Day 3 & 4 : Datathon

Coding Week Sciences Po

Timothée Gidoin - Nicolas Vogtenberger

January 2019

Before we start : reminder

Find those slides online : https://github.com/Gidoin/Codingweek

Notepad beginners coding session : https://hackmd.io/Fg2Np-hSTZuUuCD1-okgjQ?both

Notepad advanced coding session : https://hackmd.io/N13Qq0-gQyGUeeuf3QNeiQ?both

Joel Gombin email adress // Sylvain Lapoix email adress

Before we start : feedbacks

Did you enjoy the first 2 days ?

Before we start : feedbacks

What could be improved ?

Before we start : feedbacks

What could be improved ?

Apart from getting schwifty...

Before we start : who we are

Timothée Gidoin

Nicolas Vogtenberger

Let's get started

Datathon organisation

3 goals :

1) Make you work, by team, on real public policies issues **through a data point of view**

Datathon organisation

3 goals :

1) Make you work, by team, on real public policies issues **through a data point of view**

2) **Design a possible solution** to the issue you analysed

Datathon organisation

3 goals :

1) Make you work, by team, on real public policies issues **through a data point of view**

2) **Design a possible solution** to the issue you analysed

3) **Share your results** both online (Github & Medium) and offline

Datathon organisation

3 goals :

1) Make you work, by team, on real public policies issues **through a data point of view**

2) **Design a possible solution** to the issue you analysed

3) **Share your results** both online (Github & Medium) and offline

Datathon organisation

Day 1 : data analysis

- 9h : Let's get started *(30 minutes)*
- 9h30 : Introduction to open data *(30 minutes)*
- 10h : Launch of the datathon : unveiling the topics & constitution of the groups *(30 minutes)*
- 10h30 : Looking for sources & datasets (*2 hours*)
- 12h30 : 1 hour Break
- 13h30 : Brainstorming & Data analysing (*2 hours*)
- 15h30 : Graphical data representation (*2 hours*)

17h30 - 20h : *Optional further session for volunteers*

Datathon organisation

Day 2 : designing a solution

- **9h15** : Introduction to design thinking / workshop *(2,75 hours)*
- 11h : Sprint to formalize both your analysis and solution (*4,5 hours* including 1 hour break)
    - Prepare few slides,
    - Write a Medium article,
    - Tidy and put your code on Codingweek repo Github
- 15h30 : Official restitution in front of the jury (5 minutes per project + 5 minutes Q/A) (*2 hours*)
- 16h45 : Grand Jury decision and prize giving *(15 minutes)*

Introduction to open data

Open data principles

1. Completeness

Datasets released by the government should be as complete as possible, reflecting the entirety of what is recorded about a particular subject.

All raw information from a dataset should be released to the public, except to private information and information that may be sensitive for national safety

Open data principles

2. Primacy / Raw data

Datasets released by the government should be primary source data

Open data principles

3. Timely data

Datasets released by the government should be available to the public as soon as possible

Open data principles

4. Ease of Physical and Electronic Access

Datasets released by the government should be as accessible as possible, with accessibility defined as the ease with which information can be obtained, whether through physical or electronic means

Open data principles

5. Machine readability

Machines can handle certain kinds of inputs much better than others. Information shared in the widely-used PDF format, for example, is very difficult for machines to parse

Thus, information should be stored in widely-used file formats that easily lend themselves to machine processing.

Open data principles

6. Non-discriminatory access to data

“Non-discrimination” refers to who can access data and how they must do so

non-discriminatory access to data means that any person can access the data at any time without having to identify him/herself or provide any justification for doing so.

Open data principles

7. Open standards

Open standards refer to who owns the format in which data is stored

Open data principles

7. Open standards

Open standards refer to who owns the format in which data is stored

Microsoft Excel is a fairly commonly-used spreadsheet program which costs money to use. Freely available alternative formats often exist by which stored data can be accessed without the need for a software license

Open data principles

8. Open Licence

Maximal openness includes clearly labeling public information as a work of the government and available without restrictions on use as part of the public domain

In France two type of licences : Licence Ouverte (CC-BY) ou ODBL (CC-BY-SA). Do you know the difference ?

