

## DATA VISUALISATION

Année scolaire 2018/2019 Support de cours

Alison PATOU

Data Scientist chez ECONOCOM
Alison.patou@econocom.com

## SOMMAIRE

- 1. LA DATA VISUALISATION
- 2. TABLEAU SOFTWARE
- 3. TRAVAUX PRATIQUES SUR TABLEAU SOFTWARE

# LA DATA VISUALISATION

#### Quelle est la plus grande valeur ?

	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
Janvier	62405	74505	60370	11981	69052	53961	34195
Février	62914	47190	50828	30369	85128	20144	65602
Mars	24686	30565	51608	63073	48626	72855	48669
Avril	86809	11457	80951	58427	91895	38992	11324
Mai	12793	87570	95553	60865	29808	45423	69133
Juin	41580	99474	53347	33245	11694	51940	95770
Juillet	54924	16225	81434	89454	61081	75905	82631
Août	60217	66368	30667	65971	48806	58906	14935
Septembre	55941	18541	56107	52343	14971	20181	24319
Octobre	54438	94302	74714	44129	24986	12539	49630
Novembre	72945	92634	57760	51066	41845	80109	83700
Décembre	50897	91724	19291	94308	18279	74317	58655

#### Quelle est la plus grande valeur ?

	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
Janvier	62405	74505	60370	11981	69052	53961	34195
Février	62914	47190	50828	30369	85128	20144	65602
Mars	24686	30565	51608	63073	48626	72855	48669
Avril	86809	11457	80951	58427	91895	38992	11324
Mai	12793	87570	95553	60865	29808	45423	69133
Juin	41580	99474	53347	33245	11694	51940	95770
Juillet	54924	16225	81434	89454	61081	75905	82631
Août	60217	66368	30667	65971	48806	58906	14935
Septembre	55941	18541	56107	52343	14971	20181	24319
Octobre	54438	94302	74714	44129	24986	12539	49630
Novembre	72945	92634	57760	51066	41845	80109	83700
Décembre	50897	91724	19291	94308	18279	74317	58655

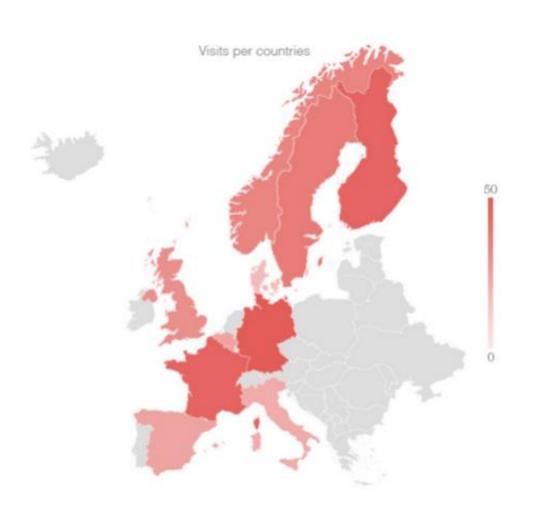
#### Quelle est la plus grande valeur ?

	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
Janvier	62405	74505	60370	11981	69052	53961	34195
Février	62914	47190	50828	30369	85128	20144	65602
Mars	24686	30565	51608	63073	48626	72855	48669
Avril	86809	11457	80951	58427	91895	38992	11324
Mai	12793	87570	95553	60865	29808	<b>45</b> 423	69133
Juin	41580	99474	53347	33245	11694	51940	95770
Juillet	54924	16225	81434	89454	61081	75905	82631
Août	60217	66368	30667	65971	48806	58906	14935
Septembre	55941	18541	56107	52343	14971	20181	24319
Octobre	54438	94302	74714	44129	24986	12539	49630
Novembre	72945	92634	57760	51066	41845	80109	83700
Décembre	50897	91724	19291	94308	18279	74317	58655

Autre exemple : quel pays a le plus de visites?

#### Visits per countries

	2010	2011	2012	2013	2014
U.K.	30	35	40	25	30
Belgium	10	15	20	20	15
France	35	40	20	20	25
Italy	10	10	15	10	20
Norway	20	25	25	35	45
Spain	5	15	10	15	20
Sweden	20	30	30	45	40
Germany	40	50	40	35	40
Finland	35	40	40	35	45
Danemark	5	5	15	20	20



#### LES APPORTS DE LA DATA VISUALISATION





Optimise et accélère la prise de décision



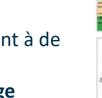
Facilite la compréhension des données



Améliore la communication



Contribue à la motivation des collaborateurs







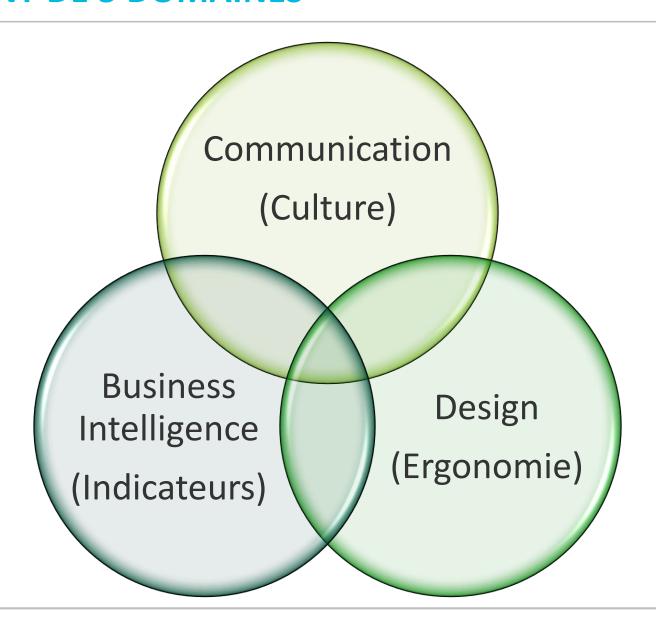


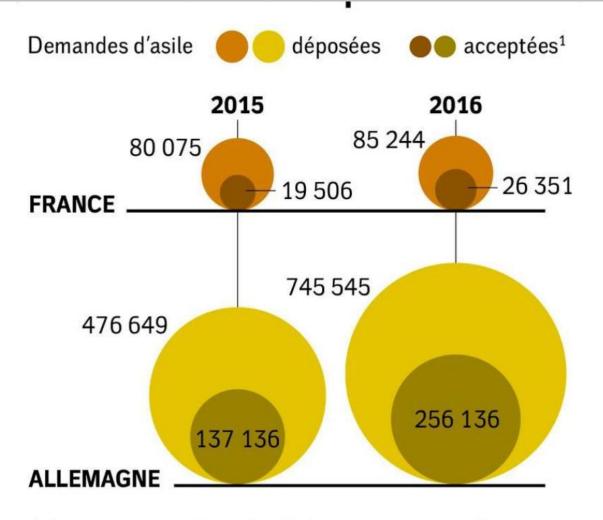
Source: IDG Research Services

77 % des clients déclarent qu'elle améliore la prise de décision grâce notamment à de **meilleures analyses** 

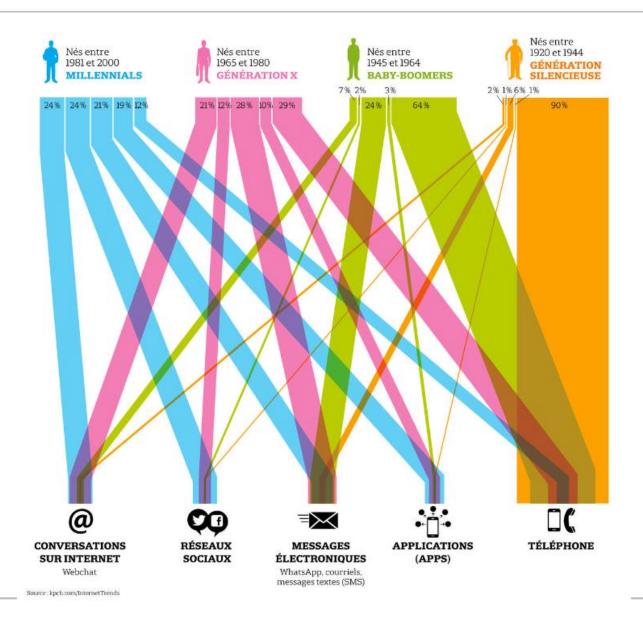
41% considèrent que la data visualisation améliore la **collaboration & le partage d'informations** 

