

# Meetup Visualisation des données Toulouse

## 13 Avril 2016

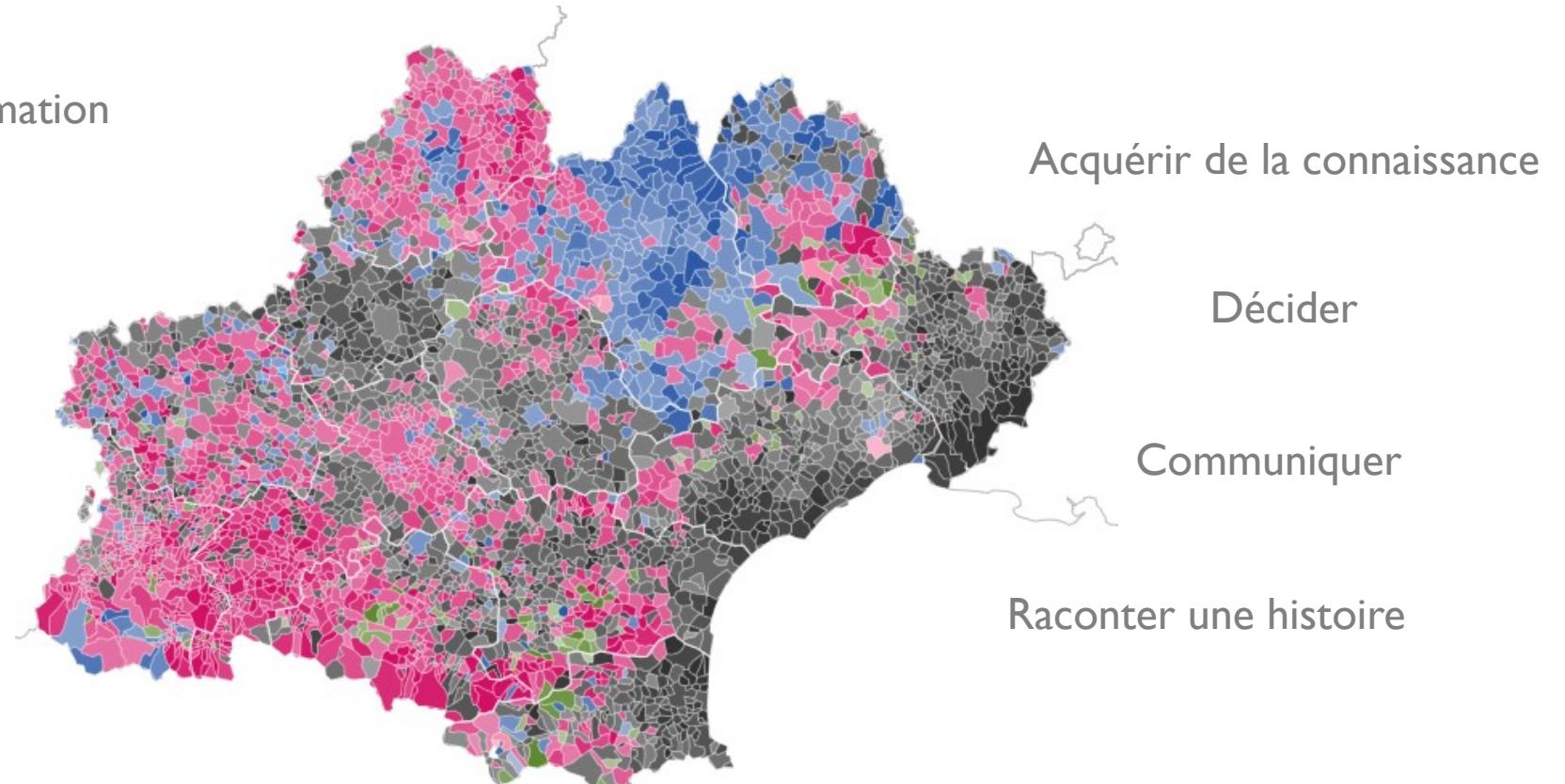
Exploration interactive de données avec les librairies d3.js ,  
dc.js et crossfilter.js : **une bonne solution ?**

Alain Ottenheimer

# Un Explorateur de données : Pourquoi faire ?

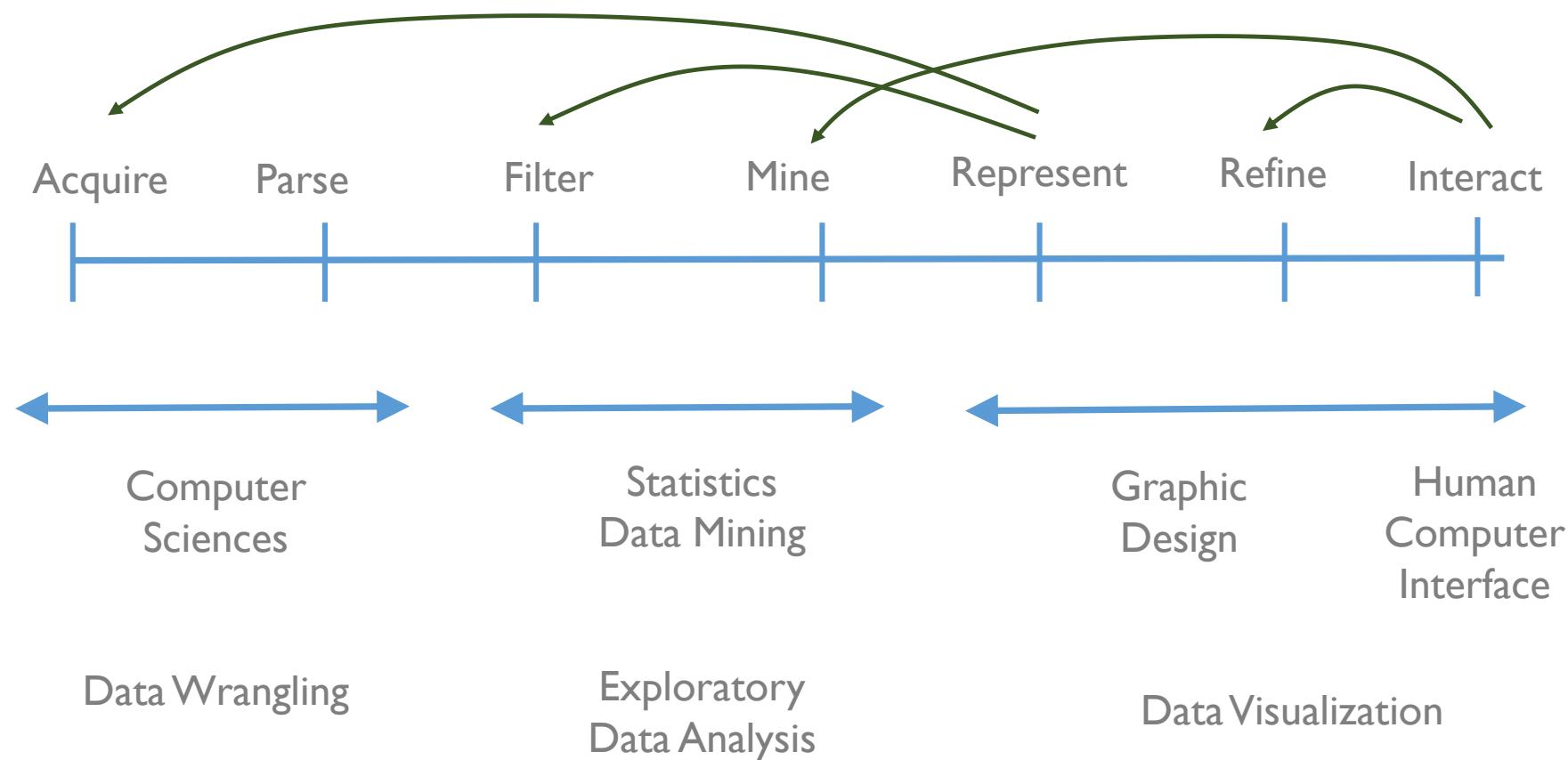
Computer based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively. Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision making. *Tamara Munzner*

