# La Data Visualisation

Meetup Big Data – Data Science Montpellier

Mardi 29 Mai 2018







### Alexia Audevart

Data & Enthusiasm

@aaudevart



Founder of datactik
Data Scientist



President of Toulouse Data Science meet-up



Co-organizer of Toulouse DevFest conference

"Il existe 3 catégories de personnes : celles qui voient, celles qui voient quand on leur montre, celles qui ne voient jamais"

Léonard de Vinci

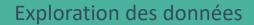
# Sommaire

- 1 Définition
- 2 L'efficacité graphique selon Jacques Bertin
- 3 L'excellence graphique d'Edward Tufte
- 4 Perception Visuelle
- 5 Le Story Telling
- 6 Good or Bad?

### Définition de la data visualisation?

Ensemble de méthodes et techniques de représentation des données sous forme visuelle

### Quand utilise-t-on la data visualisation?



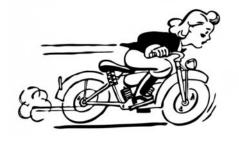


#### Représentation des données



Communication implies simplification, Data exploration implies exhaustivity

## 1 image vaut 1000 mots



# 60 000 fois plus

Rapide d'analyser une image que du texte



# 85% du cerveau

Dédié à l'analyse des images

### Le raisonnement humain

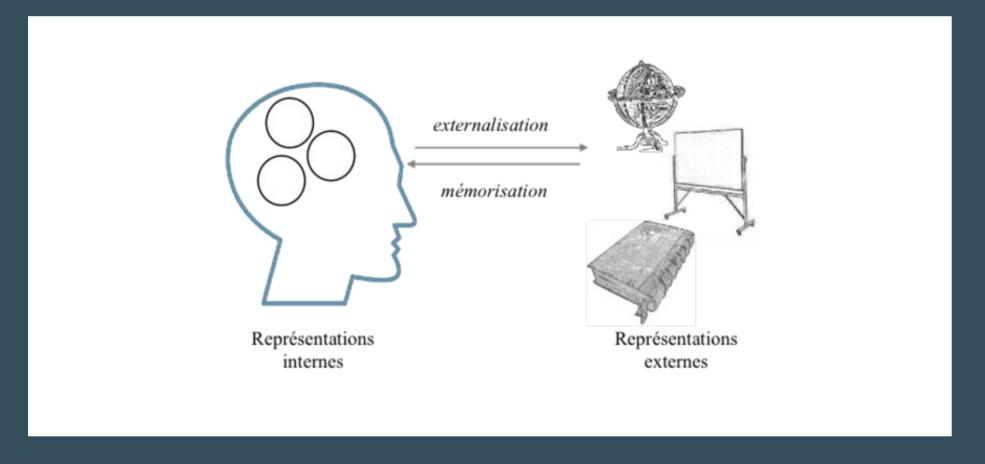
Besoin de raisonner, à un certain niveau de difficulté et pour une certaine quantité d'informations

=> utilisation de supports externes

Ce principe est nommé la cognition externe

## Le raisonnement humain

### 2 Phases



### Des premières cartes au premier graphiques

Frise chronologique: Lien

- Première représentation graphique 100 avant JC
  - Essentiellement des cartographies
- 18ème siècle : représentation de données quantitatives
  - 1765 : time line Joseph Priestley
  - 1786 : William Playfair inventeur de 3 graphiques
  - Série chronologique
  - Histogramme
  - Camembert
- 19ème siècle :
  - Apparition de la couleur
  - Diagramme en 3D
  - Diagramme de flux

A partir du 20ème siècle, opposition de 2 conceptions de représentation de données : démarche artistique vs vision plus scientifique

## Différents types de variables

### Variables qualitatives ou catégorielles

### expriment une qualité

#### Variables Nominales

=> Représente des catégories que l'on nomme avec un label

#### Exemple:

- signe astrologique
- liste des pays

Opération : = , #

#### Variables Ordinales

=> Catégories naturellement ordonnées = le rang

#### Exemple:

• un peu, moyen, beaucoup, à la folie

<u>Opération</u>: = , ≠, <, >

### Différents types de variables

#### Variables quantitatives

#### contiennent des valeurs mesurables

#### Variables Discrètes

- ⇒ Variables numériques ayant des valeurs dénombrables entre deux valeurs
- ⇒ Variables ayant des valeurs finies

#### Exemple:

- Nombre de voitures dans la rue
- Nombre de personnes majeures
- Nombre de candidats à l'examen

#### **Variables Continues**

=> Variables numériques ayant un nombre infini de valeurs entre deux valeurs.

#### Exemple:

- La vitesse d'une voiture
- La taille, le poids d'une personne
- Le temps de réalisation d'une tâche

# Sommaire

- 1 Définition
- 2 L'efficacité graphique selon Jacques Bertin
- 3 L'excellence graphique d'Edward Tufte
- 4 Perception Visuelle
- 5 Le Story Telling
- 6 Good or Bad?

### Les Travaux de Jacques Bertin



### Jacques Bertin (1918-2010)

- Père fondateur de la visualisation de l'information
- Auteur de l'ouvrage :

Sémiologie graphique

Les diagrammes, les réseaux, les cartes

1<sup>ère</sup> publication : 1967



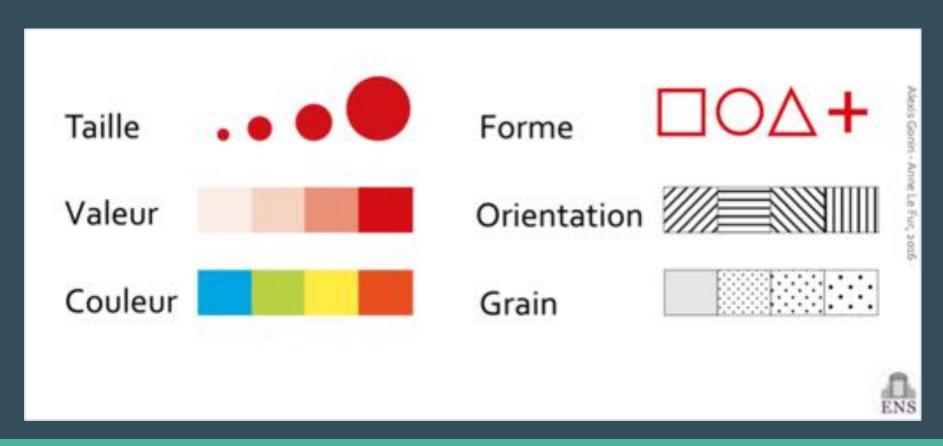
Sémiologie (étude de la signification) graphique répond à 2 questions :

- Dans quel cas faut-il faire un dessin ?
- Quel dessin faut-il faire?

Le 1er à formaliser un corpus scientifique structuré dans une perspective d'analyse de données

## Les Travaux de Jacques Bertin

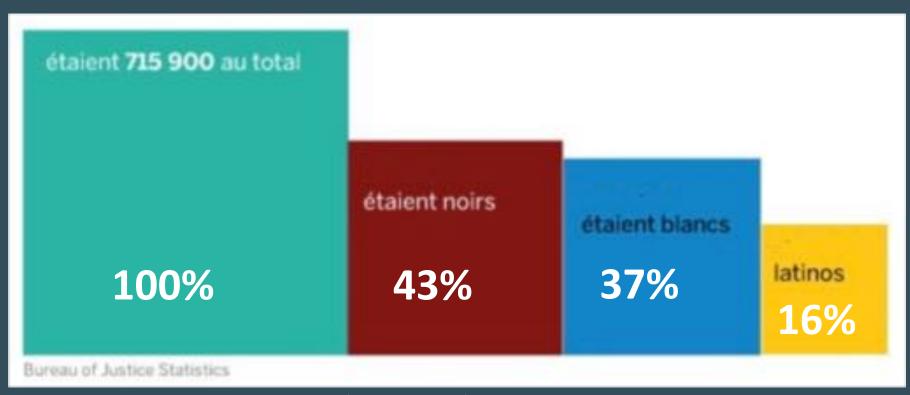
Les variables visuelles définies par Bertin



Le langage graphique s'exprime par des variables visuelles

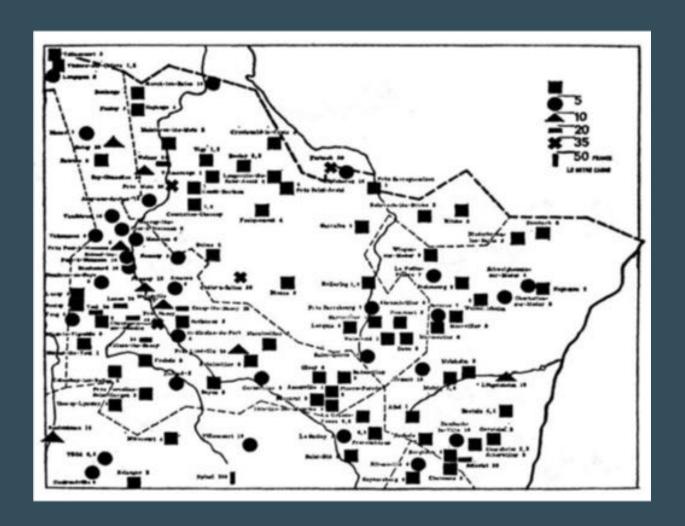
### QUIZZ

Si 100% des prisonniers US sont représentés par le carré vert, quelle est la proportion de chaque groupe ?