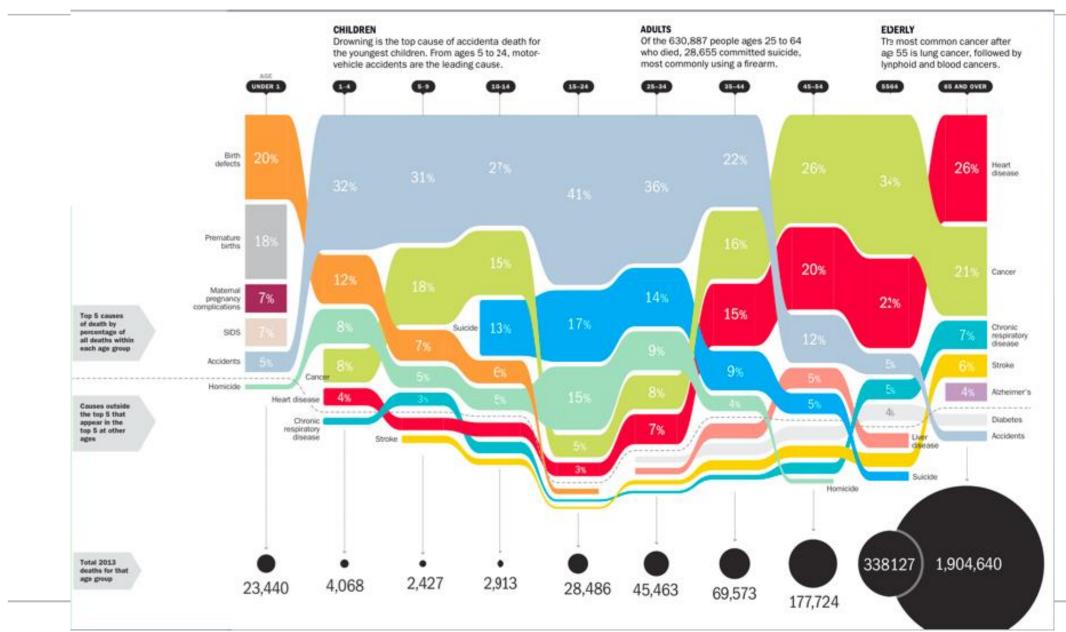
#### **AU CROISEMENT DE 3 DOMAINES**





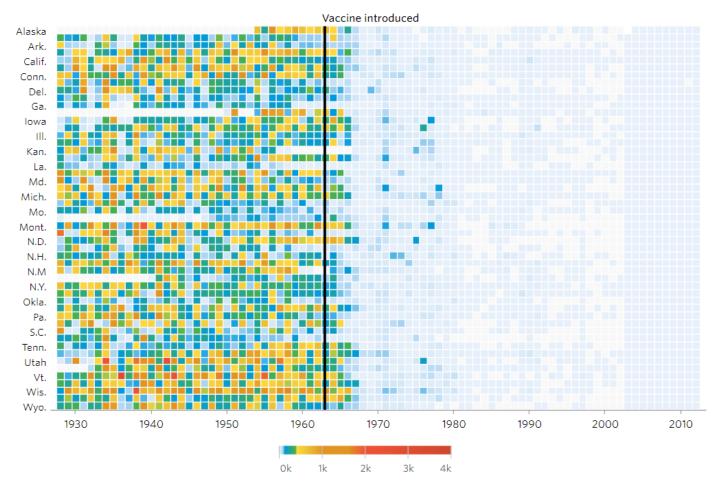
1. Les réponses aux demandes d'asile peuvent intervenir l'année qui suit leur dépôt. Le délai moyen de réponse en 2015 était de





The heat maps below show number of cases per 100,000 people.

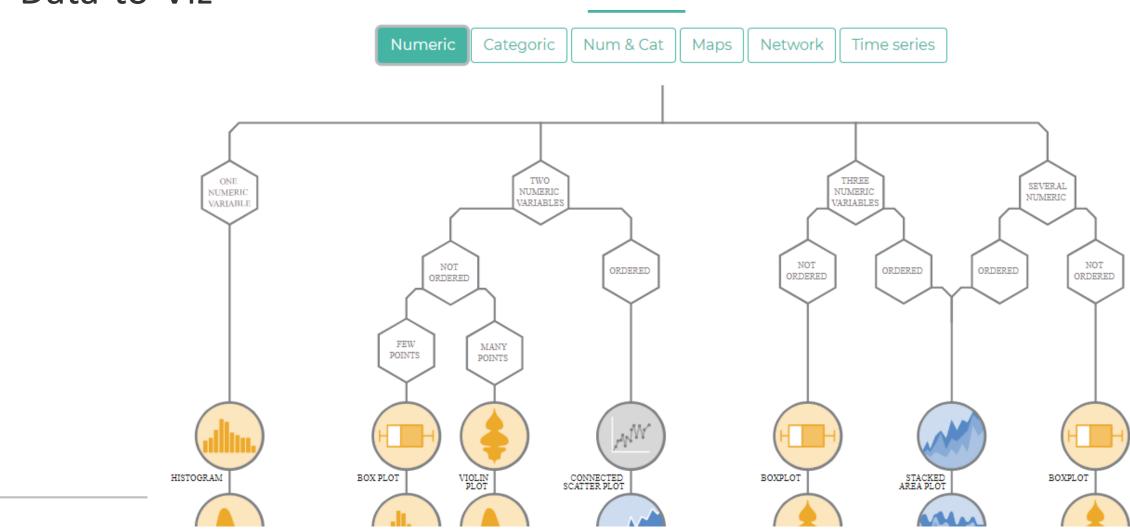
#### Measles



Note: CDC data from 2003-2012 comes from its Summary of Notifiable Diseases, which publishes yearly rather than weekly and counts confirmed cases as opposed to provisional ones.

#### LES TYPES DE DATA VISUALISATION

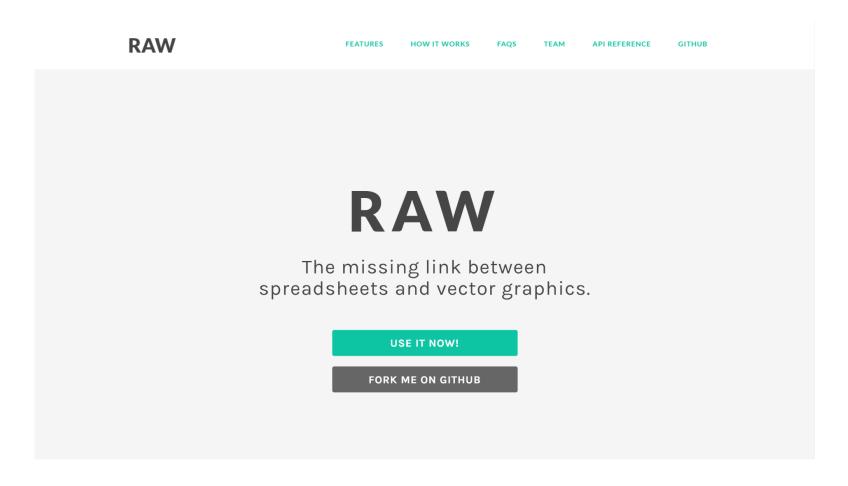
#### Data-to-Viz



#### LES ERREURS À ÉVITER

- 1. Pas de message véhiculé
- 2. Mauvais usage du graphique (pour ce que l'on cherche à démontrer)
- 3. Trop d'information (« information overload ») sur une même graphique
- 4. Manque de « goût » : police, couleurs, taille des traits, etc.
- 5. Ne pas ordonner les valeurs dans les barplots typiquement
- 6. Erreur de placement et d'ordre des éléments
- 7. 3D non justifiée
- 8. Abus de « camemberts » : l'œil humain a du mal à calculer les angles.
- 9. ...