Faciliter l'accès à l'information  
Rechercher des corrélations  
Se poser des questions  
Déetecter des signaux faibles



[Le monde: Régionales - listes en tête au 1<sup>er</sup> tour](#)

# Construire un explorateurs de données : Les 7 étapes et leurs interactions



# Quel jeu de données ?

<https://data.toulouse-metropole.fr/explore/?sort=modified&q=Toulousains>

**DATA.toulouse-métropole**  
Le portail des données ouvertes du territoire

Démarche Licence Données Cartes API Communes Réutilisations Contacts

**237** données libérées

Trier par Dernière modification ▾

Filtres

Trouver un jeu de données...

VUE

- Carte 142
- Analyse 51
- Image 2
- Calendrier 1

MODIFIÉ

- 2012 3
- 2013 30
- 2014 34
- 2015 158
- 2016 12

Récup' Ordures Ménagères Localisation des colonnes enterrées pour le dépôt des Ordures Ménagères sur le territoire de Toulouse Métropole.

Producteur: Mairie de Toulouse  
Licence: Open Database License (ODbL)  
Données: 152

déchet ordure ménagère sac poubelle recyclage orne pav points d'apport volontaire

Tableau Carte Export API

Localisation des systèmes de co... Localisation des systèmes de contrôle d'accès gérés par Toulouse Métropole. Ce sont des bornes escamotables, chaînes ou barrières actionnées sur autorisation réparties sur le territoire de la Métropole.

Producteur: Mairie de Toulouse  
Licence: Open Database License (ODbL)  
Données: 50

contrôle accès autorisation borne barrière chaîne escamotable circulation donnée ouverte

Tableau Carte Export API

Implantation des modules com... Localisation des modules fitness en accès libre sur le territoire de la ville de Toulouse.

Producteur: Mairie de Toulouse  
Licence: Open Database License (ODbL)  
Données: 30

sport fitness musculation module donnée ouverte

Tableau Carte Export API

Recensement population 2012 - ... Données du recensement de la population 2012 à l'échelle des grands quartiers sur le thème Logement.

Producteur: Mairie de Toulouse  
Licence: Licence Ouverte (Etalab)  
Données: 60

logement maison appartement construction HLM ménage résidence principale

Tableau Carte Analyse Export API

Recensement population 2012 - ... Données du recensement de la population 2012 à

Recensement population 2012 - ... Données du recensement de la population 2012 à

**DATA.toulouse-métropole**  
Le portail des données ouvertes du territoire

Démarche Licence Données Cartes API Communes Réutilisations Contacts

**1** donnée libérée

Trier par Dernière modification ▾

Filtres Tout effacer

Trouver un jeu de données...

VUE

- Analyse 1

MODIFIÉ

- 2015 1

PRODUCTEUR

- Ville de Toulouse 1

THÈME

- Culture 1
- Patrimoine 1

Tous les jeux de données sont affichés.

Toulousains morts pour la France... Liste des Toulousains morts pour la France entre le 2 août 1914 et le 24 octobre 1919, soldats de l'armée de terre, inscrits dans la version du livre d'or du ministère des pensions conservée aux Archives municipales de Toulouse sous la référence 5H292.

Producteur: Ville de Toulouse  
Licence: Open Database License (ODbL)  
Données: 3 899

Tableau Analyse Export API

Jeu de données des Toulousains morts pour la France 14-18 :

*Nom, Prénoms, Date de naissance, Ville de naissance, Département de naissance (n°), Pays, Régiment, Grade, Date du décès, Département de décès (n°), Lieu du décès, Pays de décès*

# Licence d'utilisation

Les données publiques de Toulouse métropôle sont sous licence Open Database (ODbL)

L'ODbL est un contrat licence de base de données favorisant la libre circulation des données qui permet à chacun d'exploiter publiquement, commercialement ou non, des bases de données; à condition néanmoins de maintenir la licence sur la base de données, et éventuellement, sur les modifications qui y sont apportées, et de mentionner expressément l'usage, s'il génère des créations à partir de celles-ci.