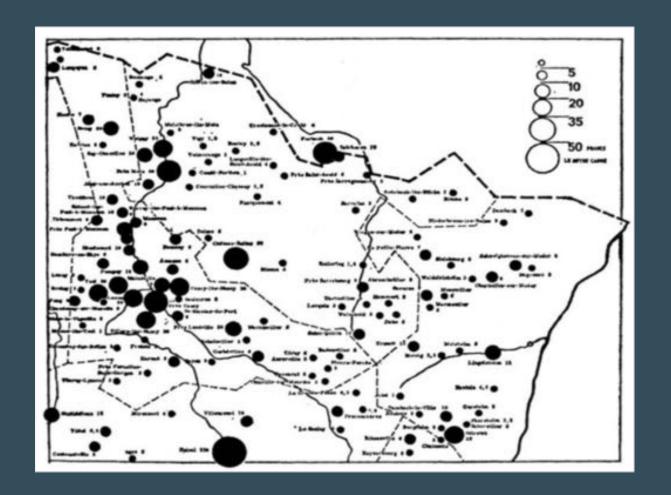
Ethic composition of prisoners in Jail in 2008 in the USA. (Le Monde 5/12/2014)

## Les Travaux de Jacques Bertin : Les formes



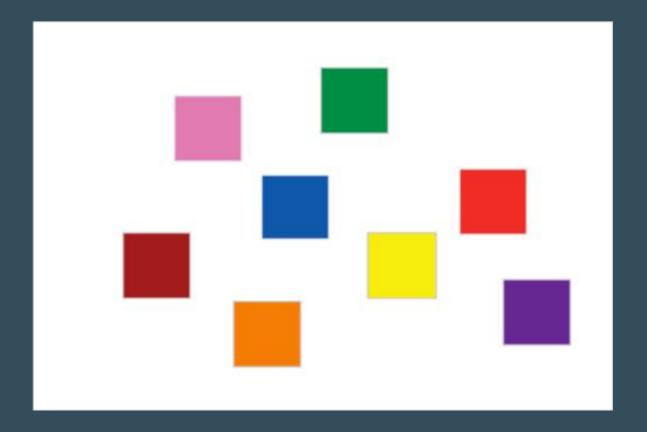
Les formes ne sont pas adaptées pour représenter la proportionnalité

## Les Travaux de Jacques Bertin : La Taille



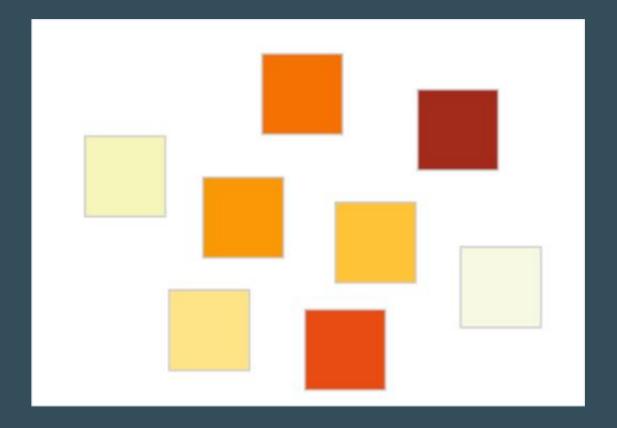
En revanche, les tailles sont adaptées pour représenter la proportionnalité (variation quantitative)

## QUIZZ: Trier par ordre croissant



Les couleurs traduisent des différences non ordonnées!

### QUIZZ: Trier par ordre croissant



Few (2008) a défini des palettes de couleurs (utilisées dans les heatmaps)

#### Relation entre les données

 ressemblance = différence = information qualitative

Pas de notion d'ordre

ordre, classement, hiérarchie

Des données numériques ou non

information ordonnée

proportionnalité Q

--- information quantitative

Quantités absolues

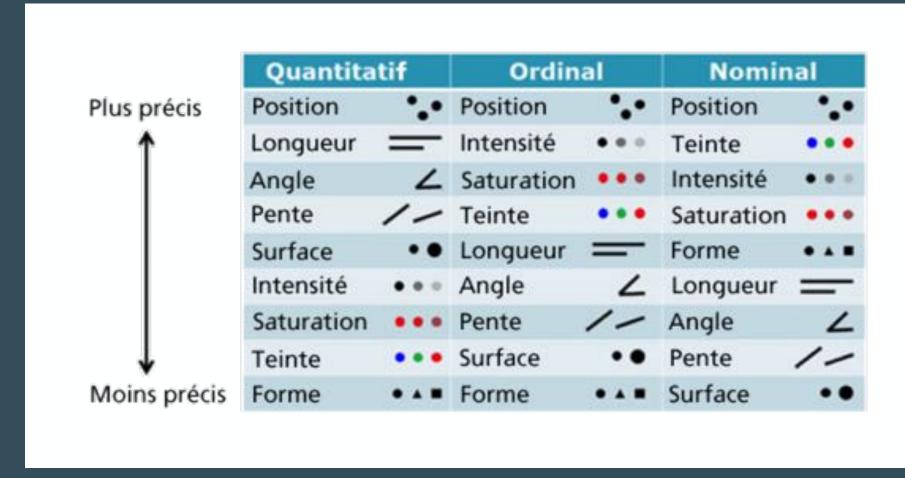


MOOC COURSERA - ENS – Cartographie thématique

# Les variables visuelles définies par Bertin

	Implantation ponctuelle	Implantation linéaire	Implantation zonale
=/=	FORME COULEUR Orientation Grain  □ □ ○ + ☆ ❖ ■ ○	COULEUR	COULEUR Forme Orientation Grain
0	TAILLE Valeur	TAILLE Valeur	VALEUR Grain + Couleur
Q	TAILLE	TAILLE	TAILLE TAILLE

### Les variables visuelles définies par Bertin



 $https://interstices.info/jcms/n\_51807/la-visualisation-analytique-pour-comprendre-des-donnees-complexes\\$ 

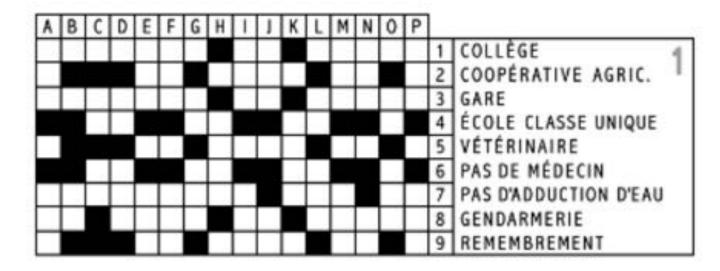
Les variables les plus couramment utilisées sont la valeur, la couleur, la taille et la forme.

\_\_\_\_\_

Les variables sont fréquemment associées lorsqu'elles ont les mêmes propriétés pour gagner en efficacité et en longueur.

### Les Travaux de Jacques Bertin - Tables & Matrices

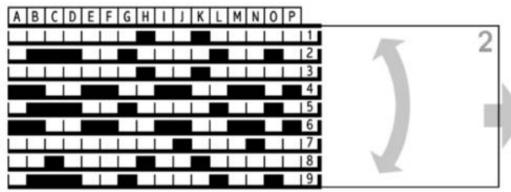
Data with many 0/1 variables: Facilities indicators for 16 towns (A-P)

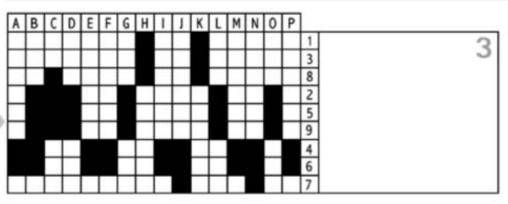


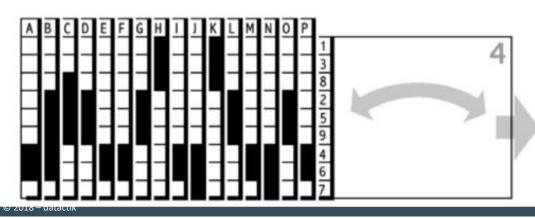
Bertin (1981)

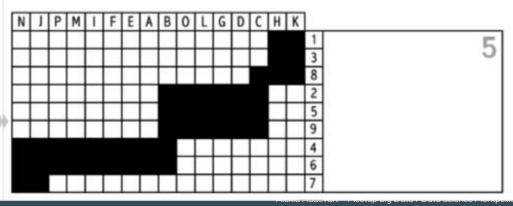
## Les Travaux de Jacques Bertin - Tables & Matrices



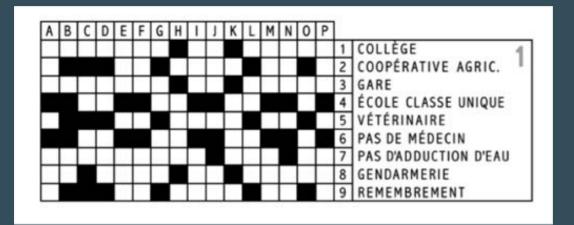


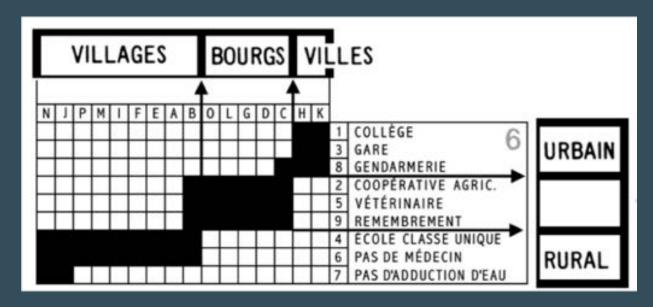






### Les Travaux de Jacques Bertin - Tables & Matrices





# Sommaire

- 1 Définition
- 2 L'efficacité graphique selon Jacques Bertin
- 3 L'excellence graphique d'Edward Tufte
- 4 Perception Visuelle
- 5 Le Story Telling
- 6 Good or Bad?

