Filtrage interactif de données multi-dimensionnelles

Patrick Brockmann
LSCE (UMR 8212 / CEA-CNRS-UVSQ)

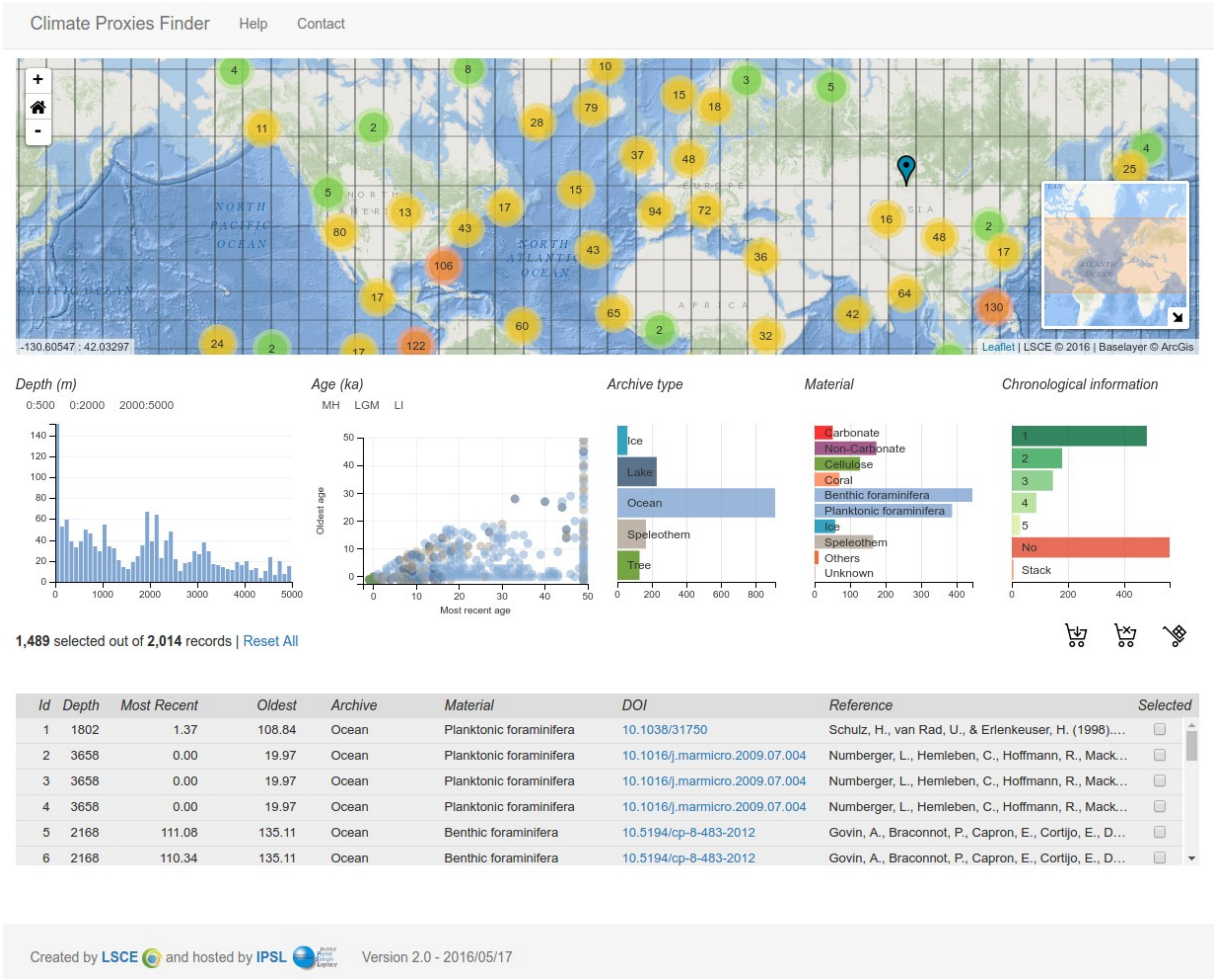
SIST 2016 : Séries Interopérables et Systèmes de Traitement
29-30 sept. 2016 Montpellier (France)

Un partage d'expérience sur la réalisation d'une application d'exploration de données

Contexte
- Données paléoclimatiques
- Renfort sur un projet en cours

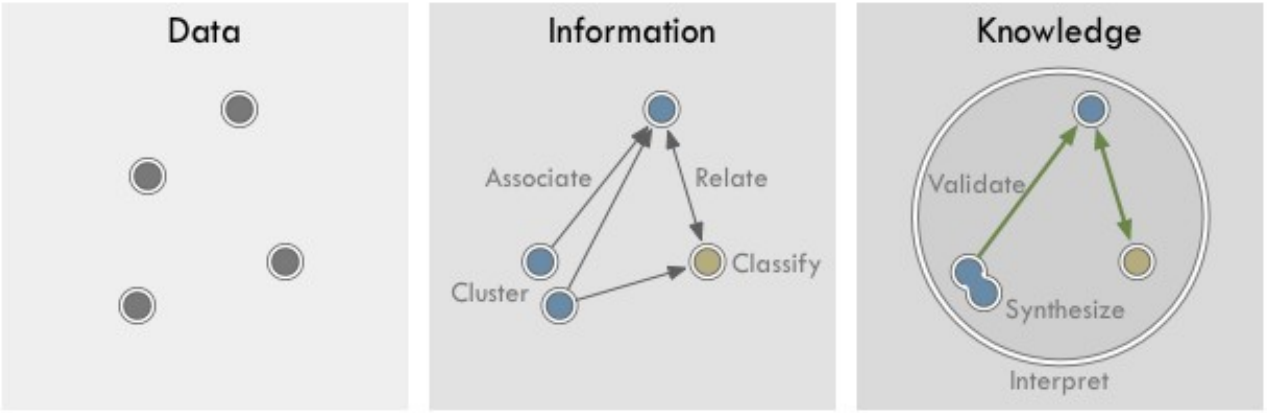
Problèmatiques
- Vérification des traitements
- Analyse exploratoire
- Diffusion des résultats