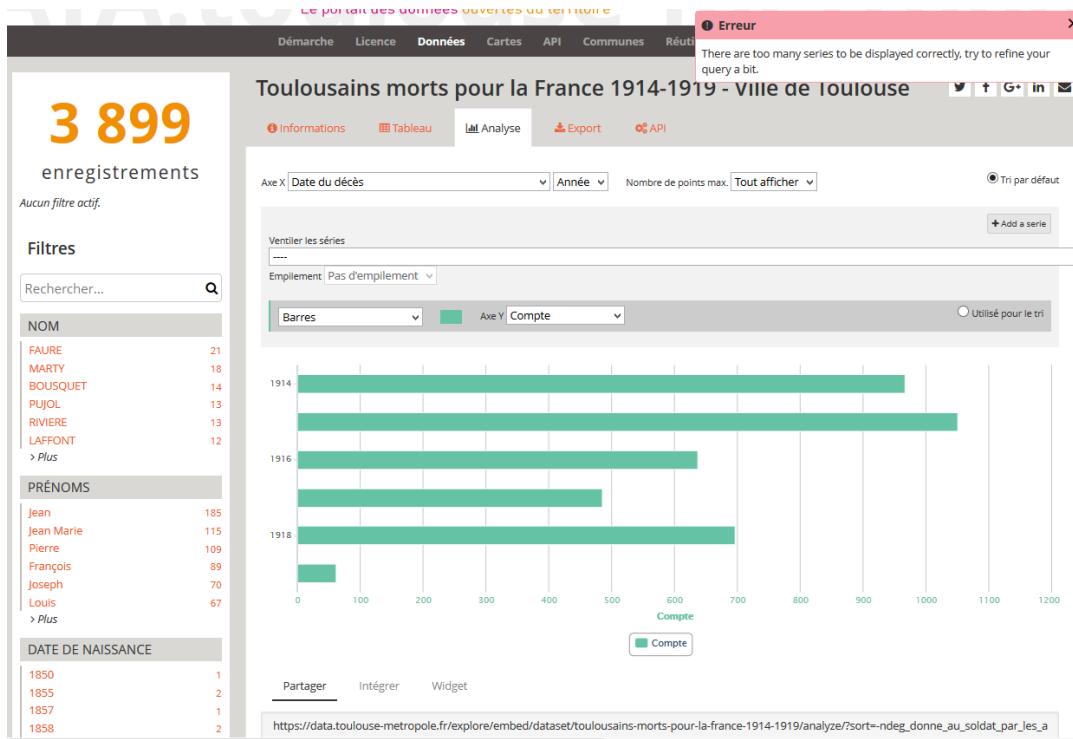
C'est un licence libre est une licence s'appliquant à une œuvre de l'esprit par laquelle l'auteur concède tout ou une partie des droits que lui confère le droit d'auteur, en laissant au minimum quatre droits considérés fondamentaux aux utilisateurs :

- usage de l'œuvre ;
- étude de l'œuvre pour en comprendre le fonctionnement ou l'adapter à ses besoins ;
- modification (amélioration, extension, transformation) ou incorporation de l'œuvre en une œuvre dérivée ;
- redistribution de l'œuvre, c'est-à-dire sa diffusion à d'autres usagers, y compris commercialement.

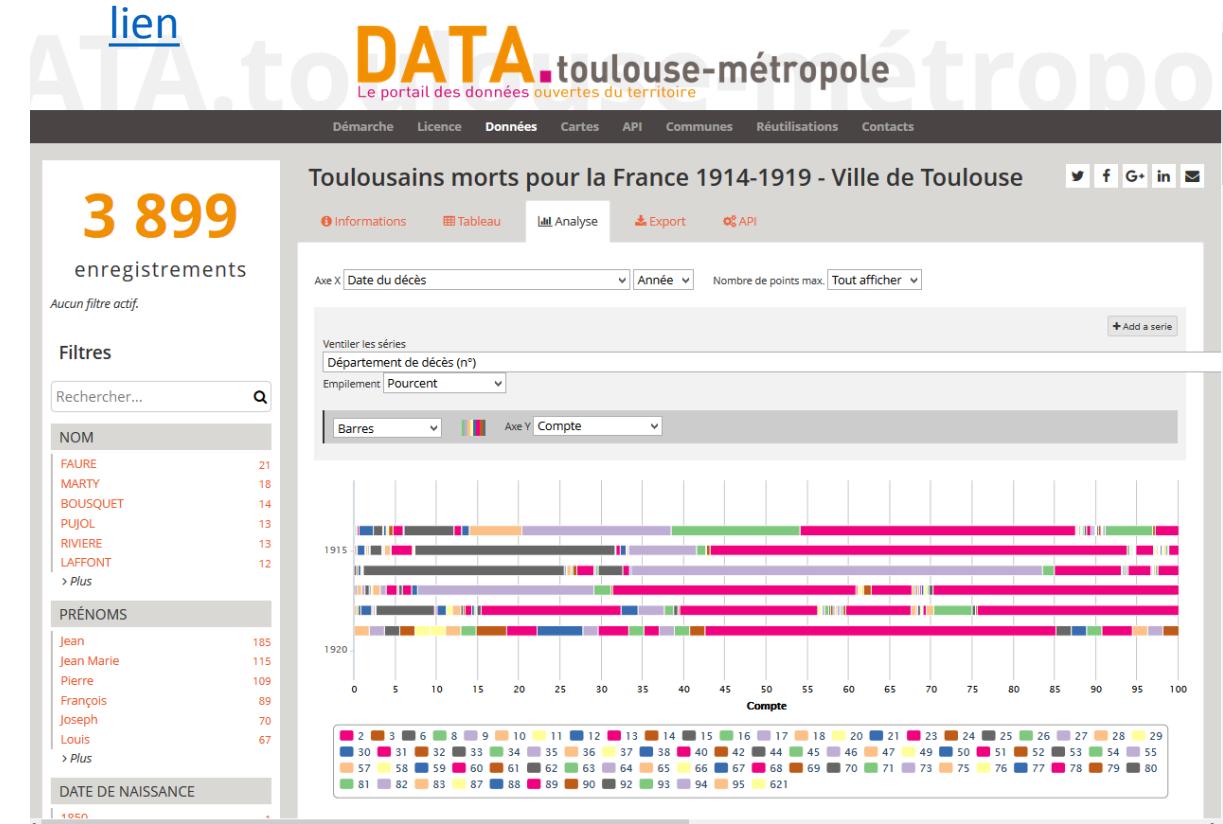
# Quel questionnement suggère ce jeu de données ? Quelles réponses apportent les outils standards ?