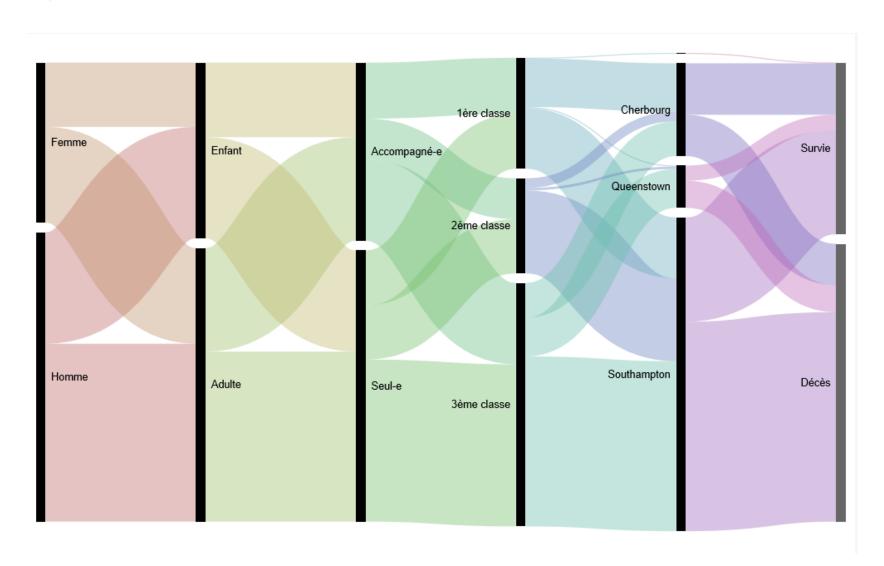
## **SE CONNECTER SUR SITE RAW GRAPHS**



http://raw.densitydesign.org/

## **ALLUVIAL DIAGRAM**

Profil des passagers du Titanic selon la survie ou le décès



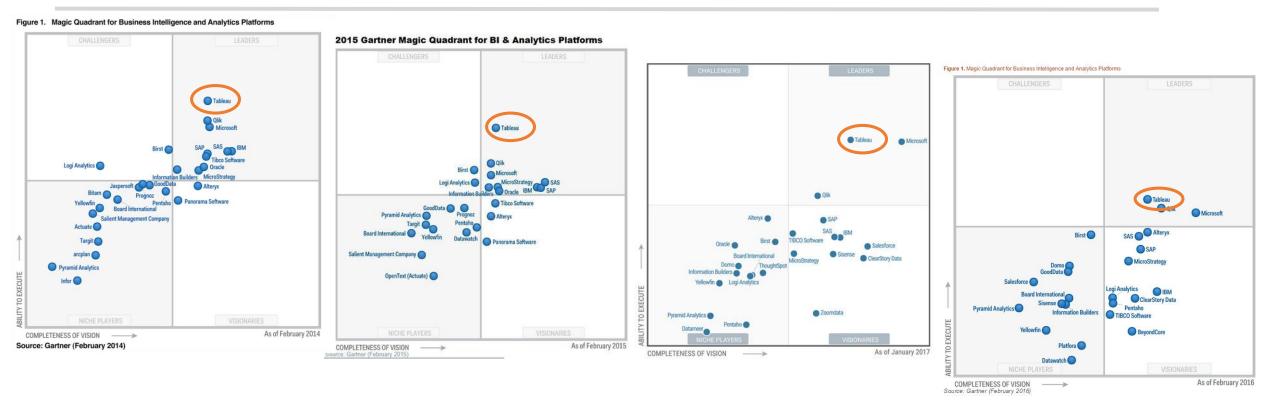
# TABLEAU SOFTWARE

#### **DÉCOUVRIR TABLEAU SOFTWARE**



- ✓ Créé en 2003 en Californie
- ✓ En 2013 la société est introduite en bourse au New York Stock Exchange
- ✓ Des versions majeures régulières
- ✓ Actuellement : Tableau 10.5

## **DÉCOUVRIR TABLEAU SOFTWARE**



2014 2015 2016 2017

Tableau est leader sur ce marché depuis 2014.

#### **DÉCOUVRIR TABLEAU SOFTWARE**

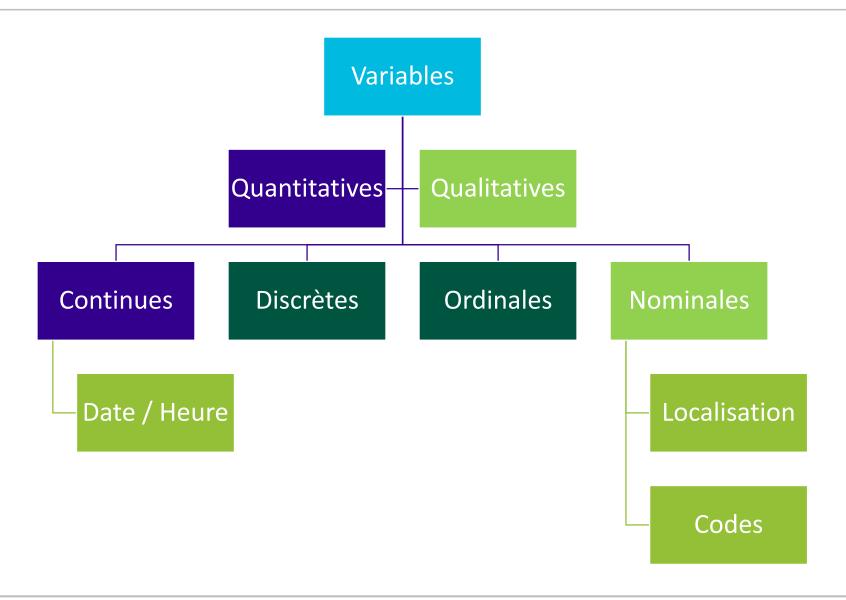
#### Ses concurrents:

- ✓ PowerBI
- ✓ Spotfire TIBCO
- ✓ Qlik









#### 1. Continue (ou quantitative ou métrique)

Exemple: x, y, temps, température, argent

#### 2. Ordinale

On peut mettre les valeurs en ordre, mais on ne peut pas dire qu'une telle valeur est N fois plus grande qu'une autre valeur

Exemple: D.E.S., D.E.C., Baccalauréat (en ordre d'années de scolarité)

#### 3. Catégorique (ou nominale)

Il n'y a pas d'ordre naturel (sauf peut-être alphabétique, mais cela est arbitraire et dépend de la langue) Exemple: groupe d'aliments (viandes, lait, légumes et fruits, produits céréaliers) Bac en génie mécanique, Bas en génie de construction, etc. Honda, Toyota, GM, Chrysler, etc.

#### 4. Binaires

Une sorte de dimension nominale (ou ordinale) ayant deux valeurs possibles

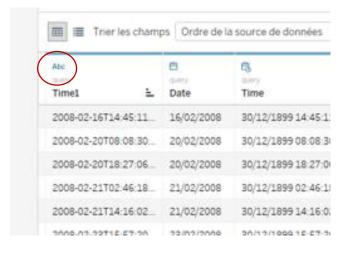
Tableau lors de la récupération de données effectue de lui-même un typage de données:

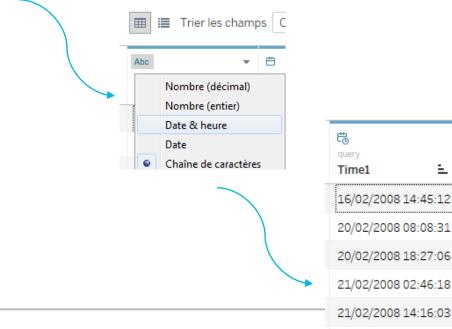
- Toute information de type qualitative sera traitée en tant que dimension : texte, date.
- Toute information de type quantitative sera traitée en tant que mesure.