#### Les Travaux de Tufte



#### **Edward Tufte:**

- Professeur Américain
- "Leonard de Vinci des données" par le NY Times
- Auteur de 4 livres : <a href="http://www.edwardtufte.com/tufte/">http://www.edwardtufte.com/tufte/</a>

"Graphics are instruments for reasoning about quantitative information.

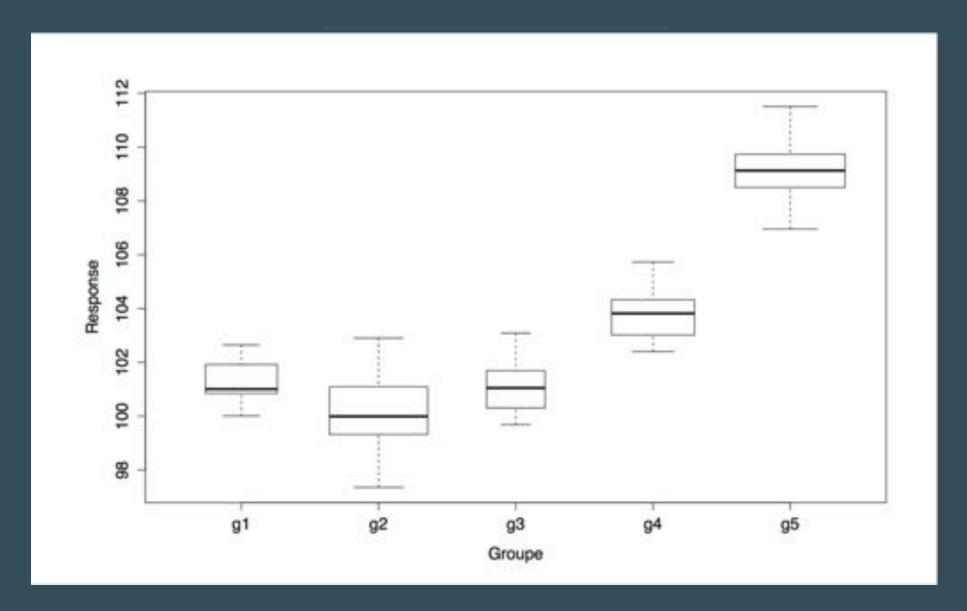
Often the most effective way to describe, explore and summarize a set of numbers - even a large set - is to look at pictures of those numbers"

### Les Travaux de Tufte

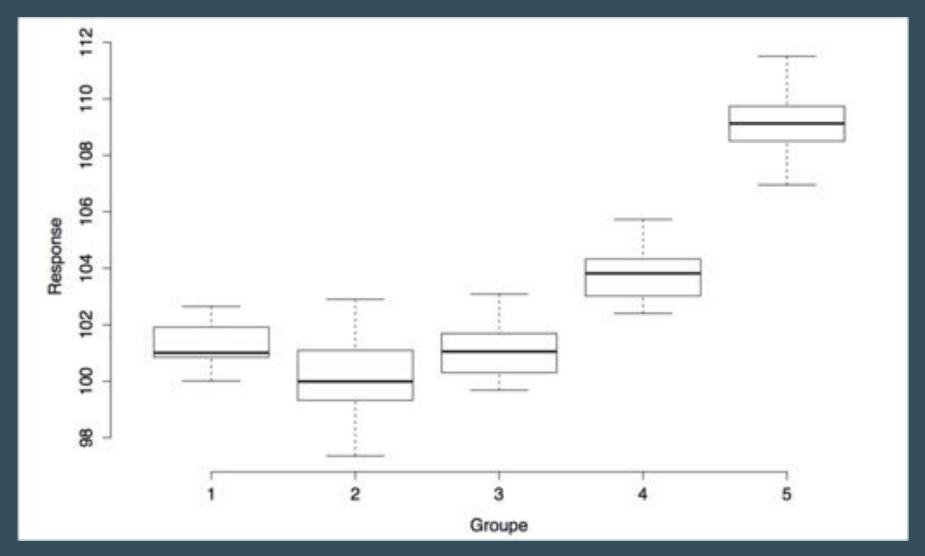
Data : Avant tout montrer les données !

- Question:
  - Inciter le spectateur à réfléchir à la substance plutôt qu'à la méthodologie ou à la conception graphique.

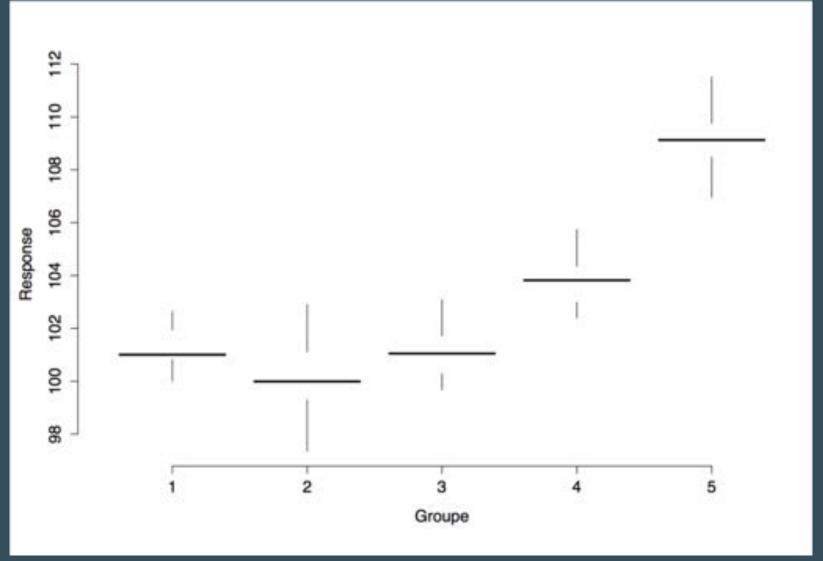
- Encourager l'œil à comparer différentes données.
- Concept du data-ink ratio
  - => Maximiser la data densité
- Intégrité : Eviter les manipulations / tordre les données
- General to Specific
- Le Contexte



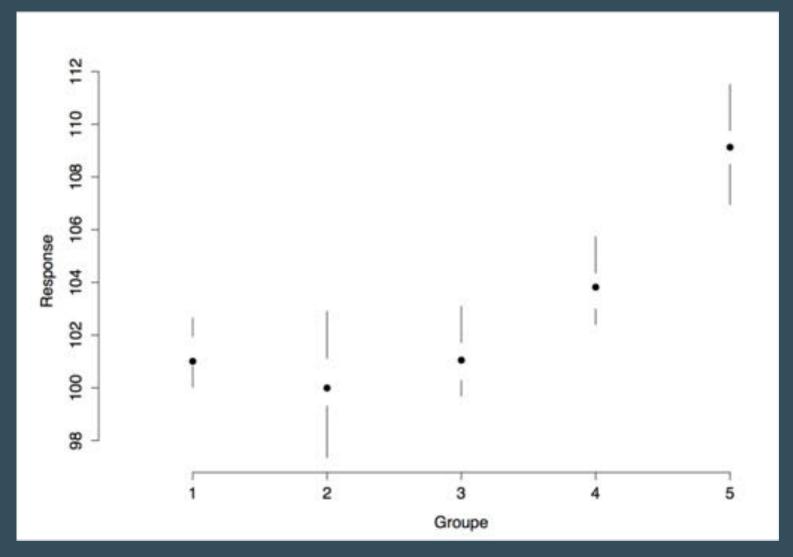
Suppression de l'encre ne représentant pas les données