## L'outil générique du site open data Toulouse métropole

[Lien](#)

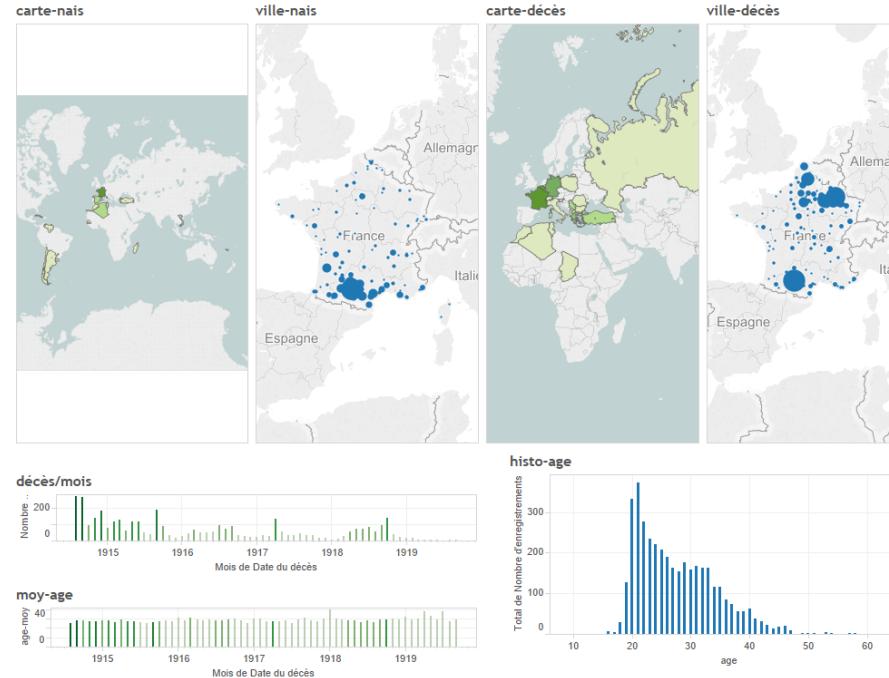


[lien](#)



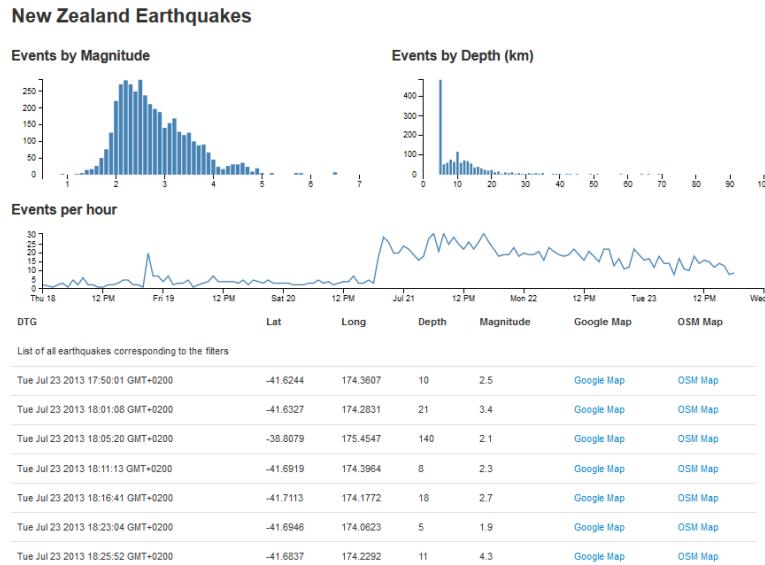
# Peut-on faire mieux avec Excel ou Tableau Software.....?

Tableau software



# Et avec les librairies javascript d3.js+dc.js+crossfilter.js !

Exemple



d3.js : Library released under [BSD license](#). Copyright 2015 [Mike Bostock](#).

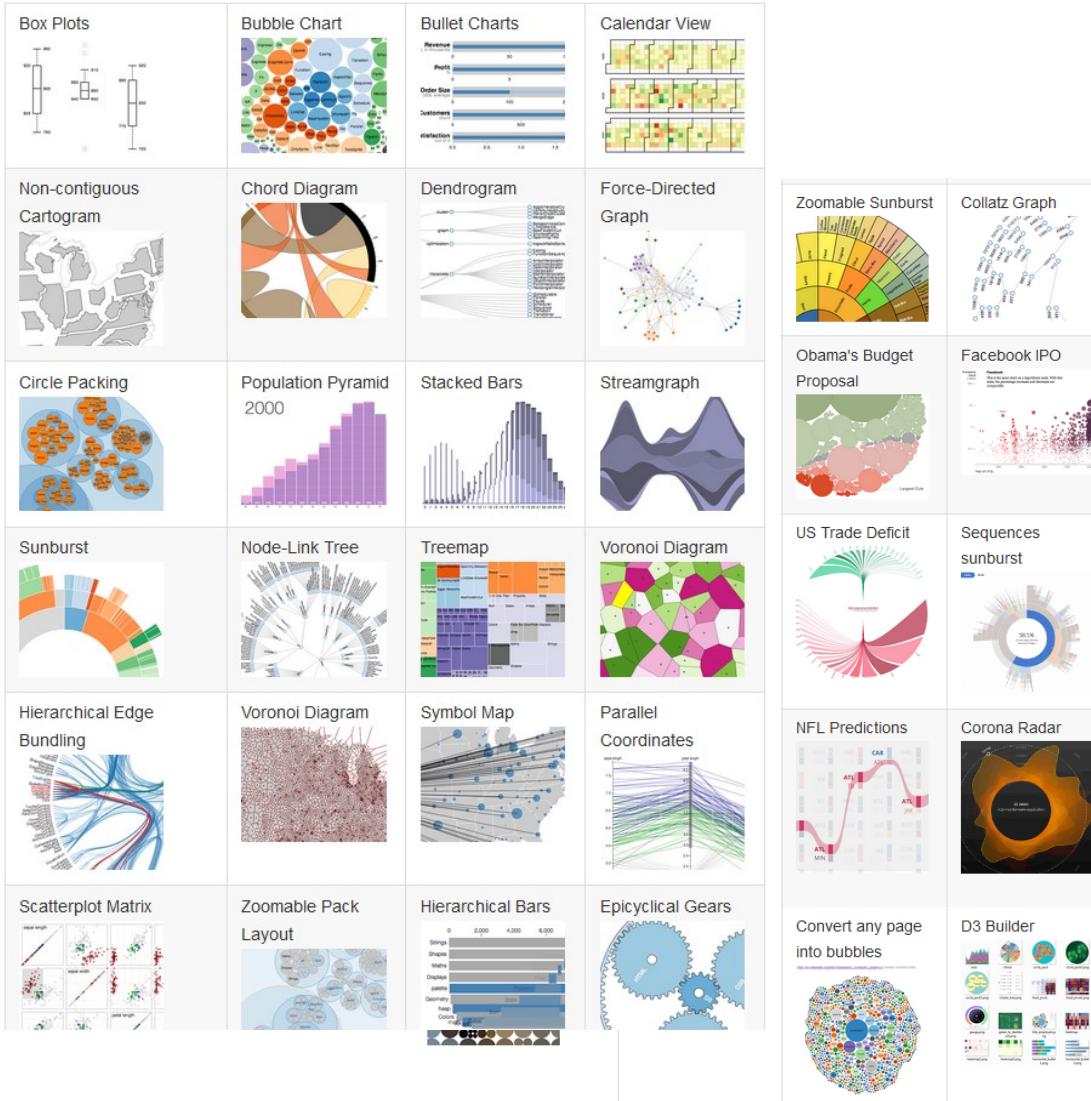
dc.js : Copyright 2012-2016 Nick Zhu & the dc.js Developers Licensed under the Apache License, Version 2.0