Open data principles

8. Open Licence

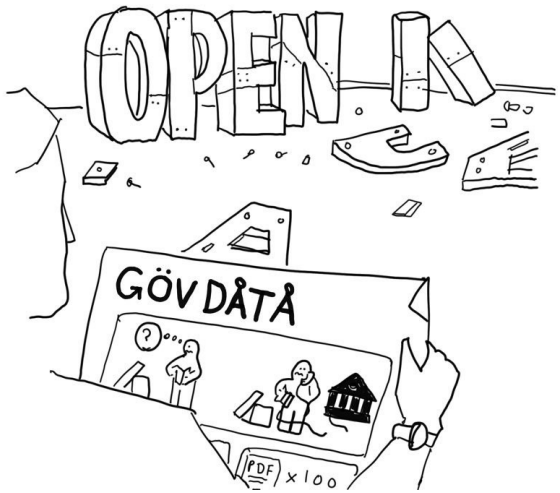
Maximal openness includes clearly labeling public information as a work of the government and available without restrictions on use as part of the public domain

In France two type of licences : Licence Ouverte (CC-BY) ou ODBL (CC-BY-SA). Do you know the difference ?

- **LO (from Etalab) / ODBL** : with both you can share, edit the database, create derived products and had a commercial usage
- **LO** : more "permissive" : you just have to mention the source and the date of update
- **ODBL** : you have to share and open your database at the same conditions

Challenge 1 : data findability

> Data findability is a major challenge. We have data portals and registries, but government agencies under one national government still publish data in different ways and different locations.(...) **Data findability is a prerequisite for open data to fulfill its potential and currently most data is very hard to find.**