On peut ensuite transformer les dimensions en indicateurs et vice versa.

Un cas courant est la présence d'une clé sous forme d'entier qui sera placée côté indicateur.

23/02/2008 15:57:20





Afin de modifier un type de données, il faut cliquer sur le logo et sélectionner le format choisi.

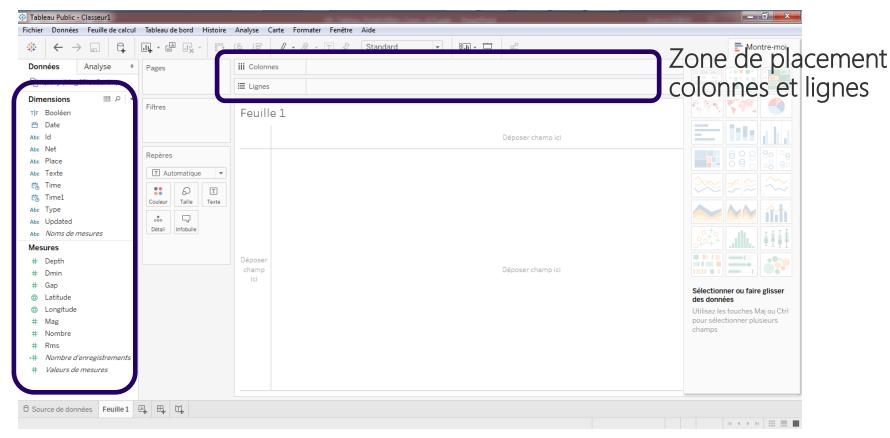
Exemple : la première colonne Time identifiée en texte, doit être convertie en donnée temps.

- 1. Cliquer sur Abc au dessus du nom de la colonne et de sa provenance
- 2. Sélectionner le nouveau type souhaité
- 3. Laisser Tableau rafraîchir les données

#### **CRÉER UNE VUE SIMPLE**

Les champs sont divisés en :

- Dimension
- Mesures.



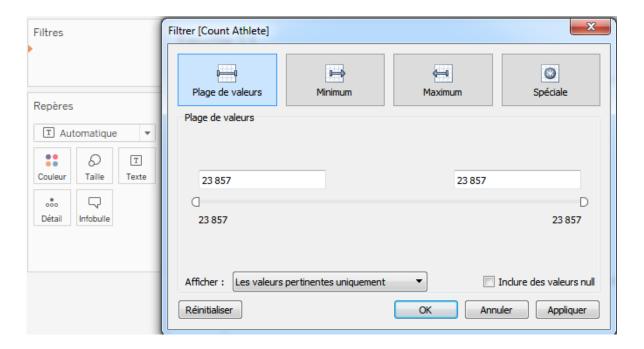
Plusieurs solution afin de créer une vue simple :

- ✓ Glisser déposer des éléments dans la zone colonne / ligne
- ✓ Glisser déposer des éléments dans la zone centrale
- ✓ Sélectionner un objet > Ajouter à la feuille

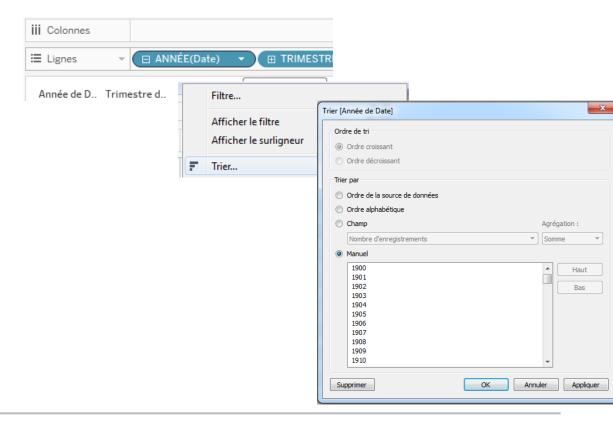
#### **CRÉER UNE VUE SIMPLE**

## Mettre en forme les objets

> Filtrer les données

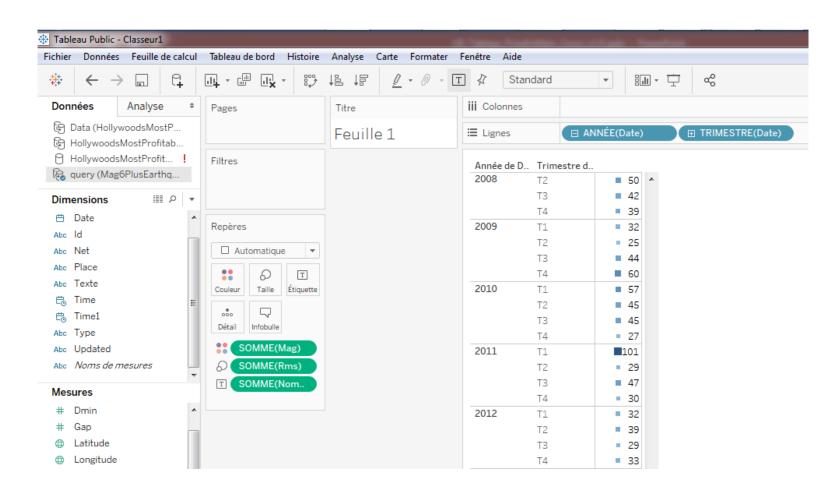


> Trier les données

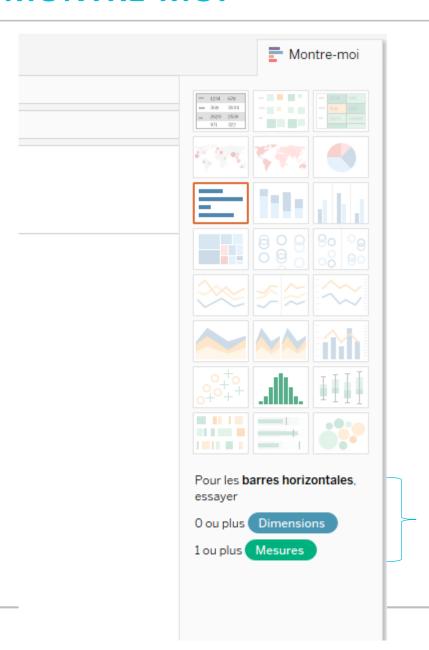


#### CRÉATION MANUELLE DE VUE

- 1 . Glisser/ Déposer les champs
- 2. Organiser les lignes et les colonnes
- 3. Naviguer dans les hiérarchies



#### **LE « MONTRE-MOI »**



A droite de Tableau, l'onglet « montre-moi »

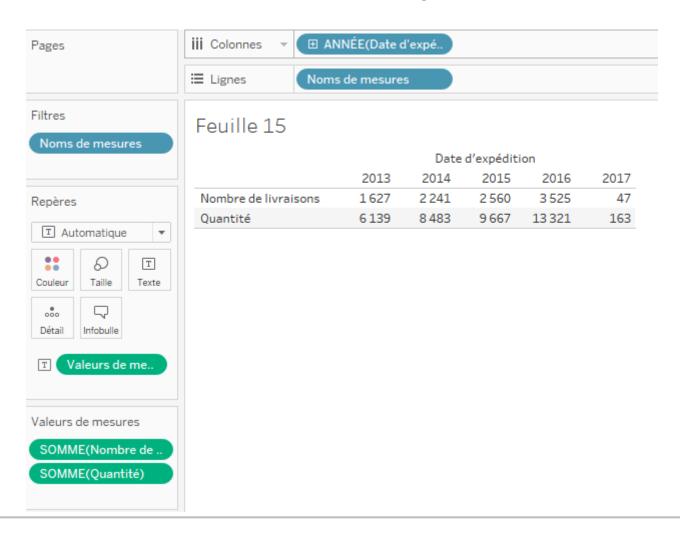
Chaque visualisation indique quelles types de données avoir

