

OpenData

Rdv avec la Data

CEREMA Med > Mathieu Rajerison

02/08/2018

Accéder



Les données OpenData sont disponibles sur les portails OpenData (ce qui paraît évident).

Elles sont généralement classées par thématiques, mais dans certains cas, on peut aussi y accéder :

- par **producteur** : pour avoir toutes les données créées par le Conseil Départemental des Bouches-du-Rhône, par exemple
- par **format** : pour n'avoir que des données que l'on sera en mesure d'exploiter dans ses outils du quotidien.
- par **licence** : pour ne retenir que les données que l'on peut exploiter en toute liberté
- parfois, par **emprise géographique** : si on ne veut que les données de son territoire propre

Faire connaissance



Méta-données

Quand on clique sur un jeu de données du portail, par exemple [celui-ci sur la localisation des défibrillateurs en Bouches-du-Rhône](#), une description plus ou moins longue de la donnée s'affiche sous la forme de texte ou de tableau (voire les deux).

Ces informations sur la donnée, ou **données sur la donnée**, s'appellent les *méta-données*.

Elles permettent entre autres de savoir **qui** a produit la donnée, **quand** et **comment**.

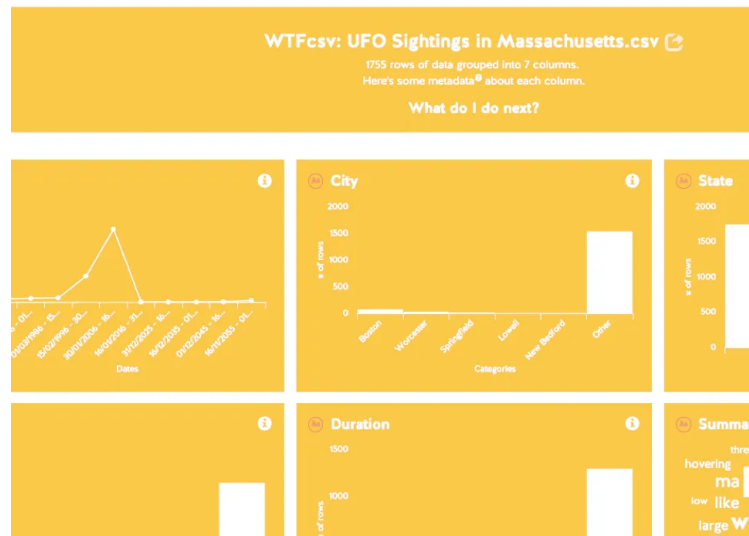
Ces informations sont indispensables afin de **fiabiliser** la source d'une donnée, connaître sa fraîcheur, connaître la licence affectée à la donnée, pouvoir **contacter** le producteur le cas échéant.

 [Lire l'article "Questions à se poser la première fois qu'on consulte une donnée"](#)

Des outils simples et rapides pour faire connaissance

Un certain nombre d'outils permettent d'explorer les données afin d'en évaluer la qualité, et repérer d'éventuelles coquilles, parmi lesquels :

- Excel ou LibreOffice Calc
- **WTFCSV**
- **CSVLint**



Synthétiser la donnée



Agréger

L'agrégation de données consiste à réaliser des **opérations d'englobage** telles que le comptage, la moyenne, le minimum, le maximum, somme à des niveaux d'organisation hiérarchique supérieurs.

L'agrégation des données simplifie l'information en la rendant **moins granulaire** et dans certains cas, permet d'**anonymiser** la donnée individuelle.

“

Par exemple, je pourrais vouloir le nombre d'habitants par département à partir de celui par commune ou le revenu médian par commune à partir de celui par quartier.