https://index.okfn.org/insights/

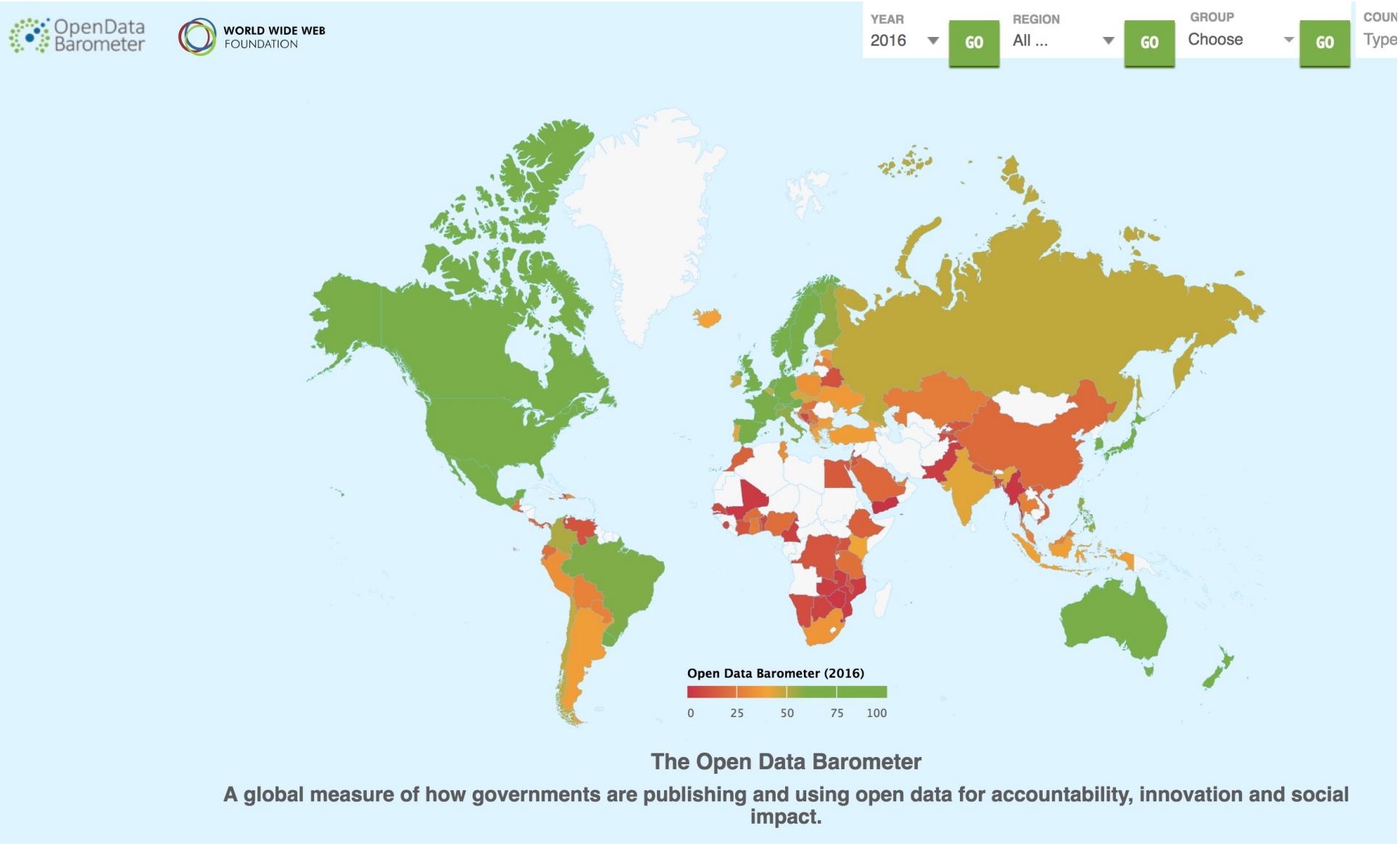
Challenge 1 : data findability

Défi 2 : data quality

> **Government data is usually incomplete, out of date, of low quality, and fragmented.** In most cases, open data catalogues or portals are manually fed as the result of informal data management approaches. **Procedures, timelines, and responsibilities are frequently unclear among government institutions tasked with this work.**

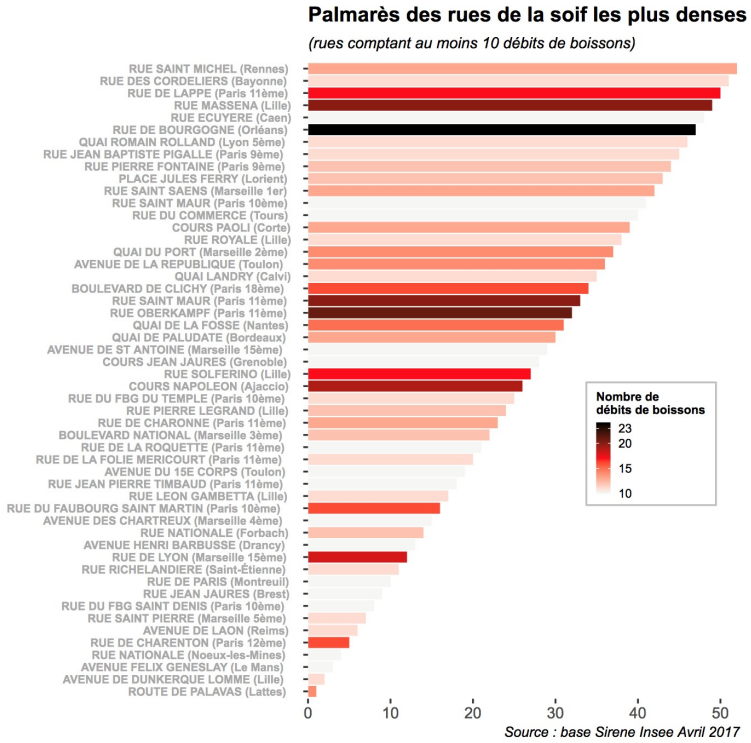
http://opendatabarometer.org/4thedition/r