#### LE « MONTRE-MOI »

#### Méthodologie pour construire une visualisation :

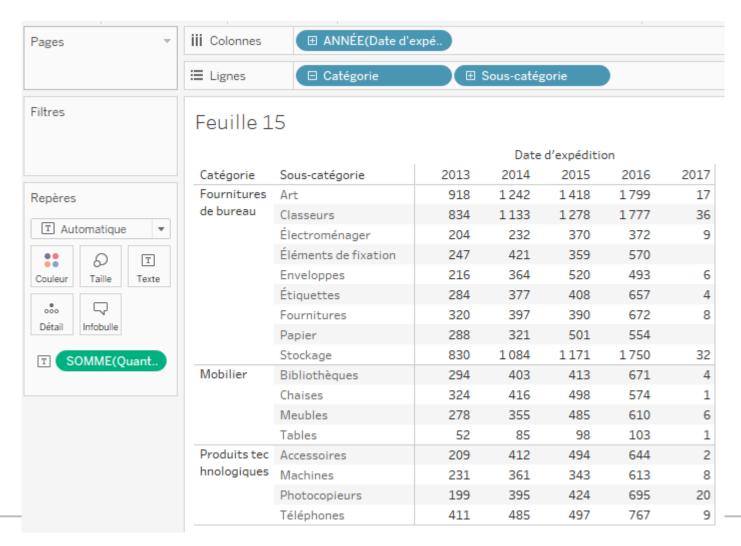
- 1. Sélectionner les champs que l'on souhaite visualiser possibilité d'en sélectionner plusieurs en maintenant la touche CTRL
- 2. Choisir la représentation via le « montre-moi »
- 3. Mettre en forme (filtres, infobulles, couleurs, taille, ....)

## Tableau Simple



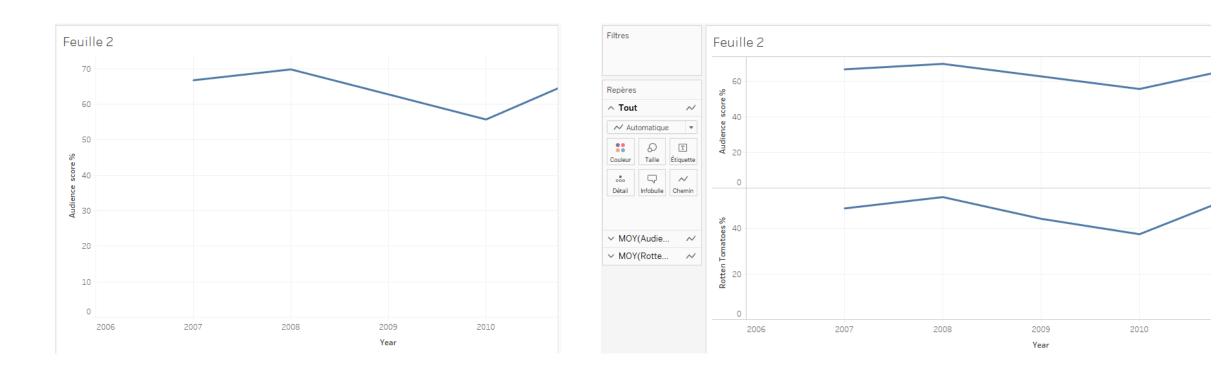
#### LES TABLEAUX

## Tableau Croisé Dynamique (TCD)



#### **LES COURBES**

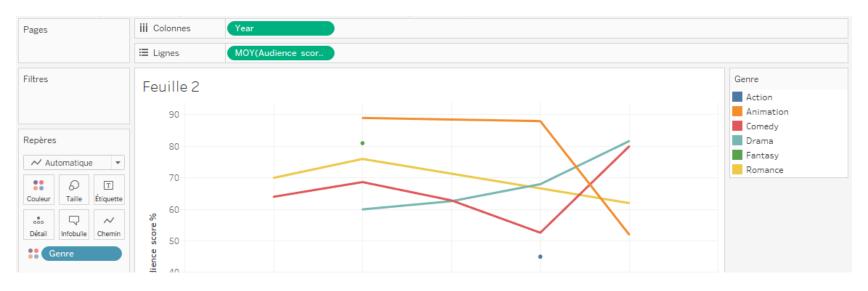
Les courbes sont souvent utilisées pour présenter les données dans le temps.



Quelle différence entre ces deux courbes ?

#### **LES COURBES**

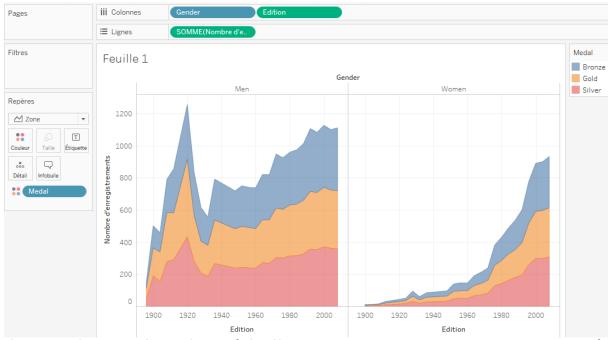
Utilisation de la couleur pour la répartition d'une dimension:



Audience par genre et par année.

#### **LES COURBES**

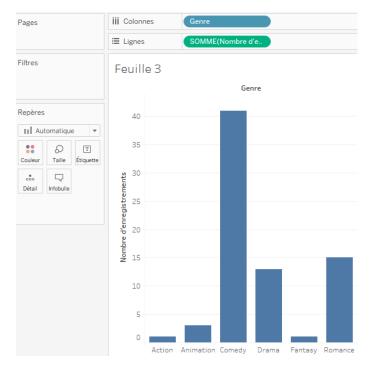
Utilisation de la couleur sur des densités:



Evolution du nombre de médaille par type, par genre et par année.

### **LES HISTOGRAMMES**

L'histogramme présente les valeurs sous forme de barres.



### Histogrammes simples :

- 1 valeur continue sur un axe.
- 1 valeur discrète sur le second axe

### **LES HISTOGRAMMES**

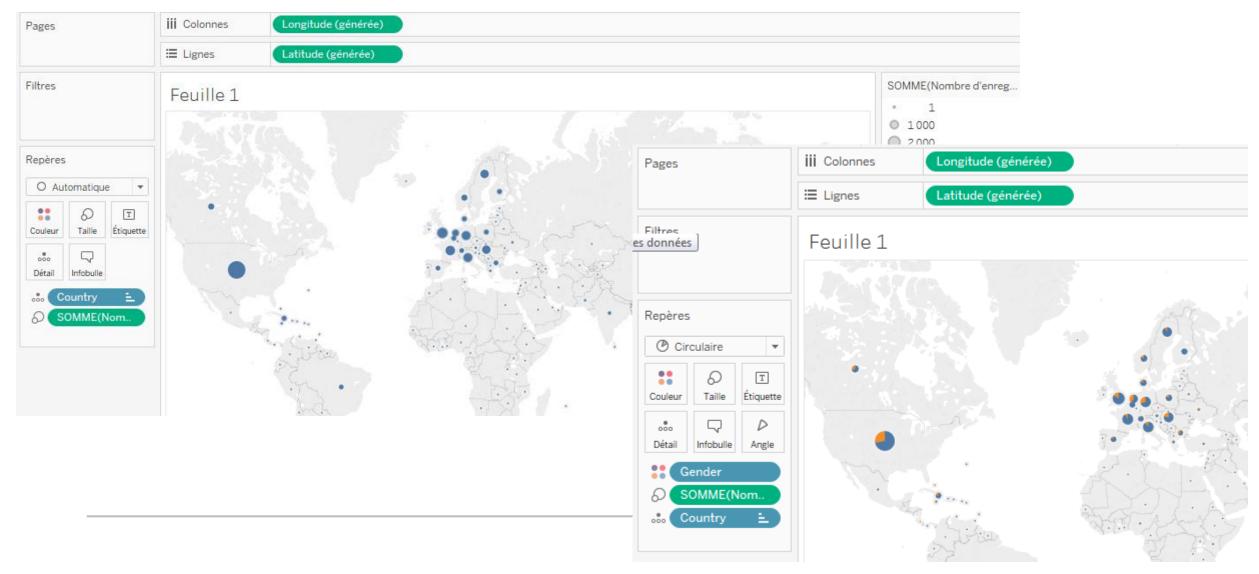
Il existe plusieurs « types » d'histogrammes :