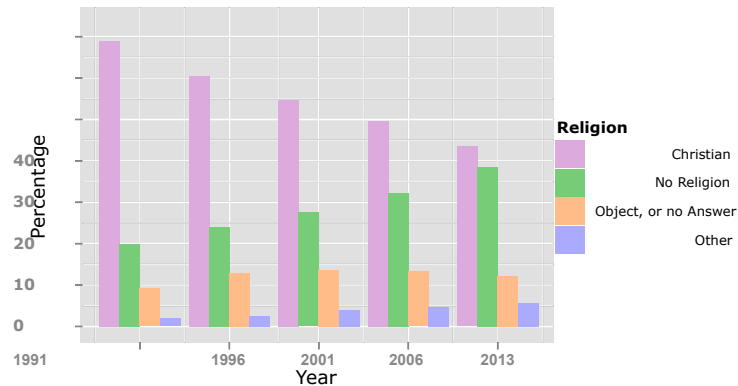
Visualiser

Infographie

Les données statistiques sont naturellement difficilement interprétables quand elles sont présentées de façon brute sous forme de tableau.

Qui a déjà vu de grands tableaux dans la partie Économie du journal "Le Monde", par exemple ?

C'est pourquoi il est d'usage de les retranscrire de façon visuelle sous forme de graphiques. Cette pratique s'appelle l'**infographie** ou **Data-visualisation**.

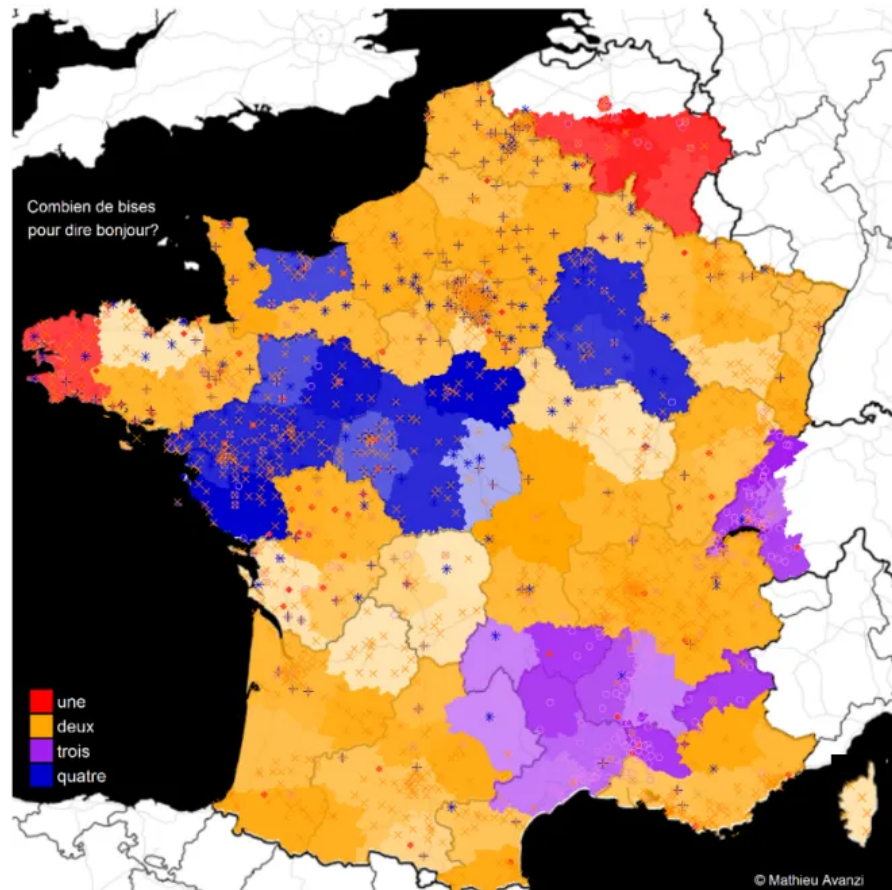


Tracer des **courbes à partir de séries temporelles**, représenter des **camemberts** ou des diagrammes en **barres** permet très rapidement de repérer des motifs intéressants dans les données.