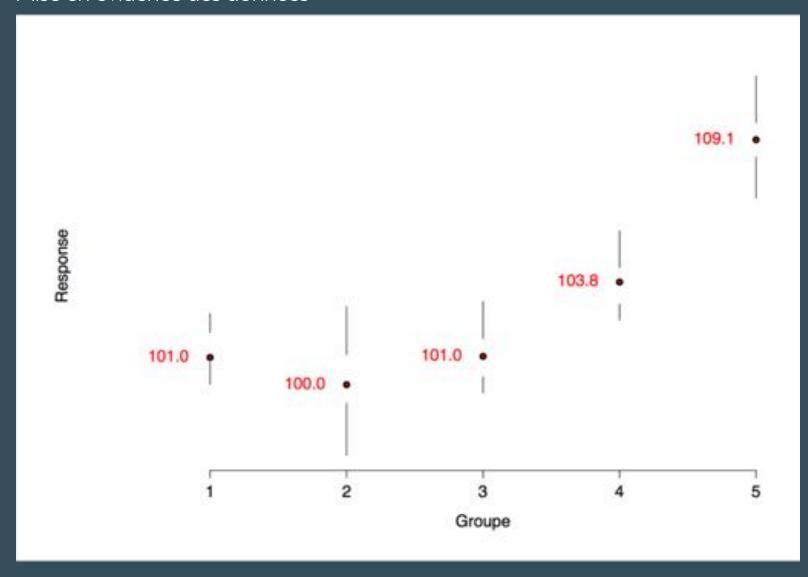
Suppression de toutes les informations redondantes



### Suppression de toutes les informations redondantes

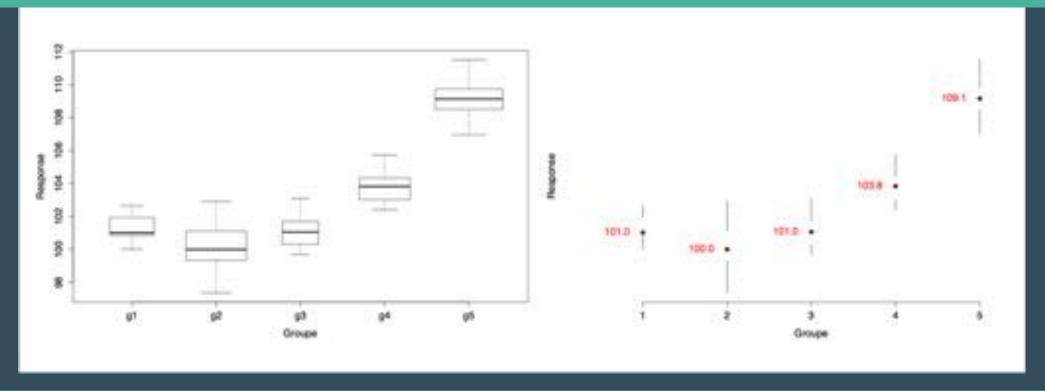


#### Mise en évidence des données



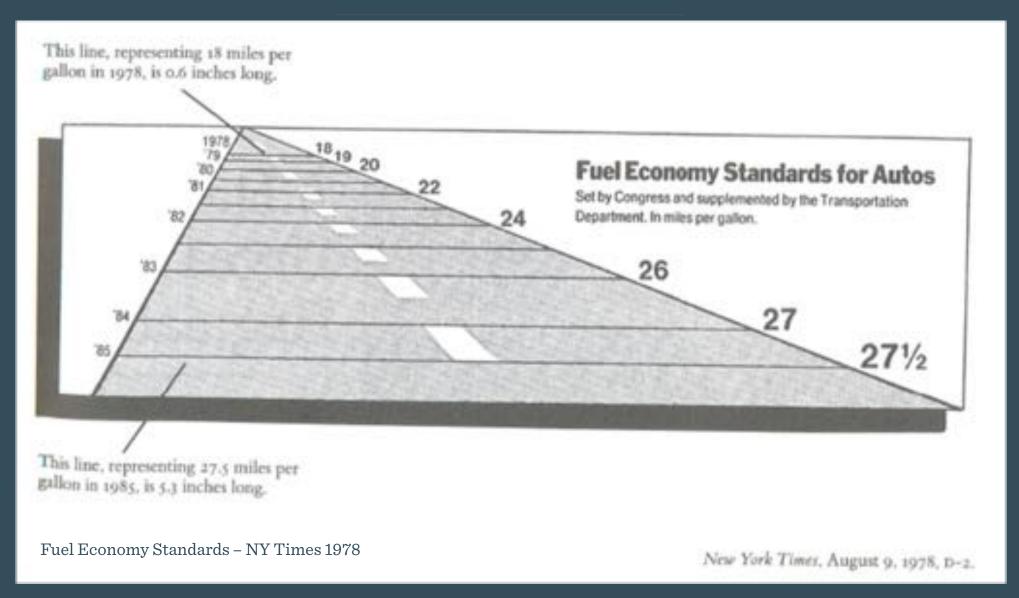
Avons nous perdu des informations?

Quelle information est mise en évidence?



L'excellence graphique se détermine donc par la conjonction de la densité de l'information associée à une faible densité visuelle

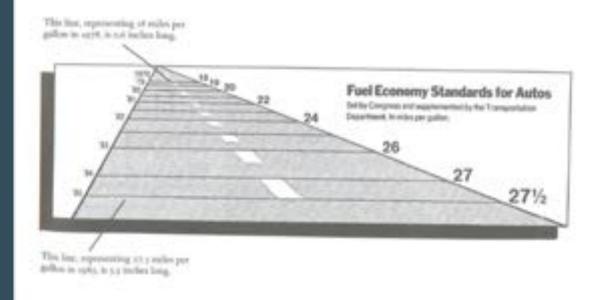
#### Les Travaux de Tufte : Intégrité – Lie Factor



#### Les Travaux de Tufte : Intégrité – Lie Factor

The "18 mpg" line measures 1.5 cm (in 1978); the "27,5 mpg" measures 13 cm (in 1985)

 $\longrightarrow$  Lie factor = 14.5%!!!



New York Timer, August 2, 1578, 21-2.

Fuel Economy Standards – Revisited



#### Attention aux manipulations : Intégrité – Lie Factor





Le Point: https://twitter.com/lepoint/status/657611278832287744

- Salaires : 28 % représentés par... 200 %
- Heures de vol : 200 % de différence pour représenter 23 %

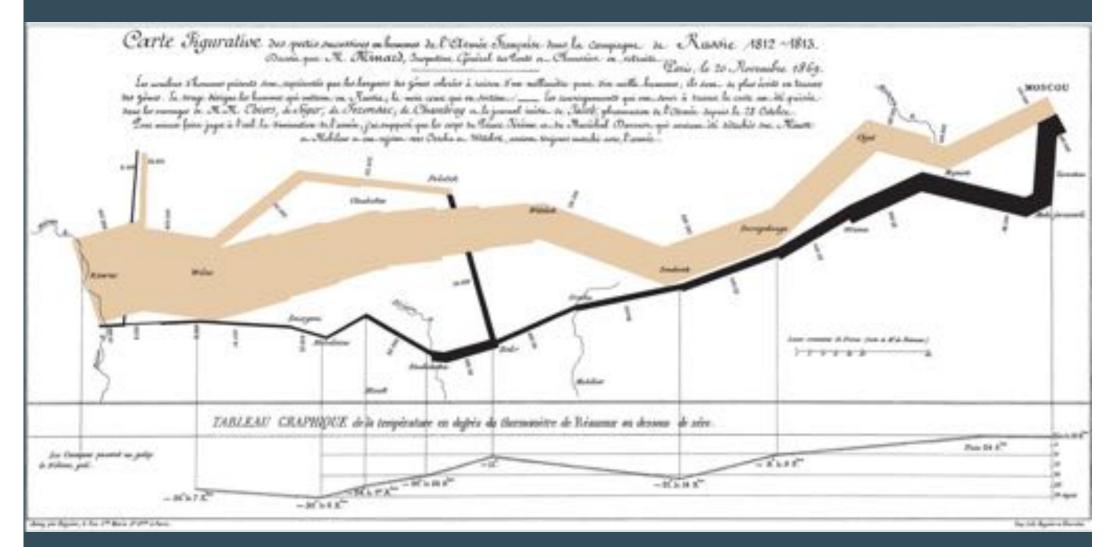
Les Travaux de Tufte : Intégrité – Lie Factor

$$LieFactor = \frac{Size \ of \ effect \ shown \ in \ graphic}{Size \ of \ effect \ in \ data}$$

L'effet visuel doit être égal à l'effet dans les données Lie Factor = 1 sinon DISTORSION

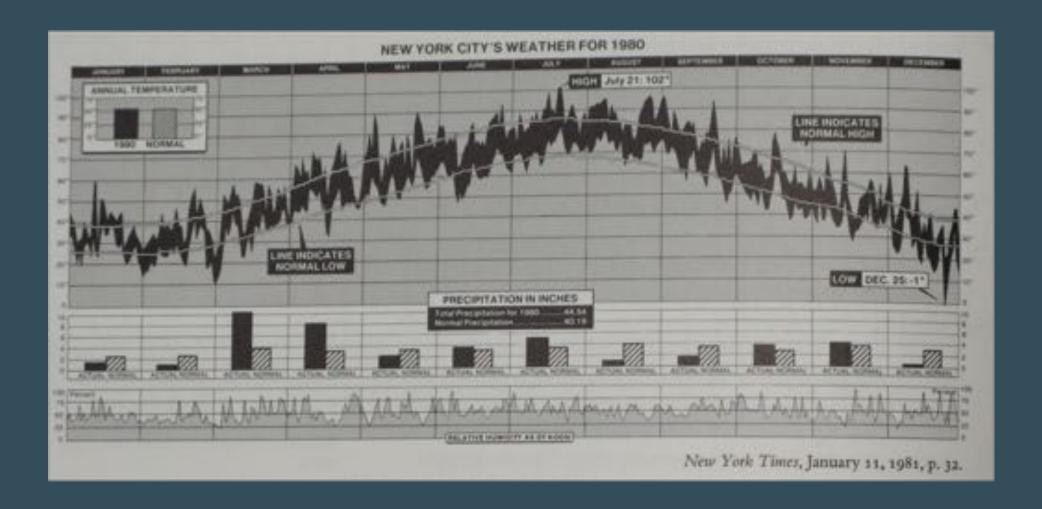
# Carte figurative des pertes successives en hommes de l'Armée française dans la campagne de Russie en 1812-1813