- -> ces 3 histogrammes ont la même construction :
  - 1 valeur continue sur un axe
  - 1 valeur discrète sur le second axe
  - 1 autre dimension sur un axe ou en couleur

### **CARTOGRAPHIE**

## Carte de symbole

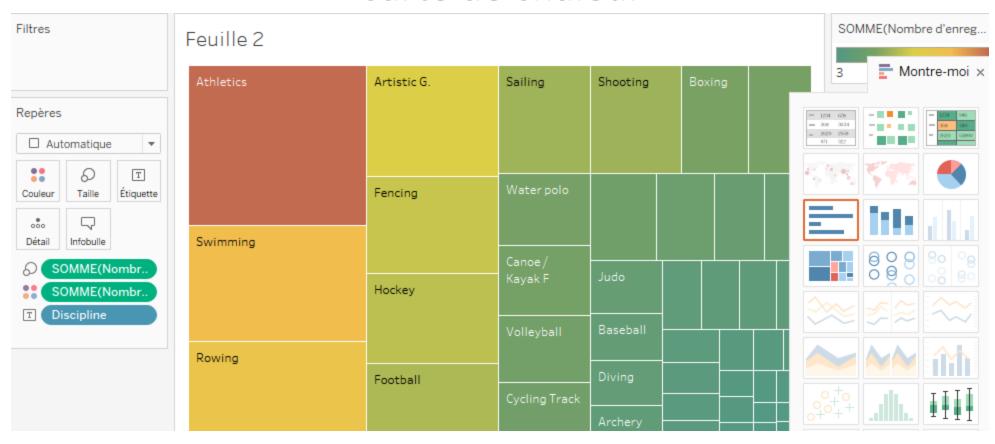


### **CARTOGRAPHIE**

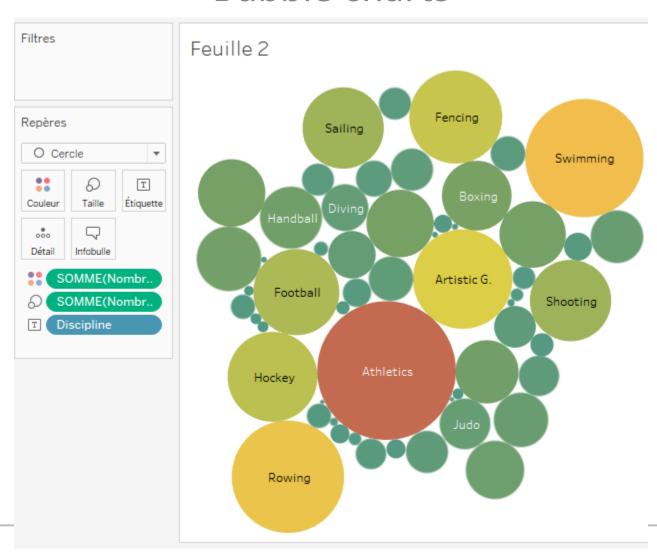
## Carte pleine



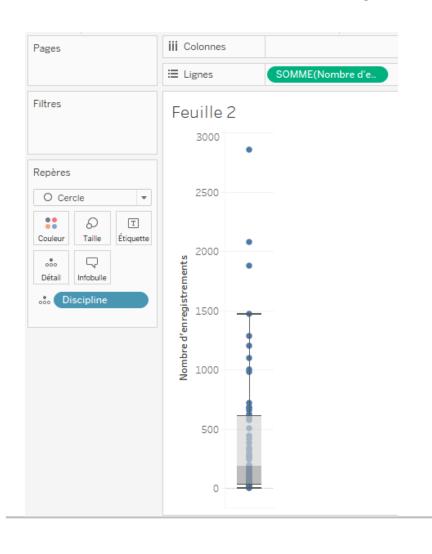
### Carte de chaleur



## **Bubble charts**



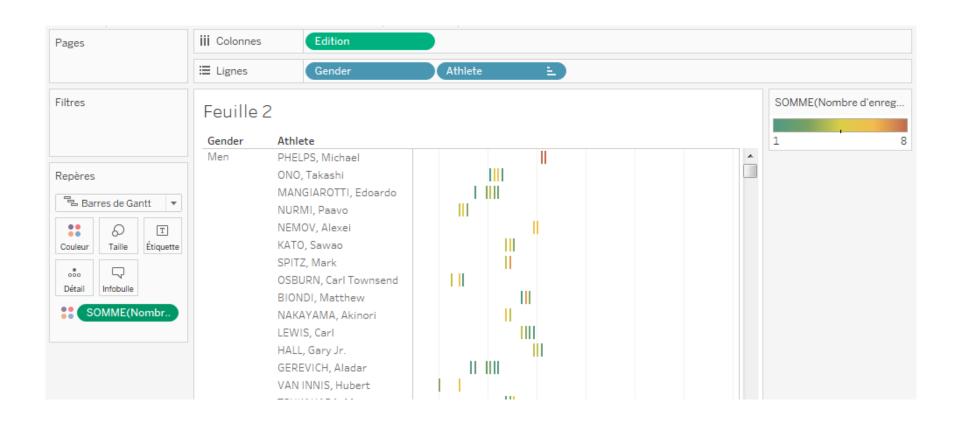
## Boxplot / Boite à moustaches



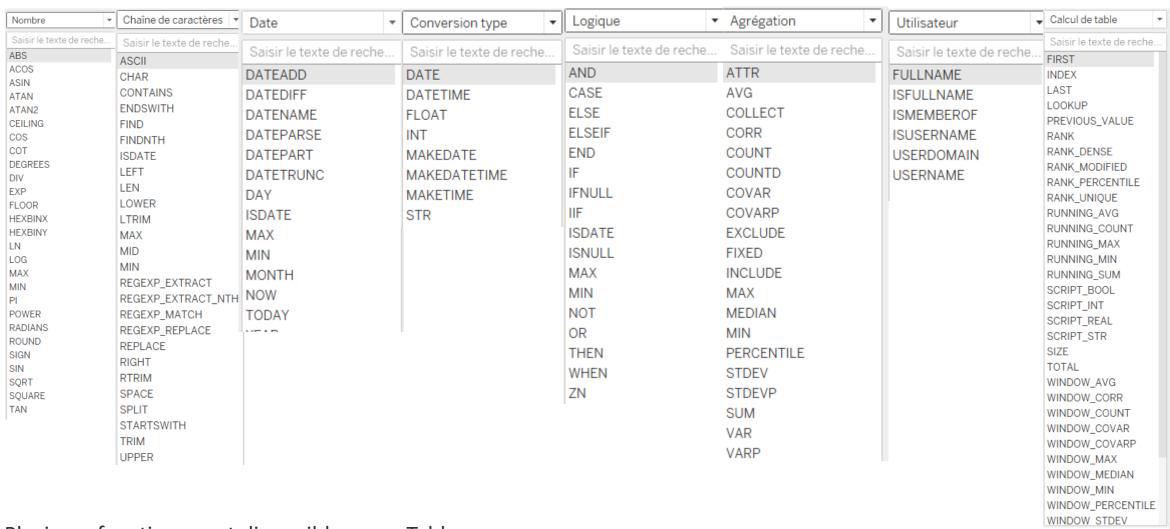
Rappel : la boîte à moustache trace un rectangle allant du premier quartile au troisième quartile et coupé par la médiane.