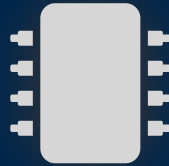
Selon le type de données rencontré, on n'utilisera pas forcément le même type de graphiques (voir l'excellent site [Data To Viz](#)) et il existe des façons de représenter la donnée très originales (voir [xeno.graphics](#)).

Cartographie

Aussi, quand cela est possible, la cartographie permet de voir des **dynamiques spatiales** à l'oeuvre et de **croiser des informations territorialisées** par le jeu des superpositions de couches d'informations



Calculer, détecter et prédire



Stats

Les statistiques correspondent à l'emploi de procédures mathématiques appliquées aux données, afin de **révéler des motifs** difficilement visibles "à l'oeil nu".

Machine learning

Le machine learning correspond à l'emploi d'algorithmes qui permettent de **prédire** des valeurs (par exemple dans le futur), voire de **regrouper** des données entre elles sur la base de leur ressemblance.

Deep learning

L'apprentissage profond est une sous-branche du machine learning, qui sous-tend l'**intelligence artificielle**, se base sur des **réseaux de neurones** qui miment le fonctionnement du cerveau humain.

Il consiste à créer un réseau capable de prédire une valeur après avoir appris depuis une multitude d'expériences et d'échantillons de données.

Métiers de la data

(La data, c'est leur dada)



Une société submergée par des masses d'informations, pour certaines à haute valeur ajoutée, a besoin de personnes sachant s'y retrouver et jongler avec ces dernières.

Des métiers commencent à apparaître, tournés autour de l'exploitation et la valorisation des données

Data scientist

Le Data scientist est un profil à la mode qui correspond à des personnes en mesure d'**accéder**, **traiter** et **manipuler** n'importe quel type de données et d'en révéler certains motifs par le biais de la visualisation ou de la statistique

CDO ou Chief Data Officer

On connaissait les CEO (Chief Economy Officer) ou CTO (Chief Technical Officer). Les CDO (pour Chief Data Officer) sont chargés de la **politique** d'ouverture des données d'une institution.

Le rôle du CDO est généralement d'**animer** un réseau de correspondants données, de recenser l'ensemble des données et d'en assurer la diffusion.

Au MTES, ce rôle est assuré par la Superviseure Générale des Données.

Chief Privacy Officer

Le Chief Privacy Officer, ou chef de la protection des données personnelles, doit **garantir la mise en oeuvre du RGPD**, soit que les données mises à disposition ne comportent pas d'informations à caractère sensible et personnel.

Communautés



Evènements

Les hackathons, data camps

Les data camps sont des évènements où des personnes intéressées par la Data **échantent** autour des données, que ce soit au niveau des **méthodes**, de **jeux de données** qui sont intéressants ou dont il serait utile de disposer.

Les hackathons, à dimension plus **technique**, sont des évènements où des personnes, geeks ou non, se réunissent sur une période assez courte (un ou deux jours) afin de répondre à certaines questions par l'exploration de données et/ou la création de prototypes applicatifs (POC ou Preuves de Concept)

Les Carto-parties

Les carto-parties sont des évènements organisés par des contributeurs OpenStreetMap visant à associer le plus souvent des personnes d'espaces périphériques (banlieues, ruraux) à la cartographie de leur commune sur OSM.

Les lieux

Data labs, infolabs

Les data labs ou info labs sont comparables aux fab labs dans la philosophie sauf qu'on n'y apprend pas à manier des imprimantes laser mais qu'on y échange autour de données ou outils relatifs à celles-ci (extraction, exploration, datavisualisation,...)

Les forums

Le forum [teamopendata](#) permet d'échanger virtuellement autour de l'OpenData.

On y apprend beaucoup de choses concernant l'actualité liée à l'ouverture des données : évènements, tournants législatifs, revue du web.

C'est le "lieu" idéal où poser ses questions sur l'OpenData.