Crossfilter.js : Licensed under the Apache License, Version 2.0

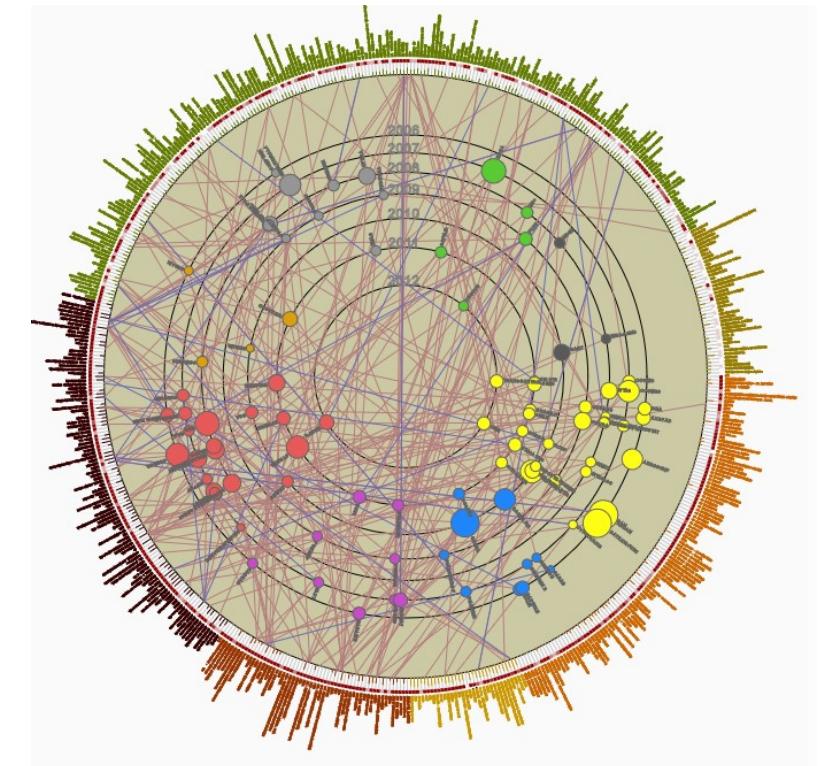
.....

# d3 : la référence !

- Toutes les sortes de graphiques

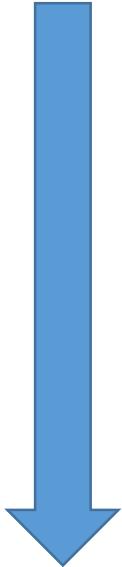


- ou créer vos propres graphiques



# En résumé

- Ranking :



|   |  |
|---|--|
| Graphiques prédéfinis                           | → Excel, Tableau software...!                        |
| Charting libraries : dc, dimple, NVD3,<br>d3.js | → Python, Ruby!<br>Ember-chart.js.....<br>→ C / C++! |
| Web GL, SVG, ...                                | → Assembleur!  |

+ de flexibilité

- Exploratoire / Explicatif
- d3 : Data Driven Document → Interactivité

# Réalisation de l'explorateur interactif avec d3.js , dc.js et crossfilter.js

Un peu de code ...

# Structure de la page html

```
<!DOCTYPE html>
<head>
Titre
Appel des librairies bootstrap.js, jquery.js, d3.js, dc.js et crossfilter.js
Appel des css associés
</head>

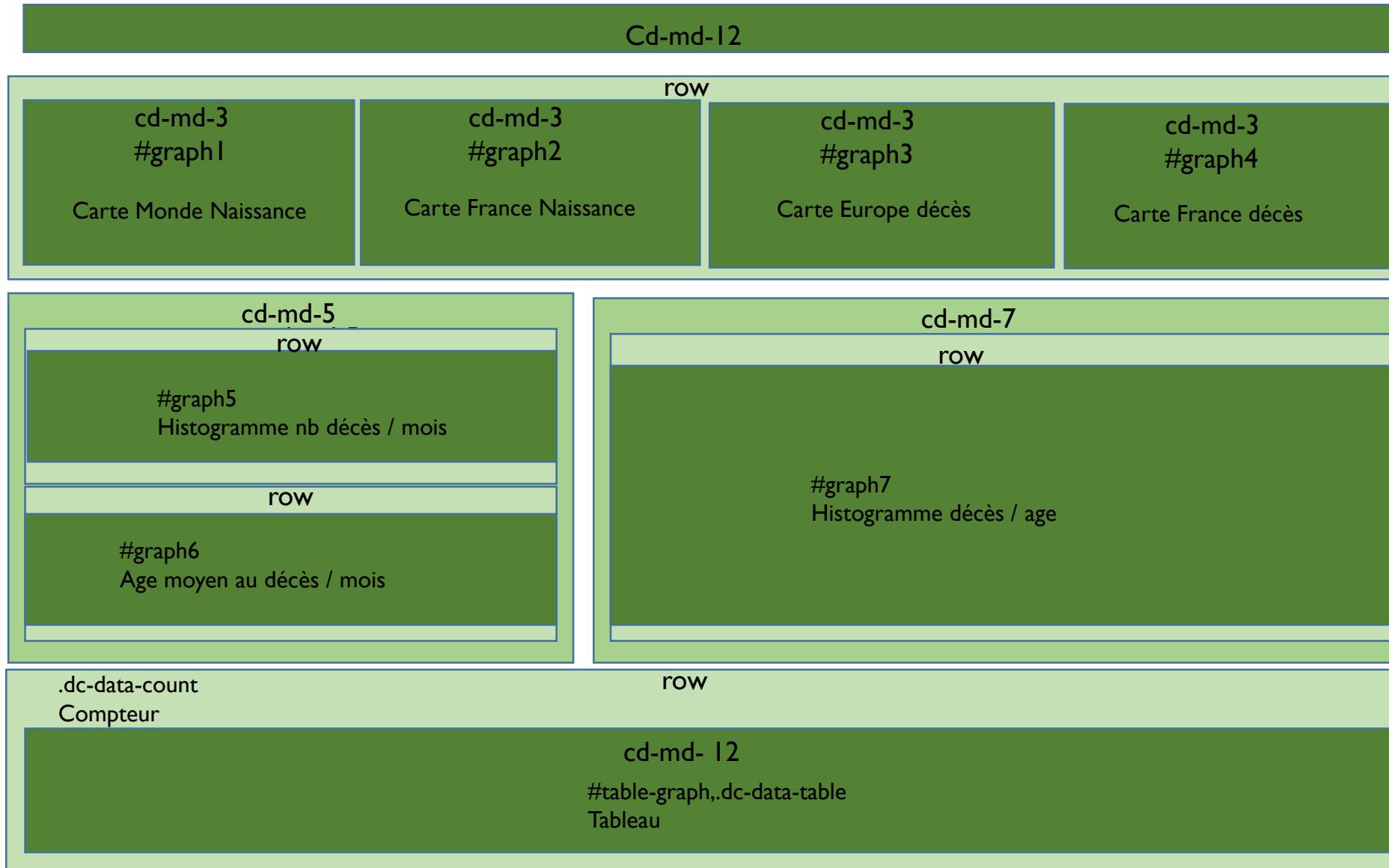
<style>
css de la page
</style>

<body>
Pagination (avec Bootstrap.js)

<script type="text/javascript">
code javascript de l'explorateur (avec d3.js, dc.js et crossfilter.js)
</script>

</body>
</html>
```