Elle permet d'analyser la répartition d'une population et d'isoler les cas extrêmes.

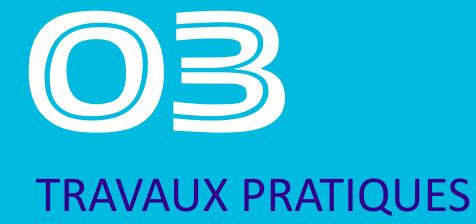
## Diagramme de GANT



### **CHAMPS CALCULÉ**



-> Plusieurs fonctions sont disponibles sous Tableau



## TABLEAUX DE BORDS ET HISTOIRE DE DONNÉES

- 1. Mise en forme simple des tableaux de bords
- 2. Gestion des filtres
- 3. Gestion des actions
- 4. Utilisation des conteneurs
- 5. Les histoires de données

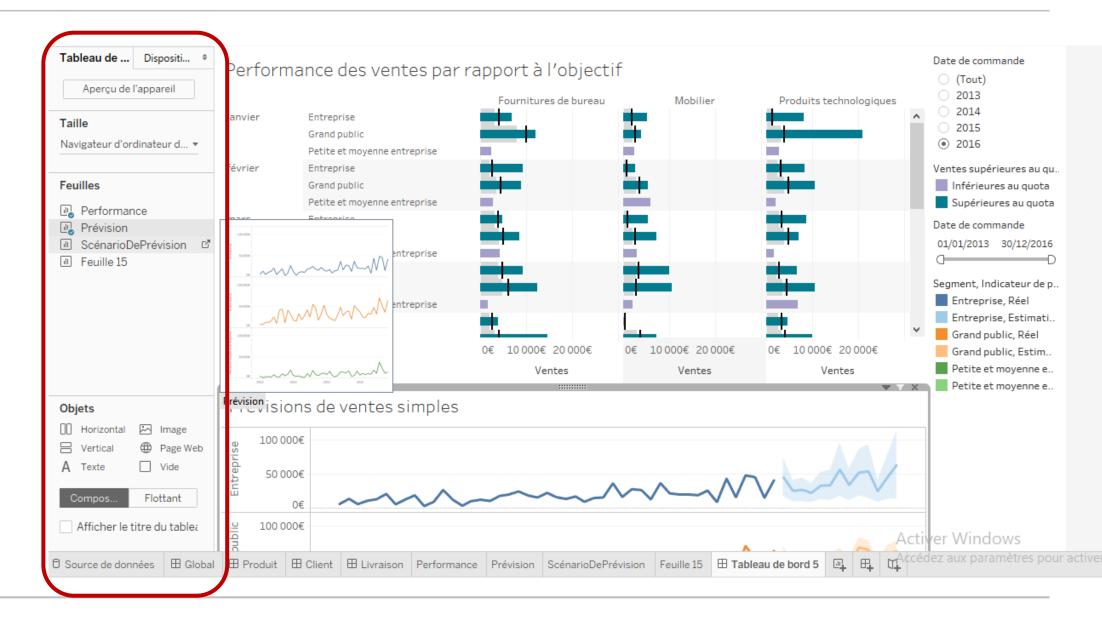
## Qu'est-ce qu'un tableau de bord ?

- ✓ Le tableau de bord est une combinaison de vues.
- ✓ Le placement se fait par glisser déposé de vues ou de conteneurs.
- ✓ Les filtres et paramètres sont automatiquement ajoutés sur la partie droite.

### Processus itératif de construction d'un TdB

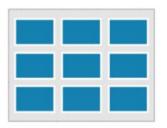
- 1. Définir le besoin
- 2. Définir les messages (1 graphique = 1 message)
- 3. Schématiser sur papier
  - Avec de vraies valeurs
  - Pour échanger avec l'audience attendue
- 4. Intégrer les données
- 5. Apporter une contextualisation
  - Commentaires
  - Légende
  - Couleurs adaptées
  - Glossaire
  - Etc.
- 6. Donner un plan d'actions

Gestion des paramètres de mise en forme des tableaux de bord

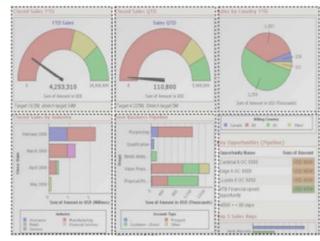


## Exemple d'organisation

#### LE SENS DE LECTURE

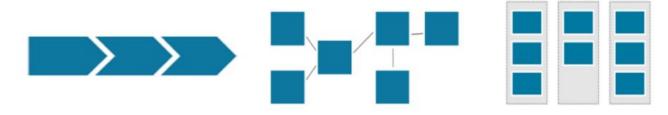


La disposition traditionnelle des éléments d'un TdB est dans une grille uniforme. Cette disposition ne véhicule aucun indice sur le début, le sens de la lecture ou les informations importantes



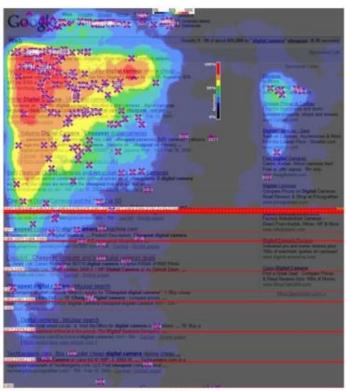
Exemple de disposition traditionnelle

D'autres dispositions sont possibles : en flux, relationnelle, groupée...
 Elles peuvent être réunies ou inbriquées les unes dans les autres, sur un tableau de bord.



- L'œil d'une personne va instinctivement être attiré par différentes zones d'un document.
- Les informations les plus importantes, les plus consultées ou le début de la lecture des données doivent se trouver dans ces zones de vues prioritaires.





Etude d'eye Tracking réalisé sur le moteur de recherche Google.





Filtres et

légendes

### 1. RÈGLES D'OR POUR LES TABLEAUX DE BORD

### 1. PRIVILÉGIEZ LES CONFIGURATIONS PAR DÉFAUT

Bien choisir (et respecter) l'outil pour aller plus vite

### 2. DESIGN FIRST, DATA SECOND

Présentez le tableau de bord avant l'étape de la connexion de données.

### 3. UN GRAPHIQUE POUR UN BESOIN

Simplicité et pertinence plutôt qu'exhaustivité

#### 4. LES HUMAINS NE SONT PAS DES ROBOTS

évitez d'écrire 10.5.2017 mais plutôt 10 Mai 2017

### 5. 3 FILTRES PAR ÉCRAN MAXIMUM

Au delà de 3 filtres, recommencez une nouvelle histoire sur une nouvelle page

### 2. GESTION DES FILTRES

Il est préférable autant que possible d'appliquer les filtres au niveau le plus bas (ex: la source de données).

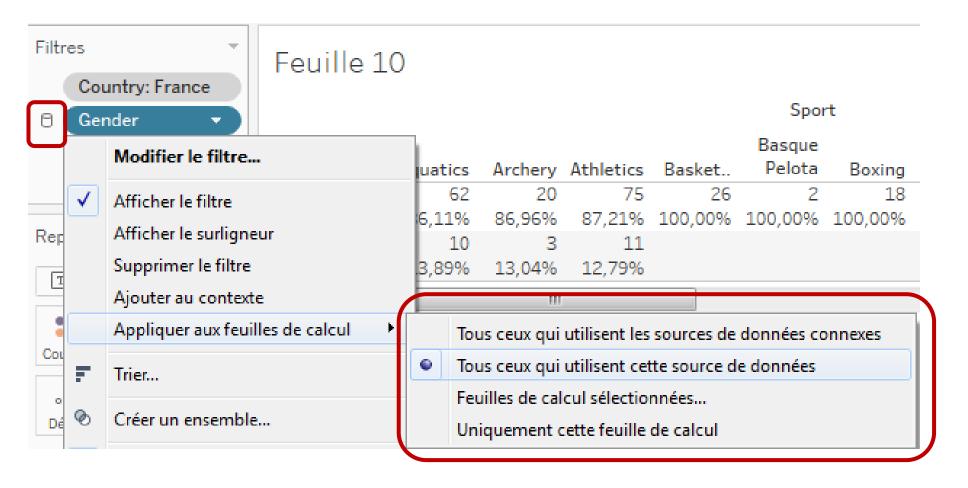
Cela permet automatiquement d'appliquer un filtre à l'ensemble des feuilles et donc du tableau de bord.

