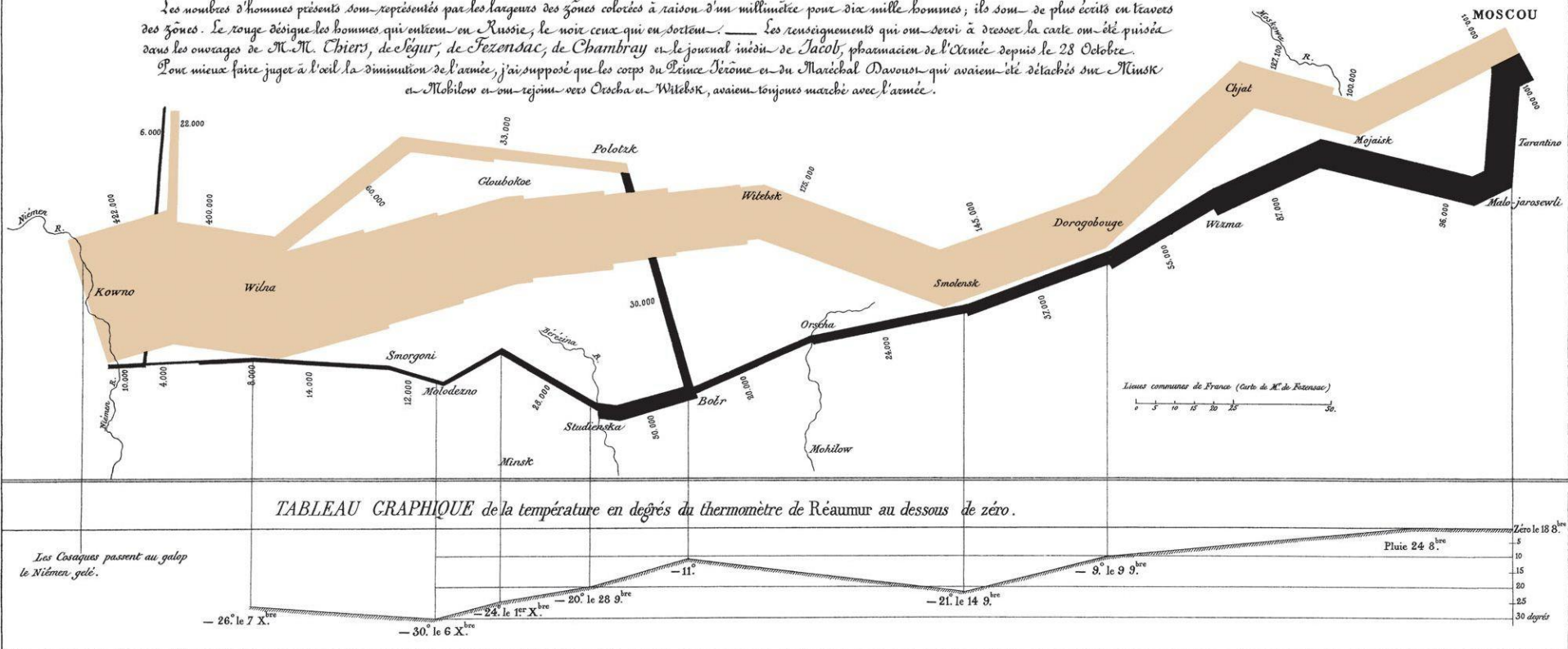
- Présentation du cours
  - Critique
  - Pourquoi visualiser ?
  - Qu'est ce que la visualisation
  - Type de données
  - Variables graphiques
  - Mapping + visualisation pipeline
  - **Un classique**
-

# 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 Fézensac, 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.



Autog. par Ragnier, 8. Par. S<sup>te</sup> Marie 5<sup>th</sup> G<sup>th</sup> à Paris.

Imp. Lith. Ragnier et Doucet.

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