Défi 2 : data quality



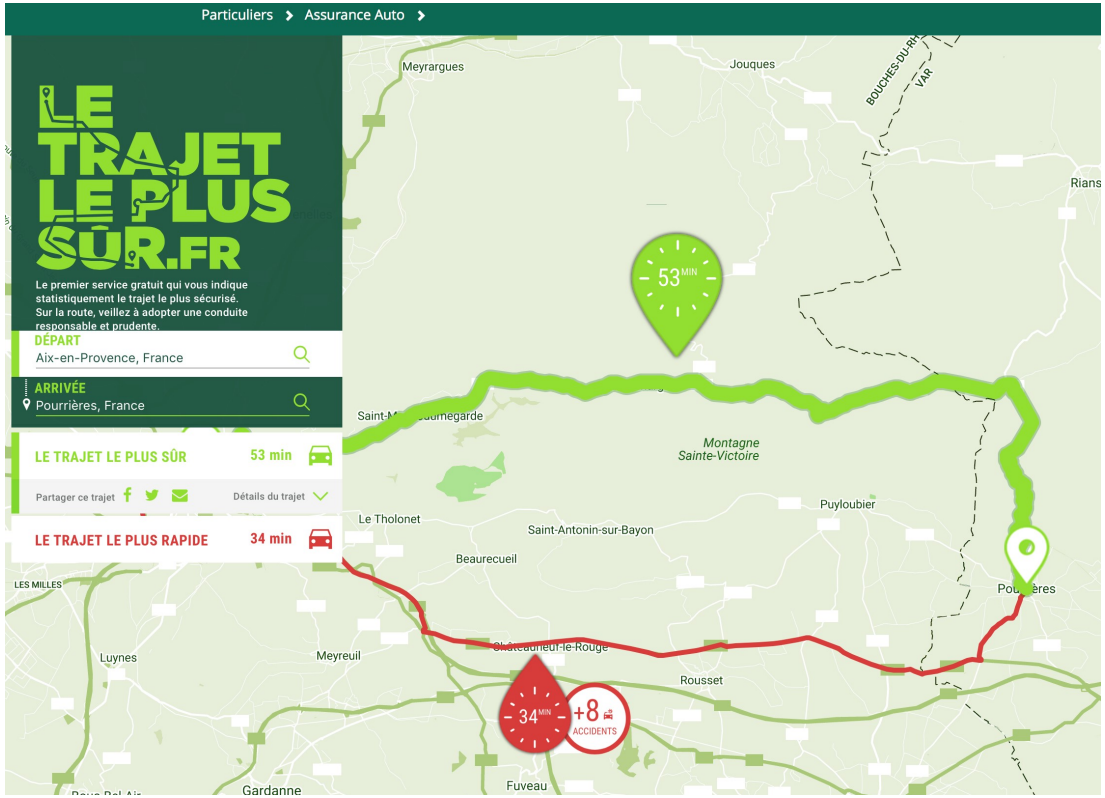
Open Data example : base SIRENE

Open Data reutilisation : base SIRENE

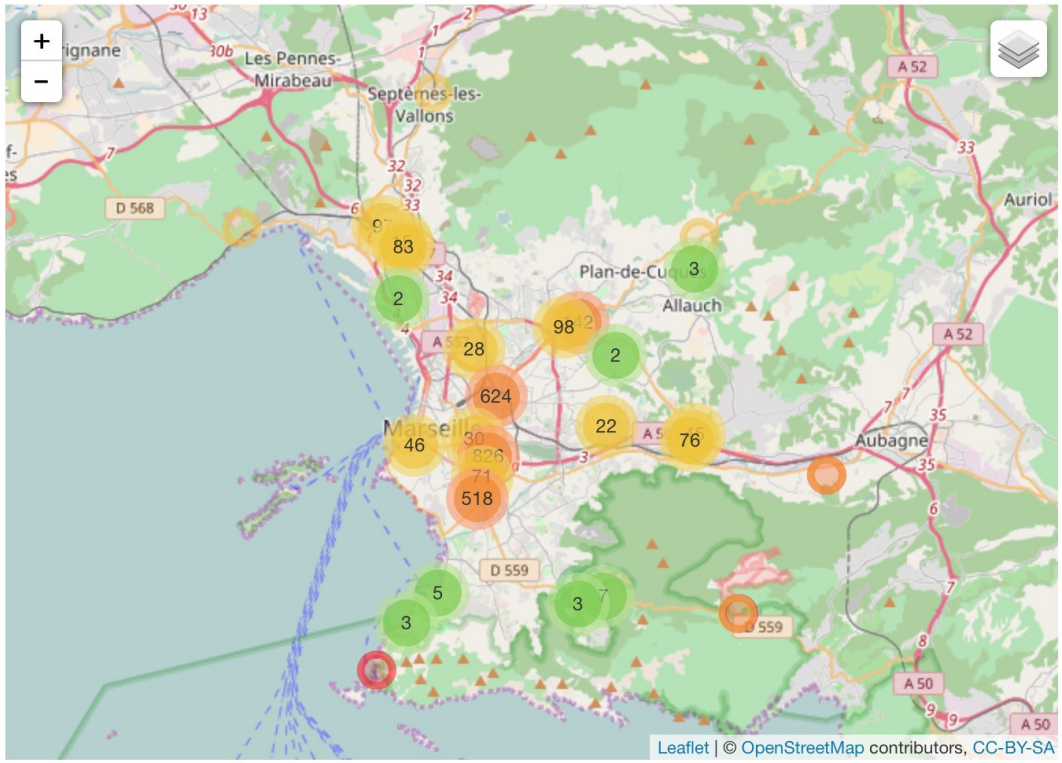


Open Data example : road accidents

Open Data reutilisation : road accidents



Open Data reutilisation : road accidents



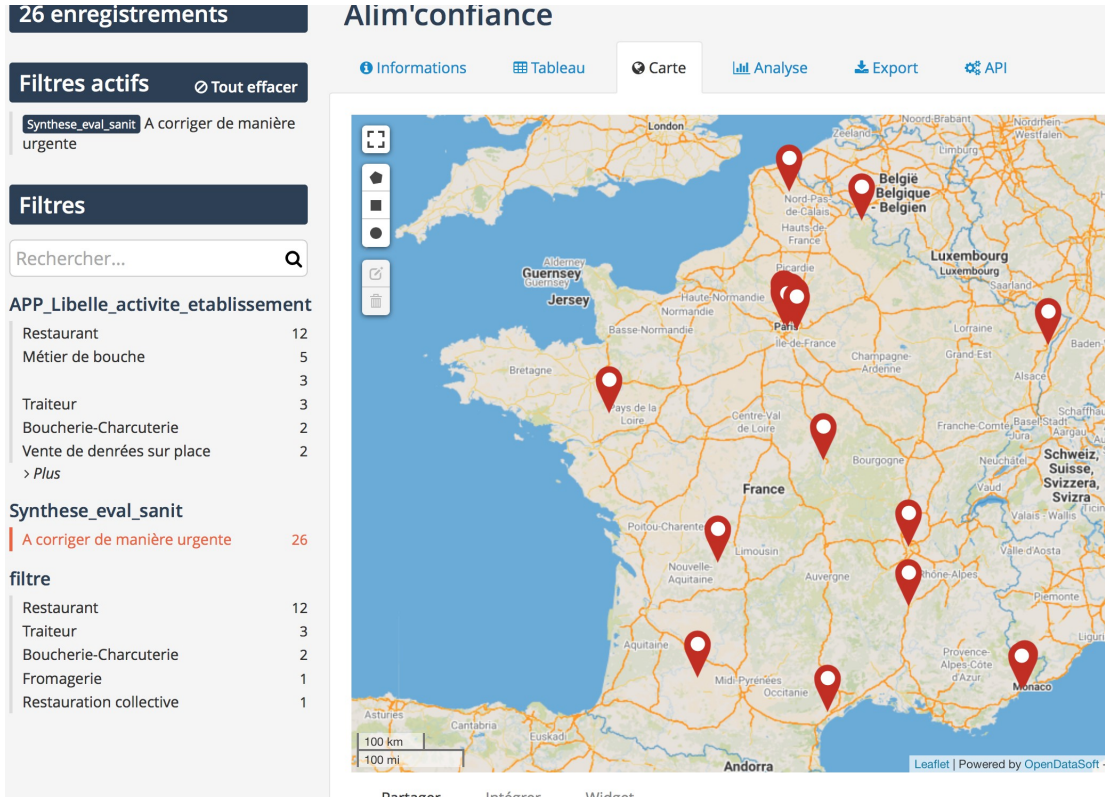
n cliquant sur le petit pictogramme de calque en haut à droite de la carte, vous pouvez

Map by Joël Gombin in Marsactu

Open Data example : sanitary inspections

Résultats des contrôles officiels sanitaires : dispositif d'information « Alim'confiance »

Open Data reutilisation : sanitary inspections



26 établissements au niveau d'hygiène à corriger de manière urgent

Open Data example : crime data

Chiffres départementaux mensuels relatifs aux crimes et délits enregistrés par les services de police et de gendarmerie depuis janvier 1996

Ce jeu de données provient d'un service public certifié

NEC MERGITUR

Ces données correspondent aux nombres de crimes et délits enregistrés mensuellement par les services de police et de gendarmerie.
Les tableaux mis à disposition forment ce que l'on appelait « l'état 4001 ».
Ils contiennent des informations, de caractère administratif, sur l'activité judiciaire des services de police et de gendarmerie, y compris celles des DOM-COM, depuis janvier 1996. Ces données sont mises à jour mensuellement.