# Pagination - bootstrap.js....



# Structure de l'explorateur

```
<script type="text/javascript">

    • Déclaration des graphes

    • Acquisition et mise en forme des données

    • Traitement des données

    • Réalisation des graphiques
        4 cartes géographiques
        3 histogrammes
        1 compteur
        1 table

</script>
```

# Les formats des données utilisées

Données .csv

```
Nom,Prénoms,Date_de_naissance,Ville_de_naissance,Département_de_naissance,Pays,Régiment,Grade,Date_du_décès,Département_de_décès,Lieu_du_décès,Pays de décès,  
CHAYNES,Louis,1886-01-17,Toulouse,31,France,281e Infanterie,Soldat,29/07/1916,80,Rosières-en-Santerre,France,  
CHAYNES,Marius Georges Jean,1897-12-23,Toulouse,31,France,327e Infanterie,Soldat,14/09/1918,2,Vailly-sur-Aisne,France,  
.....
```

Données géographiques .json

```
{"type": "FeatureCollection", "features": [  
  { "type": "Feature", "properties": { "ID_GEOFLA": 1, "CODE_DEPT": "1", "NOM_DEPT": "AIN", "CODE_CHF": "053", "NOM_CHF": "BOURG-EN-BRESSE", "X_CHF_LIEU": 8717, "Y_CHF_LIEU": 65696, "X_CENTROID": 8814, "Y_CENTROID": 65582, "CODE_REG": "82", "NOM_REGION": "RHONE-ALPES" }, "geometry": { "type": "Polygon", "coordinates": [ [ [ 5.831226, 45.938460 ], [ 5.822121, 45.930136 ], [ 5.829155, 45.917237 ], [ 5.826140, 45.903274 ], [ 5.815154, 45.877201 ], [ 5.807525, 45.860640 ], [ 5.801109, 45.845796 ], [ 5.795850, 45.829798 ], [ 5.786934, 45.823224 ], [ 5.784591, 45.797224 ], [ 5.781383, 45.780841 ], [ 5.783047, 45.772465 ], [ 5.779417, 45.762247 ], [ 5.783592, 45.745204 ], [ 5.780414, 45.742922 ], [ 5.769679, 45.740816 ], [ 5.776366, 45.727920 ], [ 5.763298, .....  
  { "type": "Feature", "properties": { "ID_GEOFLA": 2, "CODE_DEPT": "2", "NOM_DEPT": "AISNE", "CODE_CHF": "408", "NOM_CHF": "LAON", "X_CHF_LIEU": 7451, "Y_CHF_LIEU": 69406, "X_CENTROID": 7404, "Y_CENTROID": 69401, "CODE_REG": "22", "NOM_REGION": "PICARDIE" }, "geometry": { "type": "Polygon", "coordinates": [ [ [ 3.485187, 48.851908 ], [ 3.470198, 48.850547 ], [ 3.461955, 48.838474 ], [ 3.445361, 48.843232 ], [ 3.452232, 48.856300 ], [ 3.428521, 48.861676 ], [ 3.423063, 48.867568 ], [ .....
```

# L'affectation des graphiques

```
var carteNMonde = dc.geoChoroplethChart("#graph1");
var carteN = dc.geoChoroplethChart("#graph2");
var carteDMonde = dc.geoChoroplethChart("#graph3");
var carteD = dc.geoChoroplethChart("#graph4");
var NbMortMoisChart = dc.barChart("#graph5");
var moyChart = dc.barChart("#graph6");
var NbMortAgeChart = dc.barChart("#graph7");
var dataTable = dc.dataTable("#table-graph,.dc-data-table");
var count = dc.dataCount('.dc-data-count');
```

# Acquisition et mise en forme des données

```
d3.csv("data/toulousains-morts-pour-la-france-1914-1919.csv", function (csv) {  
    var dtgFormat1 = d3.time.format("%Y-%m-%d");  
    var dtgFormat2 = d3.time.format("%d/%m/%Y");  
    var numberFormat = d3.format(".0f"); // pas de chiffre après la virgule  
    var dtgFormat3 = d3.time.format("%Y/%m");  
    var dtgFormat4 = d3.time.format("%b %Y");  
  
    function _calculateAge(dDeces, dNaissance) {  
        var ageDifMs = dDeces.getTime() - dNaissance.getTime(); // ms  
        var ageDate = new Date(ageDifMs);  
        return Math.abs(ageDate.getUTCFullYear() - 1970);  
    }  
  
    csv.forEach(function (d) {  
        d.dateNaissance = dtgFormat1.parse(d.Date_de_naissance);  
        d.dateDeces = dtgFormat2.parse(d.Date_du_décès);  
        d.dateDecesM = d3.time.month(d.dateDeces);  
        d.age = _calculateAge(d.dateDeces, d.dateNaissance);  
    });  
.....
```

Format des dates

Calcul de l'âge

Mise au format des dates

Date\_de\_naissance : "1880-06-18"  
dateNaissance : Date {Fri Jun 18 1880 00:00:00 GMT+0200}

# Traitemen~~t~~ des données – crossfilter 1/2

Crossfilter est une librairie javascript permettant de traiter des gros fichiers de données très rapidement.