Par Charles Joseph Minard (1869)



« Le meilleur graphique statistique jamais tracé » [Tufte]

#### Un schéma pour 1888 nombres



"The graphics (...) tells a story "[Tufte, Visual display ... p.30]

# Sommaire

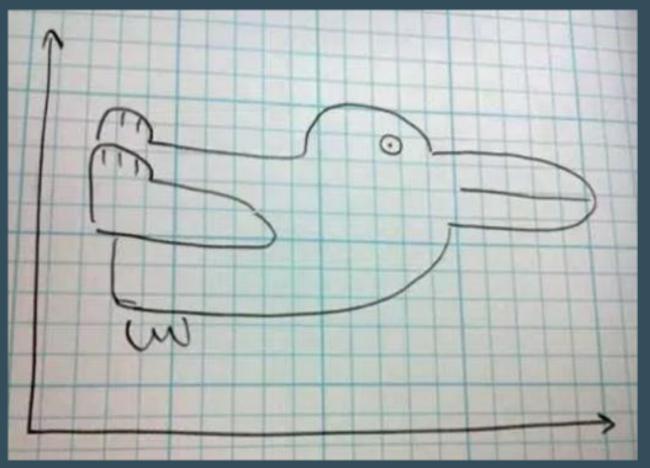
- 1 Définition
- 2 L'efficacité graphique selon Jacques Bertin
- 3 L'excellence graphique d'Edward Tufte
- 4 Perception Visuelle
- 5 Le Story Telling
- 6 Good or Bad?

#### Capacités pré-attentives

- Perception inconsciente
- Mécanismes bas niveau du système visuel
- Extrêmement rapide : réaction en 200 ms
- Montée en charge : le temps de réaction croit lentement avec le nombre d'objets
- Différents types de stimuli



