

OpenData (Exos)

Outils pour explorer et visualiser la Data

CEREMA Med > Mathieu Rajerison

06/02/2019

Ce qui fait d'un outil un bon outil

Un bon outil de traitement est capable de...

- à plein de formats de données, soit que l'on ne soit pas bloqués dès le départ.

Un bon outil de traitement est capable de...

- à plein de formats de données, soit que l'on ne soit pas bloqués dès le départ.
- de façon souple la des données

Un bon outil de traitement est capable de...

- à plein de formats de données, soit que l'on ne soit pas bloqués dès le départ.
- de façon souple la des données
- appliquer des avancés sur ces données

Un bon outil de traitement est capable de...

- à plein de formats de données, soit que l'on ne soit pas bloqués dès le départ.
- de façon souple la des données
- appliquer des avancés sur ces données
- les résultats sous plein de formats différents

Un bon outil de visualisation est capable de...

- de manière fine les graphiques : couleur, polices, position des légendes

Un bon outil de visualisation est capable de...

- de manière fine les graphiques : couleur, polices, position des légendes
- proposer des qui ne limitent pas la créativité de l'utilisateur.

Un bon outil de visualisation est capable de...

- de manière fine les graphiques : couleur, polices, position des légendes
- proposer des interfaces qui ne limitent pas la créativité de l'utilisateur.

Certains outils sont à la fois de bons outils de traitement et de visualisation.

En général, les outils de programmation combinent ces deux aspects, mais pour les maîtriser, il faut se lancer dans la ligne de code.

Tableurs

LibreOffice Calc



Tout le monde connaît Excel, mais au Ministère et dans la fonction publique, c'est son pendant OpenSource LibreOffice qui est utilisé.

En particulier, LibreOffice Calc permet de créer des tableaux statistiques et des graphiques.

LibreOffice découle de la suite OpenOffice.

Outils en ligne

Il existe un certain nombre d'outils en ligne qui permettent de "savoir ce qui se passe dans son CSV", dont voici un extrait:

- [WTFCSV](#)
- [Dive](#)
- [CSVLint](#)

[WTFCSV](#) permet de représenter simplement et de façon immédiate le contenu de ses données. [Dive](#), dans le même style, créé par le MIT, paraît beaucoup plus avancé, et devrait vous permettre de vous plonger avec facilité dans vos données. L'outil [CSVLint](#), lui, est plus ciblé contrôle Qualité des données. On l'utilisera pour savoir si une donnée est bien structurée, par exemple, avant de la transmettre à autrui.

Tad

[Tadviewer](#) est tout simplement un visualiseur de données CSV. Il permet facilement de construire des requêtes, d'explorer la donnée par de multiples facettes, de façon simple.

Graphiques

Tableau Public



Tableau Public est la variante gratuite de l'outil Tableau Desktop.

Ce dernier permet d'intégrer des fichiers de données et de réaliser sur cette base de nombreux types de graphiques de qualité, ainsi que des cartographies.

Morph

Morph est outil de dataviz en ligne qui tente de conjuguer Art et Data. Il permet de travailler sur la représentation de ses jeux de données propres ou des jeux de tests afin de créer des datavisualisations de formes variées.

Plot.ly



Plot.ly est un outil en ligne qui, après la création d'un compte, permet de créer des graphiques à partir de fichiers importés dans l'interface

Cartographie

QGIS



QGIS, logiciel opensource, est très réputé Dans le domaine de cartographie, autant topographique que statistique.

On trouve sur internet beaucoup de tutoriaux. Il rassemble une communauté d'utilisateurs très large sur laquelle s'appuyer.

Outils institutionnels

Des institutions ont développé des outils en ligne afin de produire des cartes statistiques

- **Khartis** créé par Sciences Po
- **Magrit** créé par le CNRS

kepler

kepler.gl est un outil en ligne opensource créé par la division DataViz de la société Uber et la staryup MapBox. Il permet de créer des cartographies statistiques assez sophistiquées et modernes dans l'esthétisme, notamment des rendus en 3D.

kepler

[kepler.gl](#) est un outil en ligne opensource créé par la division DataViz de la société Uber et la start-up MapBox. Il permet de créer des cartographies statistiques assez sophistiquées et modernes dans l'esthétisme, notamment des rendus en 3D.

Carto(DB)



[Carto\(DB\)](#) est un outil SIG en ligne qui permet de créer des cartes et de les partager.

ETL

Il existe une catégorie d'outils appelés ETL qui permettent de réaliser de façon visuelle et ludique des chaînes de traitement, de transformation, et de chargement de la donnée.

Parmi ces outils, on peut citer, en opensource, Talend ou Pentaho



Programmation

R



R est outil libre de programmation en statistiques qui permet de traiter et d'exploiter de la donnée dans beaucoup de contextes et de disciplines. Il est très utilisé dans le monde académique.

En particulier, les librairies issues de la suite `tidyverse` permettent d'extraire et de transformer de la donnée, et la librairie `ggplot` (pour `grammar of graphics`) permet de réaliser de beaux graphiques.

Python



Python est sans doute le langage de programmation le plus [populaire](https://www.lemonde.fr/pixels/article/2018/07/25/je-n-imaginais-pas-que-python-connaissait-un-tel-succes_5335917_4408996.html).

Il permet de traiter de la donnée à la façon de R grâce à pandas, de réaliser des calculs avec numpy et des graphiques avec matplotlib.

C'est sous python que sont développés les librairies principales de deep learning comme tensorflow (développé et maintenu par Google) ou keras

Ressources sur la data

Datavisualisation

Data-to-Viz

data-to-viz permet de savoir quelle représentation choisir en fonction du type de données que l'on a.

Datavisualisation

Data-to-Viz

[data-to-viz](#) permet de savoir quelle représentation choisir en fonction du type de données que l'on a.

The visualization Universe

[The visualization Universe](#), par Google, permet de s'y retrouver dans la jungle infographique

Blogs sur la data

FlowingData

FlowingData est un blog très intéressant qui parle de la data, surtout sous l'angle infographique. Idéal pour être courant de l'actualité et trouver de l'inspiration quand on a fait le tour des camemberts et graphiques en barres.

Blogs sur la data

FlowingData

FlowingData est un blog très intéressant qui parle de la data, surtout sous l'angle infographique. Idéal pour être courant de l'actualité et trouver de l'inspiration quand on a fait le tour des camemberts et graphiques en barres.

Matamix

Sur le **blog** de **matamix**, on trouve pas mal d'analyses de données, de cartes, et de graphiques inspirés du quotidien.