Crossfilter apporte une fonction « map-reduce » de données en utilisant les notions de « dimensions » et « groupes ».

| Nom       | Prénoms                 | Date_de_naissance | Ville_de_naissance  | Département_de_naissance | Pays    |
|-----------|-------------------------|-------------------|---------------------|--------------------------|---------|
| ABADIE    | Jean                    | 1880-06-18        | Toulouse            | 31                       | France  |
| ABADIE    | Jean                    | 1873-12-23        | Toulouse            | 31                       | France  |
| ABADIE    | Joseph Jean Baptiste    | 1883-03-21        | Pézilla-du-Conflent | 66                       | France  |
| ABADIE    | Fernand                 | 1896-10-13        | Toulouse            | 31                       | France  |
| ABADIE    | François René           | 1883-12-17        | Toulouse            | 31                       | France  |
| ABADIE    | Marius Jean             | 1887-02-28        | Toulouse            | 31                       | France  |
| ABADIE    | Henri                   | 1898-11-02        | Saint-Martory       | 31                       | France  |
| ABADIE    | Théodore                | 1895-12-28        | Madière             | 9                        | France  |
| ABADIE    | Charles François Henri  | 1890-12-09        | Paris               | 75                       | France  |
| ABBADIE   | Émile Joseph Julien     | 1895-07-10        | Toulouse            | 31                       | France  |
| ABEILLE   | Benoît Charles          | 1898-05-10        | Cenon               | 33                       | France  |
| ABEJEAN   | Albert Jules Léon       | 1888-08-13        | Bruxelles           |                          | Belgium |
| ABEL      | Joseph Antoine          | 1882-07-26        | Toulouse            | 31                       | France  |
| ABRESSENC | Étienne                 | 1885-07-29        | Colomiers           | 31                       | France  |
| ABRIBAT   | Henri Cyprien Jules     | 1894-04-04        | Bourgoin            | 38                       | France  |
| AD        | Jean Baptiste Guillaume | 1883-12-17        | Cambernard          | 31                       | France  |

NombredptNSum

```
[Object { key="31", value=2608},  
 Object { key="9", value=189},  
 Object { key="81", value=113},  
 Object { key="11", value=109},  
 Object { key="32", value=90}].....
```

```
var data = crossfilter(csv);  
var dptN = data.dimension( function (d) {return d["dptNaissance"];});  
var NombredptNSum = dptN.group().reduceCount();
```

# Traitemen~~t~~t des données – crossfilter 2/2

Calcul de l'âge moyen par mois :

```
var data = crossfilter(csv);
var anneeDecesM = data.dimension(function (d) {return d["dateDecesM"]; });
var anneeDecesMGroup = anneeDecesM.group().reduce(
    //callback for when data is added to the current filter results
    function (p, v) {
        ++p.count; // compte le nbr de décès pour chaque mois
        p.sum += v['age']; // Fait la somme des ages pour chaque mois
        p.moy = Math.round(p.sum / p.count); calcul la moyenne / mois
        return p;
    },
    // callback for when data is removed from the current filter results
    function (p, v) {
        --p.count; p.sum -= v['age']; p.moy = Math.round(p.sum / p.count);
        return p;
    },
    // initialize p
    function () { return { count: 0, sum: 0, moy: 0, };}
);
```

Résultat : anneeDecesMGroup

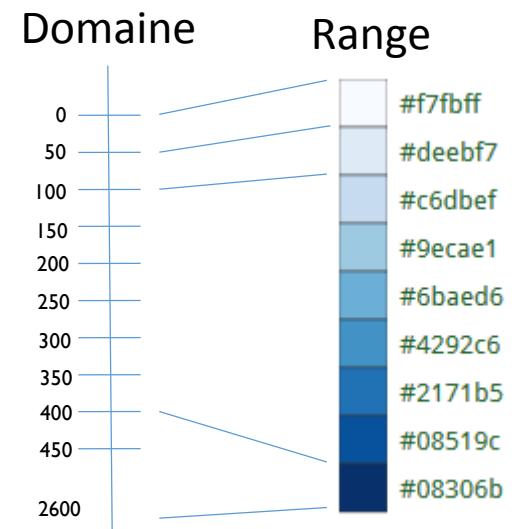
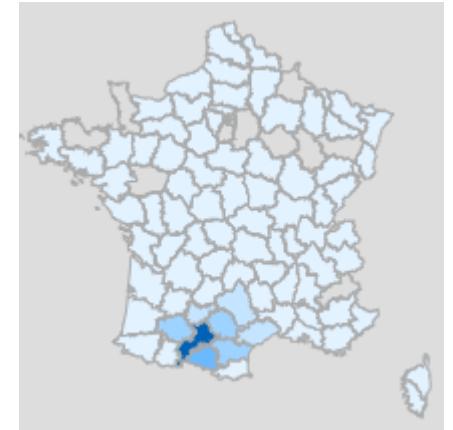
```
Key Date {Tue Sep 01 1914 00:00:00 GMT+0200}
Value Object { count=270, sum=7470, moy=28}
```

```
Key Date {Thu Oct 01 1914 00:00:00 GMT+0200}
Value Object { count=97, sum=2754, moy=28}
```

.....