#### QUIZZ: Que voyez-vous?

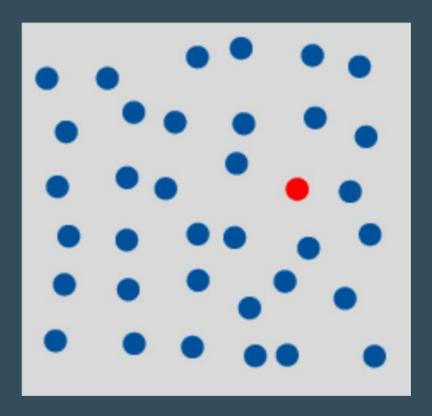


http://flowingdata.com/2014/06/25/duck-vs-rabbit-plot/

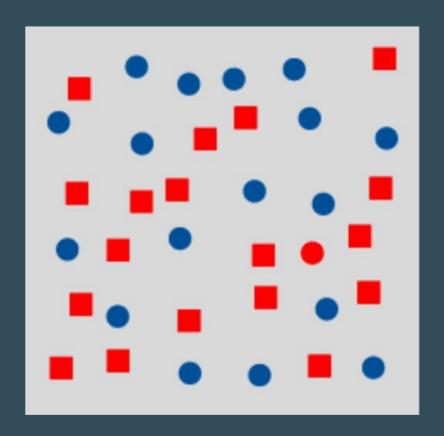
#### QUIZZ : Compter le nombre de 3

#### QUIZZ : Compter le nombre de 3

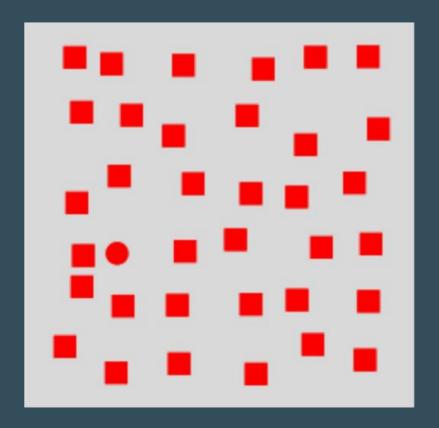
#### QUIZZ: Trouver le point rouge



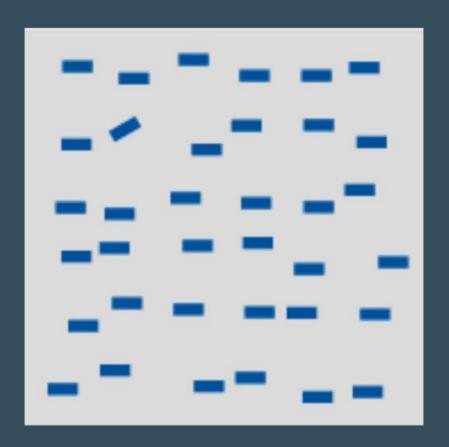
### QUIZZ : Trouver le point rouge



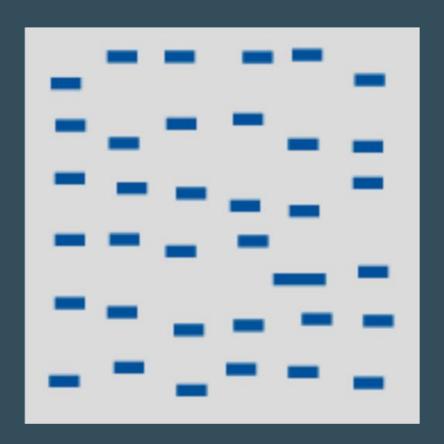
#### QUIZZ : Trouver le point rouge



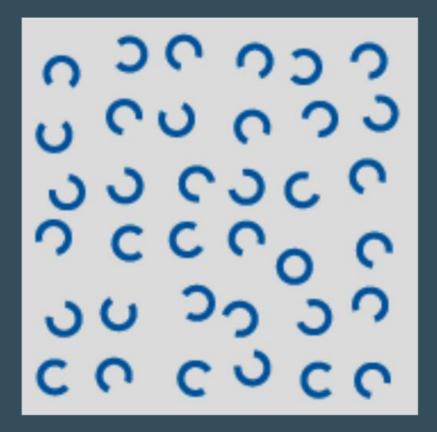
#### **QUIZZ**: Trouver l'intrus



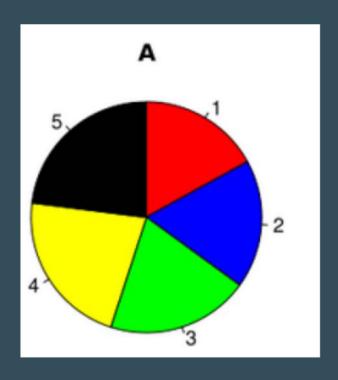
#### **QUIZZ**: Trouver l'intrus

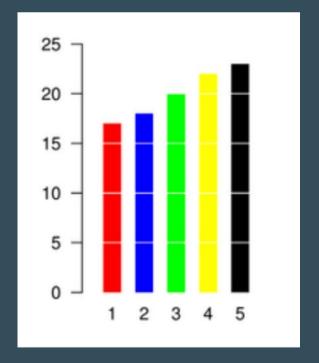


#### **QUIZZ**: Trouver l'intrus

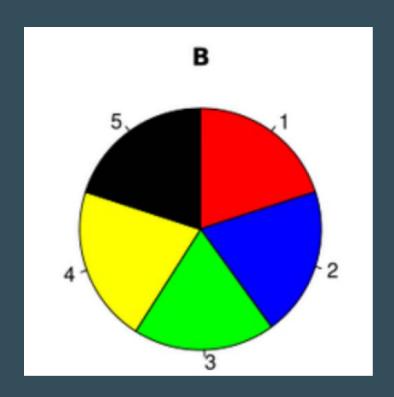


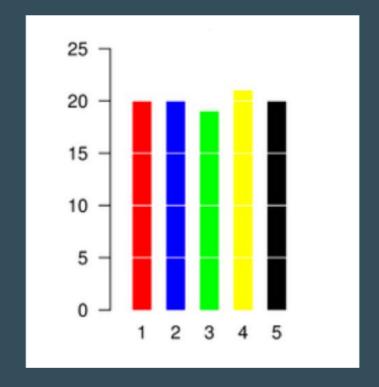
#### QUIZZ: Trier par ordre croissant



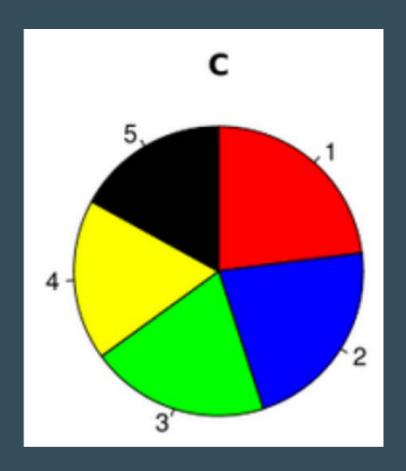


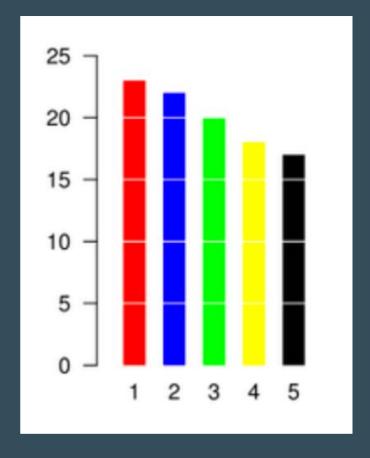
#### QUIZZ: Trier par ordre croissant



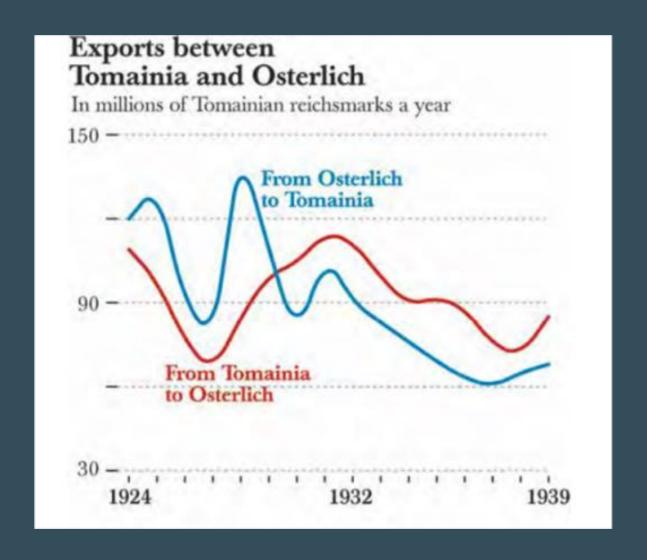


#### QUIZZ: Trier par ordre croissant

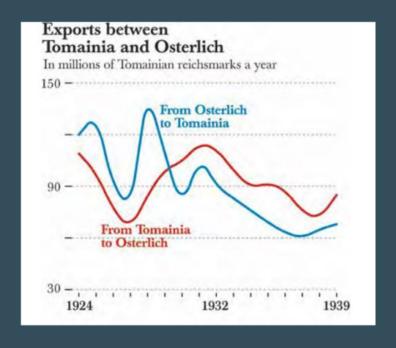


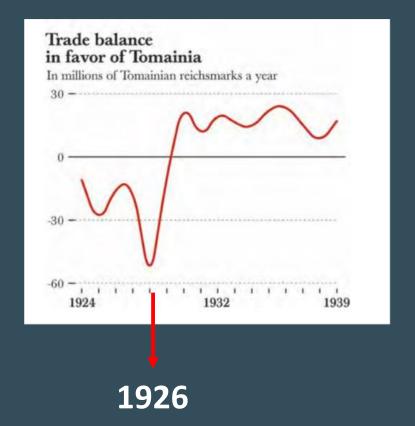


#### QUIZZ : En quelle année a eu lieu le plus grand écart ?

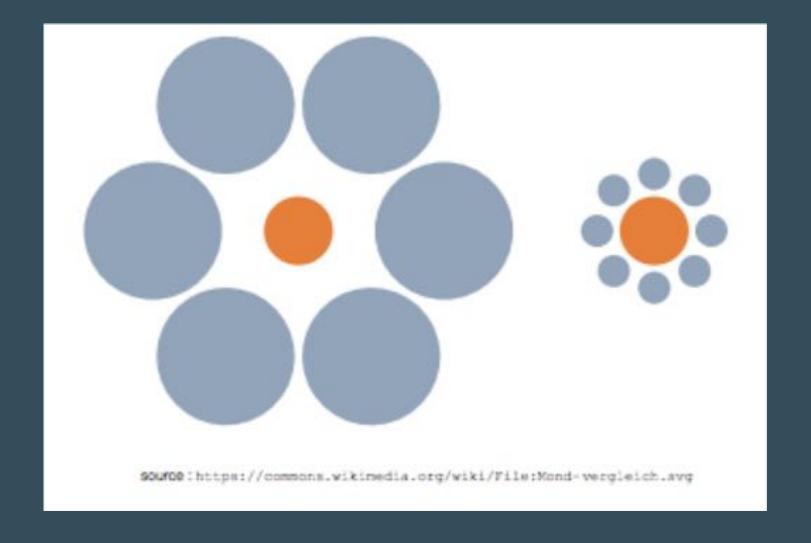


#### QUIZZ : En quelle année a eu lieu le plus grand écart ?

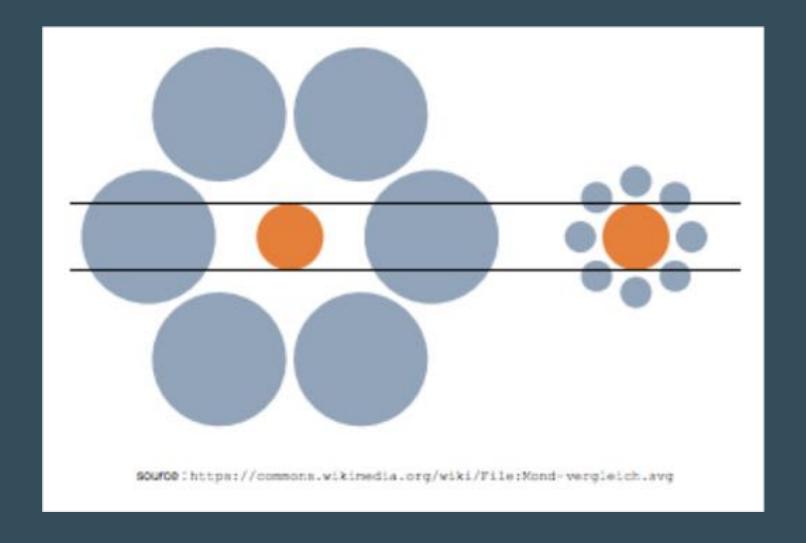




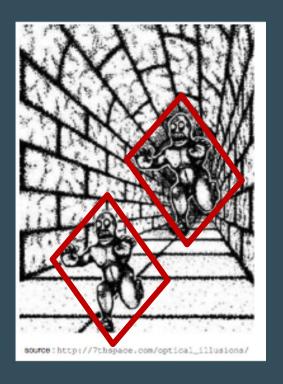
#### QUIZZ: Illusion de Ebbinghaus/Titchener



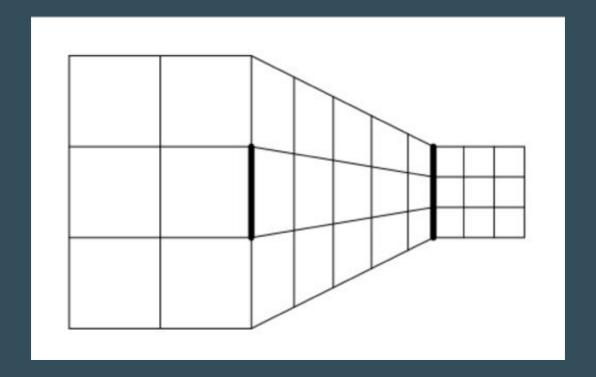
#### QUIZZ: Illusion de Ebbinghaus/Titchener



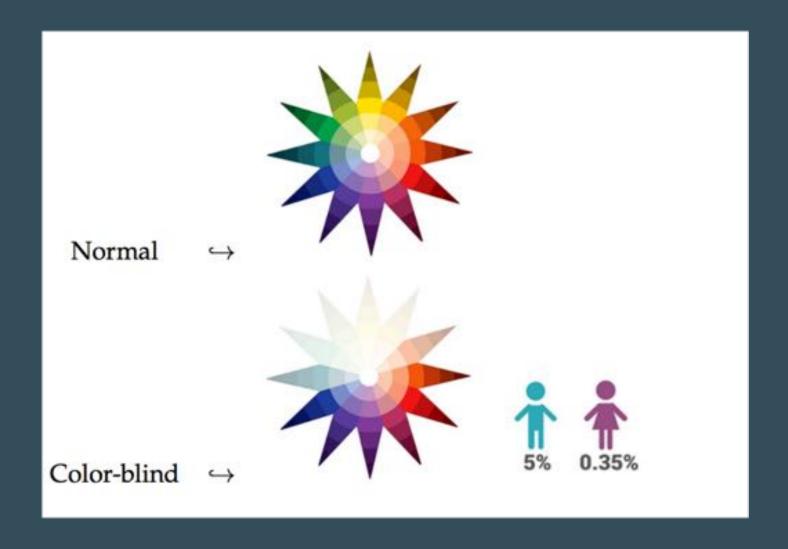
#### **QUIZZ**: Perspective



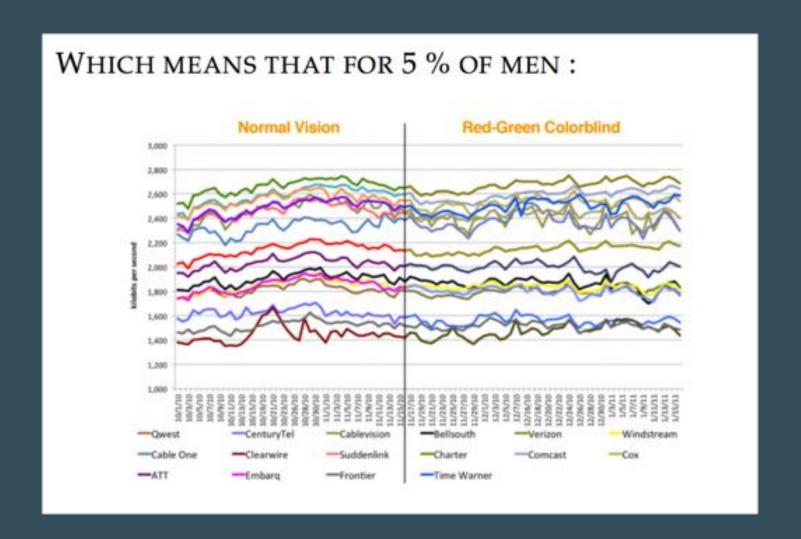
#### QUIZZ : Perspective – Variante de l'illusion de Ponzo