Cette action est matérialisée par une icone de base de données à gauche du filtre.

Note : un filtre contextuel peut également s'appliquer à la source de données.

### 2. GESTION DES FILTRES

Icône de base de données



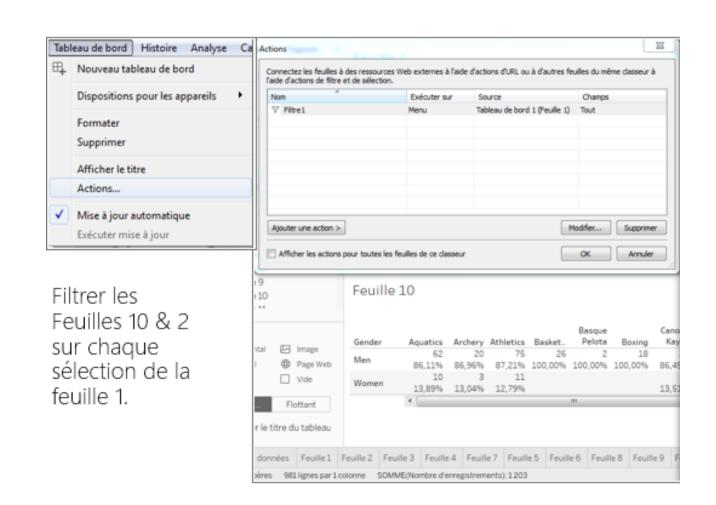
Filtres possibles

### 3. GESTION DES ACTIONS

Il existe les actions de type:

- Filtre
- Sélection
- URL

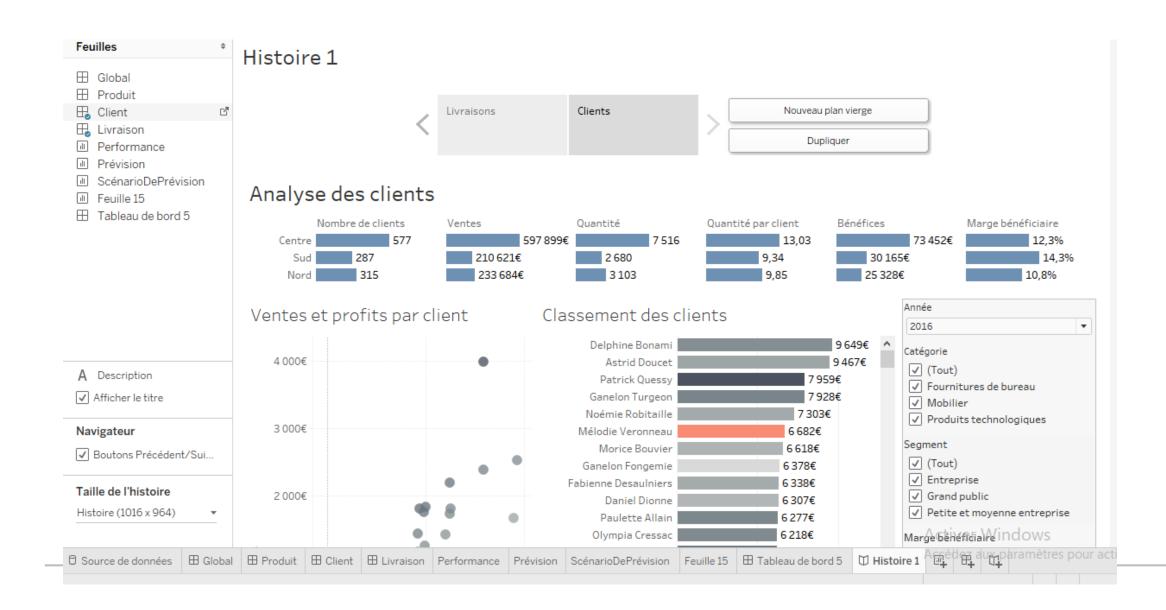
Les actions permettent d'ajouter de l'interactivité aux données



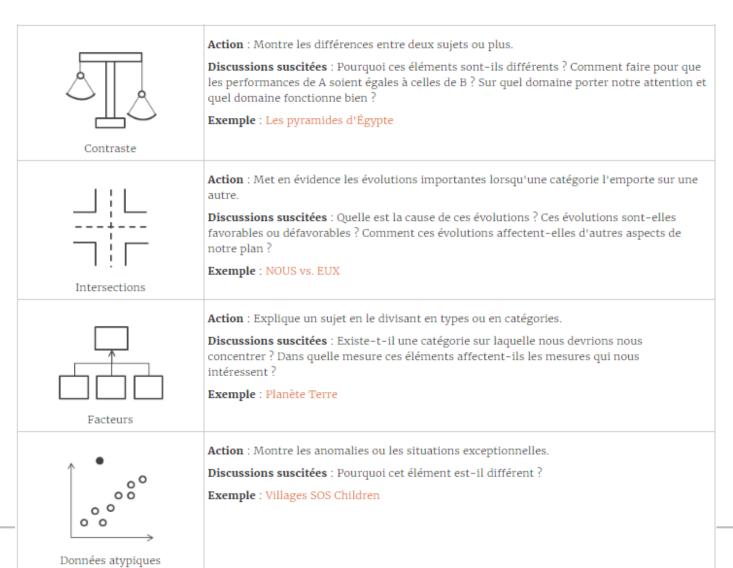
Une histoire de données est une combinaison de vues et de tableaux de bords organisées dans un ordre.

L'histoire de données est à utiliser lors de présentation. On peut le comparer à une présentation PowerPoint dans le défilement des visuels.

La construction est similaire aux Tableaux de Bords.



#### Les 7 utilités des histoires de données selon Tableau :



Type d'histoire basée sur des données	Description
Changement dans la durée	Action : Utilise une chronologie pour illustrer une tendance.  Discussions suscitées : Pourquoi cela s'est-il produit, et pourquoi la situation se répète-t-elle ? Que pouvons-nous faire pour éviter ou faire que cela arrive ?  Exemple : Blessures en série à l'Arsenal
Descendre dans la hiérarchie	Action : Définit le contexte afin que votre public comprenne mieux ce qui se passe dans une catégorie spécifique.  Discussions suscitées : Qu'est-ce qui fait la spécificité de cette personne, de cet endroit ou de cette chose ? Quelles sont les performances respectives de cette personne, de ce lieu ou de cette chose ?  Exemples : Tell Me About Will, The Simpsons Vizipedia
	Action : Décrit comment un point important pour votre public se situe par rapport au tableau d'ensemble.  Discussions suscitées : Comment un point qui vous tient à cœur se situe-t-il par rapport au tableau d'ensemble ? Quel effet un domaine a-t-il sur l'ensemble ?  Exemple : Cyclistes de Vancouver
Zoom arrière	

- ✓ On peut utiliser des histoires pour créer un business case ou simplement pour raconter un enchaînement d'événements comme on pourrait le faire avec une présentation PowerPoint par exemple.
- ✓ On peut également partager votre histoire avec d'autres personnes en publiant son classeur dans Tableau Server, Tableau Online ou Tableau Public.

### **TABLEAU SOFTWARE**

- ✓ Télécharger Tableau Software (version essai 15j) : www.tableau.com/products/trial
- ✓ Enoncé du TP : github.com/apatou/Enseignement\_ESCEN

# MERCI