# Graphiques : carte des naissances

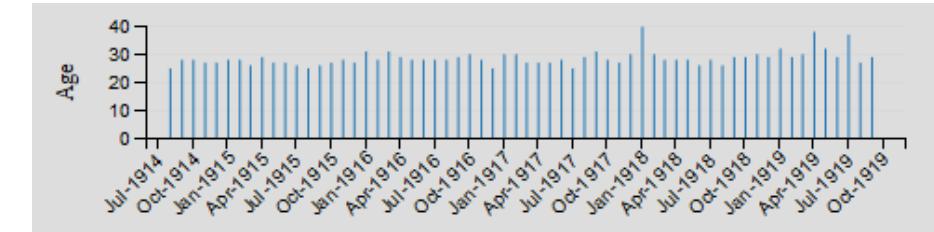
```
//#2 carte des lieux de naissance
d3.json("data/departements.json", function (dptJson) { carteN.width(250).height(250)
  .dimension(dptN)
  .group(NombredptNSum)
  .colors(d3.scale.quantize().range(["#E2F2FF", "#C4E4FF", "#9ED2FF", "#81C5FF",
    "#6BBAFF", "#51AEFF", "#36A2FF", "#1E96FF", "#0089FF", "#0061B5"]))
  .colorDomain([0,450])
  .colorCalculator(function (d) { return d ? carteN.colors()(d) : '#ddd'; })
  .projection(d3.geo.mercator()
    .center([-6,52])
    .translate([0,0])
    .scale(800))
  .overlayGeoJson(dptJson.features, "state", function (d) {
    return d.properties.CODE_DEPT;
  })
  .title(function (d) {return "N°dpt :" + d.key + "\nNb de Naissances :" +
    numberFormat(d.value ? d.value : 0);})
  .render();
});
```



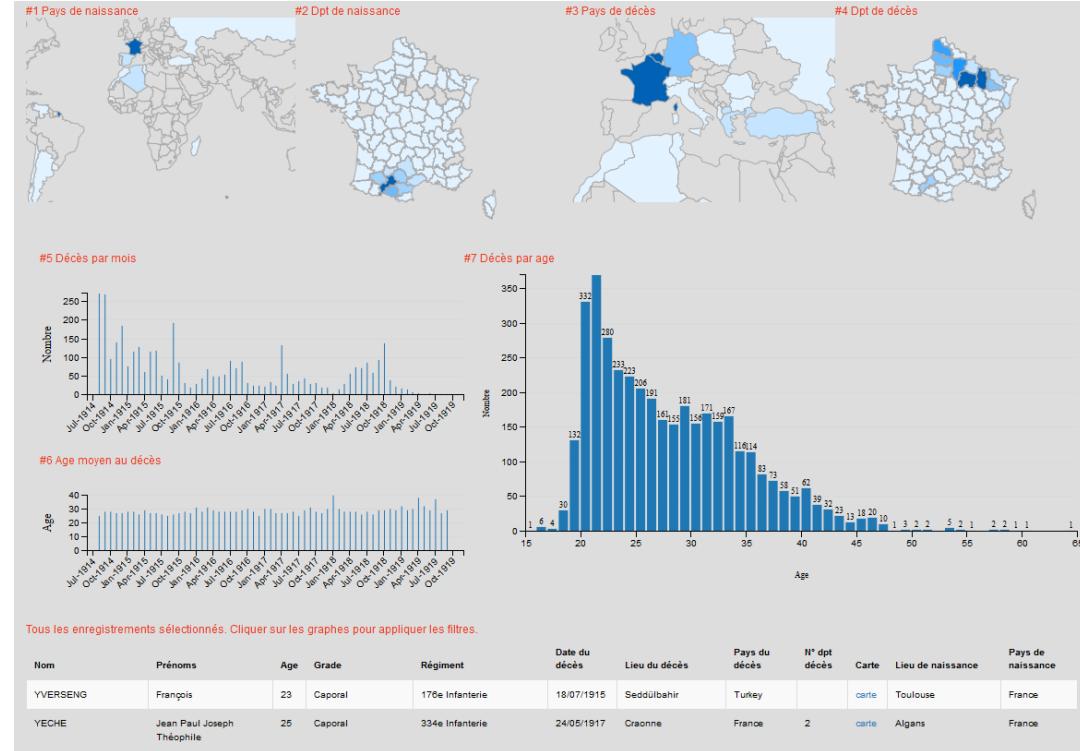
```
.title(function (d) { return dtgFormat4(d.key) + '\nMoyenne d\'age décès : ' + d.value.moy +'\\nNombre : '+ d.value.count })
```

# Graphiques : Age moyen au décès

```
//#6 Diagramme en ligne age moyen des décès
moyChart.width(510).height(150).margins({top: 30, right: 50, bottom: 60, left: 40})
  .dimension(anneeDecesM)
  .group(anneeDecesMGroup)
  .valueAccessor(function (d) {
    return d.value.moy })
  .y(d3.scale.linear().domain([0, 40]))
  .x(d3.time.scale().domain([new Date(dtgFormat3.parse("1914/06")),
    new Date(dtgFormat3.parse("1919/12"))]))
  .renderLabel(true)
  .yAxisLabel("Age")
  .title(function (d) { return dtgFormat4(d.key) + '\nMoyenne d\'age décès
    : ' + d.value.moy +'\\nNombre : '+ d.value.count })
  .renderHorizontalGridLines(true);
moyChart.yAxis().ticks(5);
moyChart.xAxis().ticks(25).tickFormat(d3.time.format("%b-%Y"));
moyChart.renderlet(function (moyChart) {
moyChart.selectAll('g.x text')
  .attr('transform', 'translate(-22,15) rotate(315)'); })
moyChart.render();
```



# Explorons !



Exploration interactive de données avec les librairies d3.js , dc.js et crossfilter.js : une bonne solution ?

# Synthèse

| outil            | +   | -  |
|------------------|---|--|
| Excel            | Simplicité d'utilisation<br>Large diffusion   | Payant (office 280€)<br>Interactivité  |
| Tableau software | Richesse des fonctionnalités<br>Pas ou peu de codage<br>Simplicité d'utilisation  | Payant (\$999 - \$1999)<br>Outil peu diffusé<br>interactivité<br>Apprentissage - |
| dc.js            | Fonctionne avec un navigateur<br>Interactivité<br>Flexibilité ++<br>Open source<br>Communauté (nb de contributions sur le web)  | Apprentissage --<br>Temps de développement -                                     |
| d3.js            | Fonctionne avec un navigateur<br>Interactivité<br>Flexibilité +++<br>Open source<br>Communauté (nb de contributions sur le web) | Apprentissage ---<br>Temps de développement --                                   |

# Bibliographie

- **D3-Tips-and-Tricks** (Malcom Maclean - ebook)
- **Interactive Data Visualization For The Web** (Introduction to Designing with d3 by Scott Murray - O'REILLY)
- **Visualizing Data** (Exploring and Explaining Data with the Processing Environment by Ben Fry - - O'REILLY)
- **Visualization Analysis & design** (Tamara Munzner)
- **dc.js - Dimensional Charting Javascript Library**
- **Over 1000 D3.js Examples and Demos**