#### Capacité pré-attentives : Couleurs - Daltonisme



#### Capacité pré-attentives : Couleurs - Daltonisme



#### Perceptions Visuelles

#### Limites et freins de la nature humaine

- Physiologie de l'oeil
- Illusions graphiques
- Fausses interprétations
- Daltonisme
- Erreurs d'appréciation
- Préjugés de formes
- Charge cognitive

# Sommaire

- 1 Définition
- 2 L'efficacité graphique selon Jacques Bertin
- 3 L'excellence graphique d'Edward Tufte
- 4 Perception Visuelle
- 5 Le Story Telling
- 6 Good or Bad?

#### Les clés du Data Storytelling







#### Les clés du Data Storytelling



Le Data Storytelling permet de démocratiser l'accès aux données et de faciliter la prise de décision

#### Posez-vous les bonnes questions!

- Quel est le message que vous souhaitez transmettre ?
- A quel public est-il destiné?
- Quelles questions se pose ce public ?
- Quelles réponses lui apportez-vous?
- Les données les plus importantes sont-elles mises en valeur ?
- La compréhension de votre visualisation est-elle rapide ?
- Est-elle intelligible par elle même?

#### Exemples

# 100 YEARS OF ROCK THAN A MINUTE

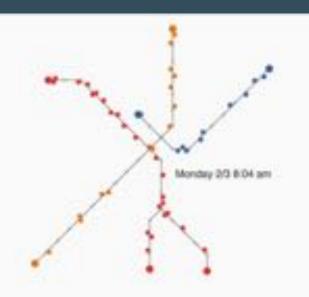
http://www.concerthotels.com/100-years-of-rock/

#### **Exemples**

# Visualizing MBTA Data

An interactive exploration of Boston's subway system

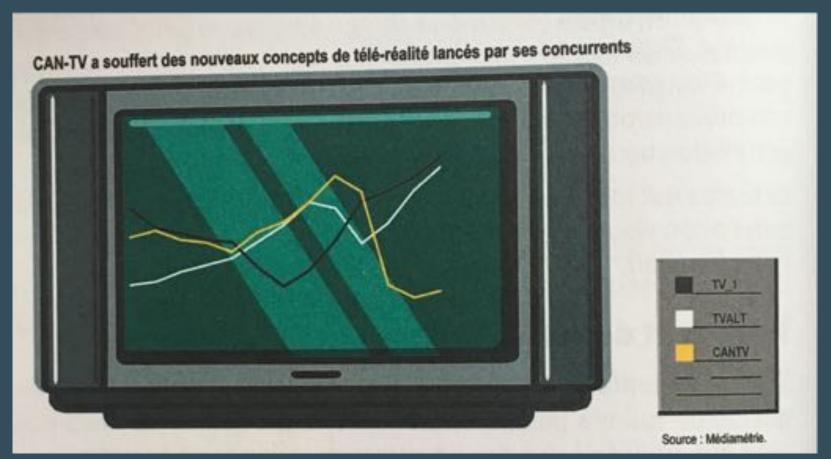
Mike Barry and Brian Card - June 10, 2014



http://mbtaviz.github.io/

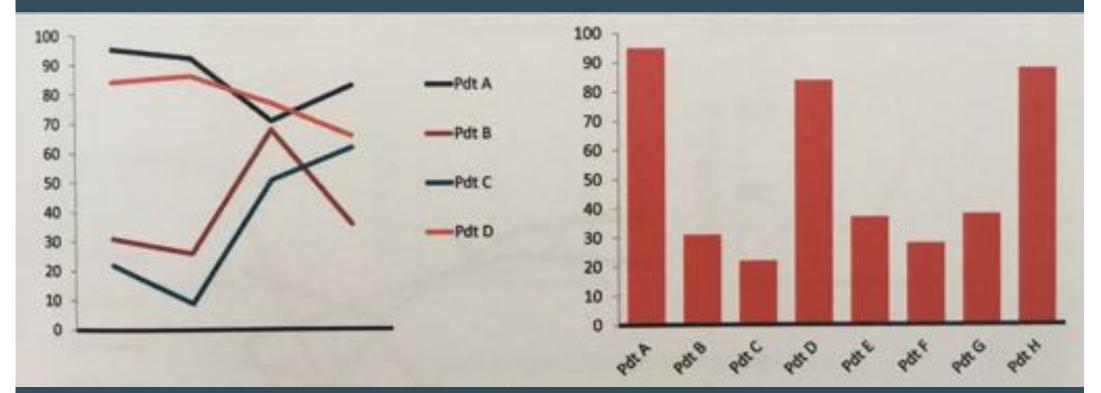
# Sommaire

- 1 Définition
- 2 L'efficacité graphique selon Jacques Bertin
- 3 L'excellence graphique d'Edward Tufte
- 4 Perception Visuelle
- 5 Le Story Telling
- 6 Good or Bad?



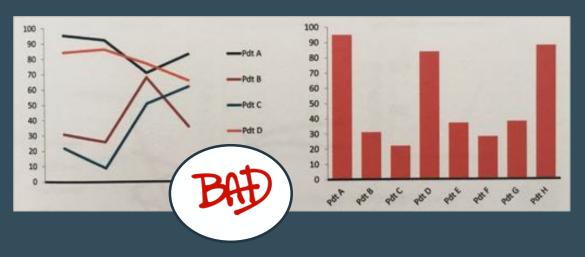
Bernard Lebelle – Convaincre avec des graphiques efficaces

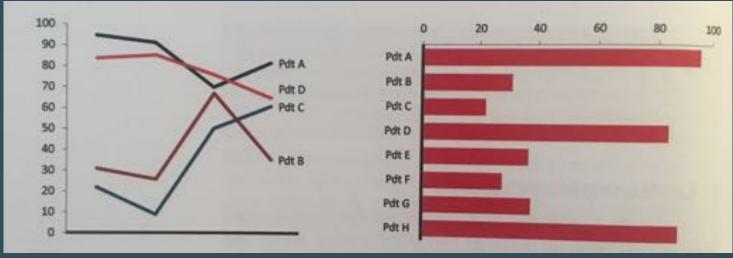
## Trop de décorum tue l'information



Bernard Lebelle – Convaincre avec des graphiques efficaces

#### QUIZZ: Good or Bad? - Le mouvement pendulaire





Définir un emplacement judicieux des légendes

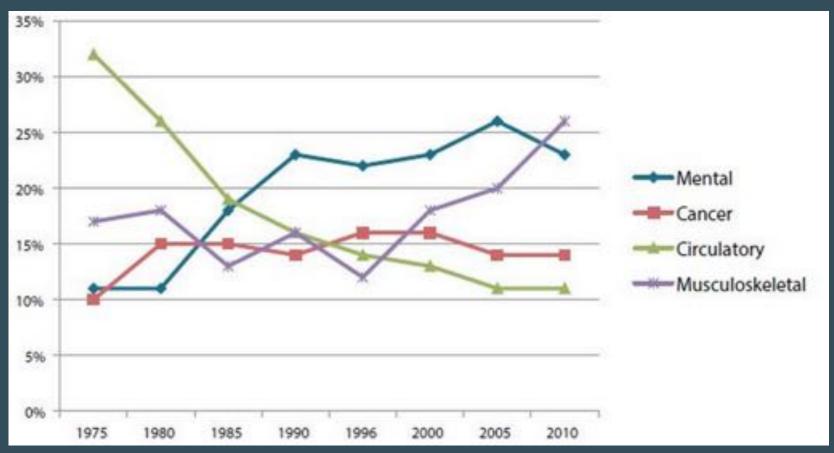
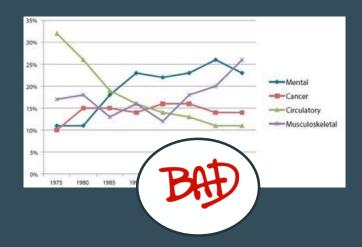


FIGURE – Major Cause of Worker Disability (1975-2010) (J. Schwabish, 2014).

- Quelle est la valeur observée la plus élevée ? C'était quand ?
- En 2010, quelle était la cause principale d'invalidité?
- Est-ce que le cancer (courbe rouge) augmente durant la période représentée ?
- Au cours des dernières années, quelles sont les causes qui ont augmenté (diminué) le plus ?