Ressources

Documentation_des_chiffres_mensuels_departementaux.docx

docx (40.8Ko) 737 Disponible — TÉLÉCHARGER

Tableaux_4001_TS.xlsx

Le fichier Tableaux_4001_TS contient les données enregistrées par les unités de la gendarmerie et de la police de chaque département.

xlsx (11.6Mo) 855 Disponible — PRÉVISUALISER — TÉLÉCHARGER

Producteur

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE

MINISTÈRE DE L'INTÉRIEUR

Ministère de l'Intérieur

Placé au cœur de l'État, le ministère de l'Intérieur assure la permanence et la continuité de l'État. Cette fonction régalienne se concrétise par le rôle majeur et les services…

VOIR LE PROFIL

CONTACTER

Launch of the datathon

Topics & groups

Here are the 6 topics

- **Do the French students cost too much ?**
- **Do the French public servants cost too much ?***Grand débat National*
- **So as to reduce the public debt and taxes, which public spendings shall be cut off in priority ?** *Grand débat National*
- **Are there too many representatives in France ?**
- **Are French representatives paid too much ?**
- **Are French public services too concentrated in the cities ?***Grand débat National*

You have 20 minutes to form a group of 5-6 students with at least 2 "advanced" students

Thank you

Contact : timothee@datactivi.st