#### QUIZZ: Good or Bad? - Les effets spaghettis



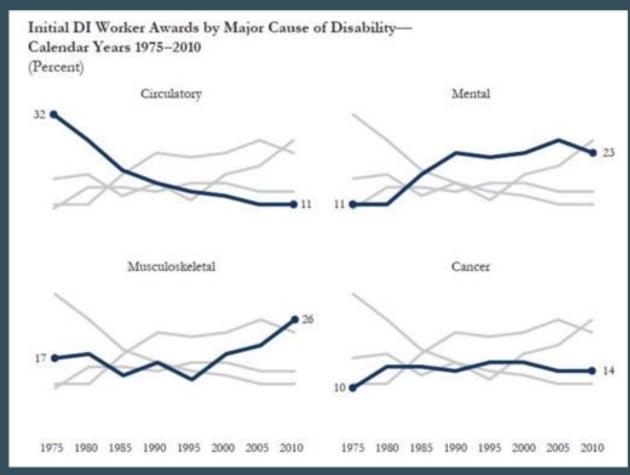
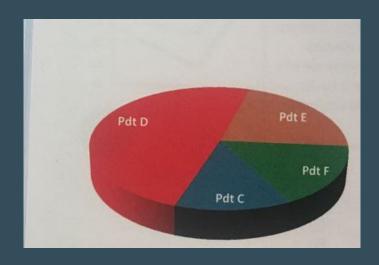
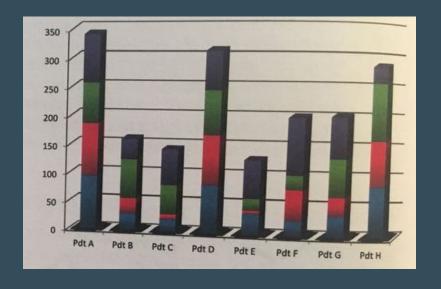


FIGURE – Major Cause of Worker Disability (1975-2010) (J. Schwabish, 2014).

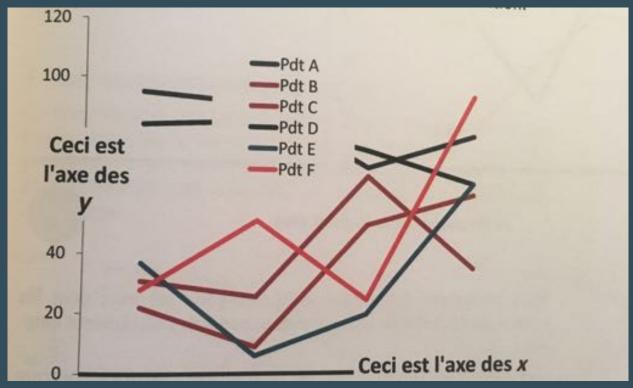
## Scinder en mini-canevas les univers complexes



Bernard Lebelle – Convaincre avec des graphiques efficaces

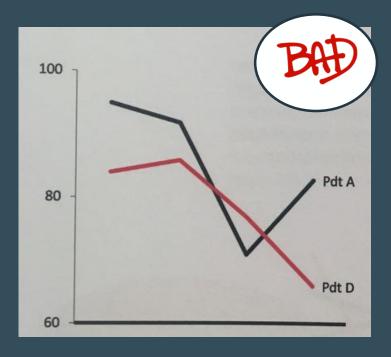


#### Eviter les 3D malsaines!

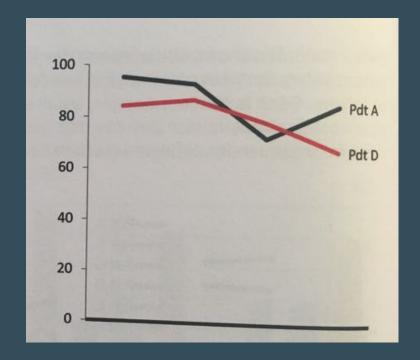


Bernard Lebelle – Convaincre avec des graphiques efficaces

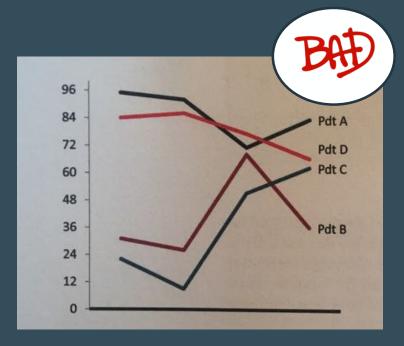
## Ne pas surcharger les graphiques!



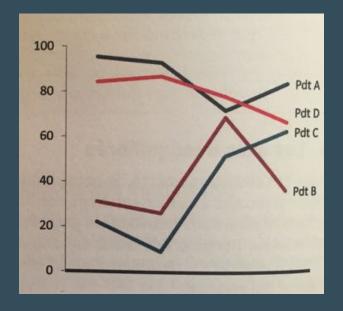
Bernard Lebelle – Convaincre avec des graphiques efficaces



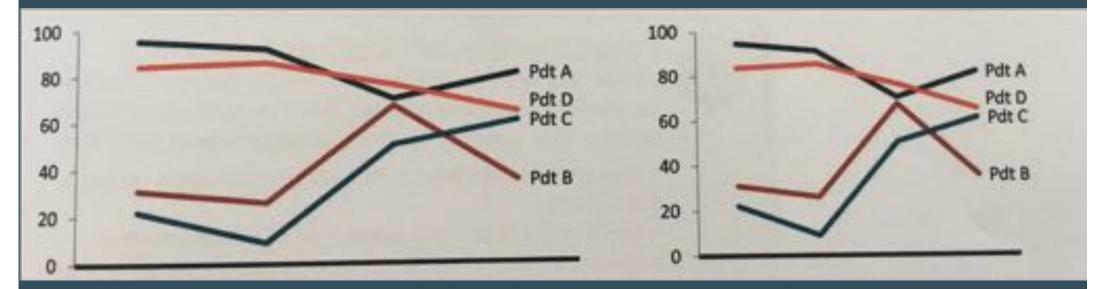
## Ne pas occulter le point zéro!



Bernard Lebelle - Convaincre avec des graphiques efficaces



#### Choisir des incréments naturels!



Bernard Lebelle – Convaincre avec des graphiques efficaces

## Ne pas étirer ou déformer les axes

#### Construction d'un bon graphique

• Utiliser les attributs graphiques adaptés

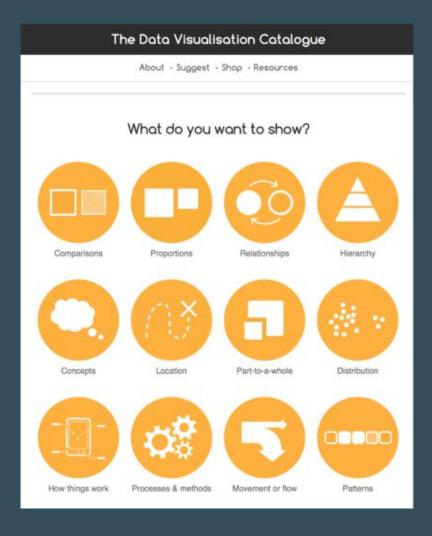
S'appuyer sur la pré-attention

• Ne pas distraire le lecteur

Ne pas mentir au lecteur

#### Construction d'un bon graphique

#### The Data Visualisation Catalogue About - Suggest - Shop - Resources Search by Function View by List **Bubble Chart** Arc Diagram Area Graph Bar Chart Box & Whisker Plot Brainstorm 00000 Bubble Map Bullet Graph Calendar Candlestick Chart Chord Diagram Choropleth Map Circle Packing Connection Map Density Plot Dot Matrix Chart Error Bars Flow Chart Flow Map **Illustration Diagram** Kagi Chart Histogram.



http://datavizcatalogue.com/search.html

# KISS



Keep It Significant & Shareable!

"There are no "good" nor "bad" graphics (...), there are graphics answering legitimate questions and graphics that do not answer question at all "

**Bertin (1981)** 

## Thank You!

Do you have any questions?

#### **Alexia Audevart**

Data & Enthusiasm



alexia@datactik.com



@aaudevart



#### Sources

Jacques Bertin: Livre Semiology of Graphics

Edward Tufte: Livre The Visual Display of Quantitative Information

Convaincre avec des graphiques efficaces – Bernard Lebelle Eyrolles : <a href="https://www.eyrolles.com/Entreprise/Livre/convaincre-avec-des-graphiques-efficaces-9782212553994">https://www.eyrolles.com/Entreprise/Livre/convaincre-avec-des-graphiques-efficaces-9782212553994</a>

MOOC ENS - Coursera - Cartographie thématique : <a href="https://www.coursera.org/learn/cartographie">https://www.coursera.org/learn/cartographie</a>

#### Christophe Bontemps:

- <a href="https://speakerdeck.com/toulousedatascience/number-26-la-data-visualisation-pour-la-data-science-et-les-data-scientists">https://speakerdeck.com/toulousedatascience/number-26-la-data-visualisation-pour-la-data-science-et-les-data-scientists</a>
- http://data.visualisation.free.fr/