Ressources
- 1 personne sur quelques semaines



Motivations

Accompagner la Recherche sur la transformation des Données en Informations et en Connaissances.



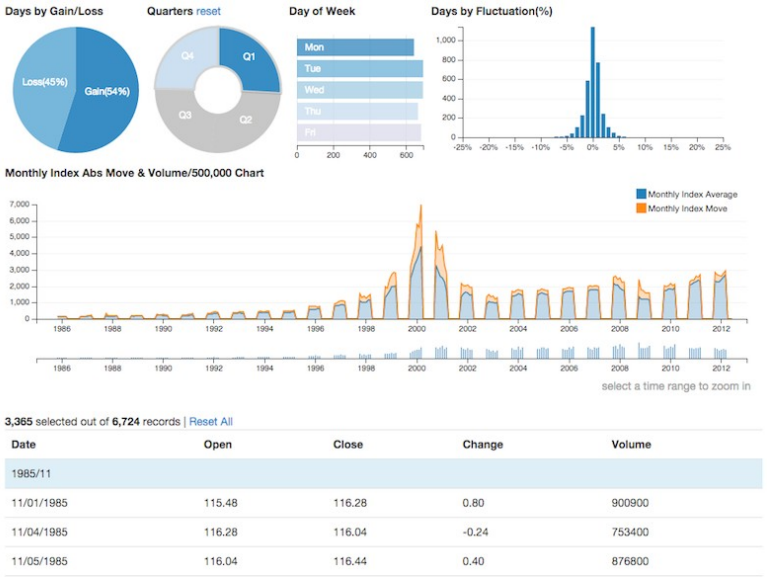
Comment ?

En concevant des applications avec des tableaux de bord interactifs

Pourquoi ?

Parce que organiser, filtrer les différentes propriétes des données (leurs *dimensions*), puis les grouper, les associer, les agréger permet de

- Trouver des relations entre les données
- Vérifier les données en les visualisant
- Interpréter l'information produite
- Etablir des connaissances

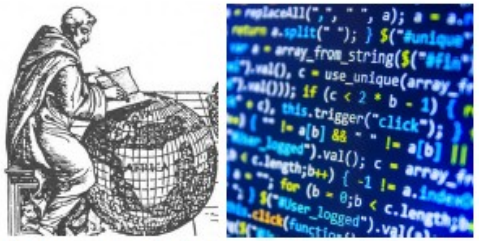


LSCE Data Visualization Group

Climate data analysis produces large-scale and multi-attribute data. How can these results be explored? What can be done to better explore these data? What are the essential findings? What is the best tecnhical way to communicate them ? These are critical questions that interactive, web-based data visualization can help answer.

Products

We use web technology to create interactive graphics that are displayed in the browser. This is especially useful to see the effect of changing the value of an attribute on a variable of interest or provide a synthesized view of a complex dataset. The visualizations we produce can be included as figures (with online links) in journal publications, websites, blogs, or can be exploratory tools hosted as web platforms.



Modern data visualization is produced with code that uses specialized libraries (e.g. d3.js, dc.js, crossfilter.js).

Does data vis replace your analysis tools?

No. Data vis comes at the end of the pipeline, after you have analyzed your data in Python, R, or Ferret, for example.

What is the role of data visualization?

Data visualization transforms data (numerical values) into information by showing how they are related. Knowledge is produced when information is interpreted, analyzed and judged to inform decision-making or further analysis.

Design process

Visualization ideas are conceived based on your needs, the technologies at our disposal, and the target users. The process then iteratively evolves by prototyping, testing, and getting feedback.

1. Define objective and target users
2. Find the story in the data

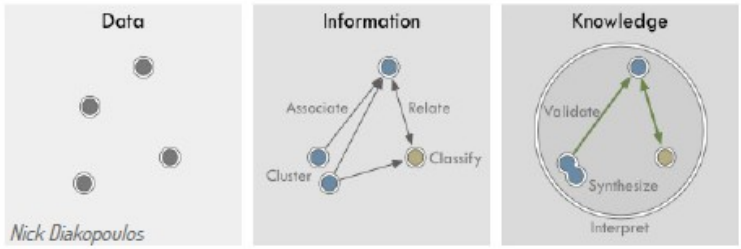
collaborative effort with scientists

3. Build web-based visualization

feedback & iteration

4. Disseminate

publications, local /public sites, blogs...



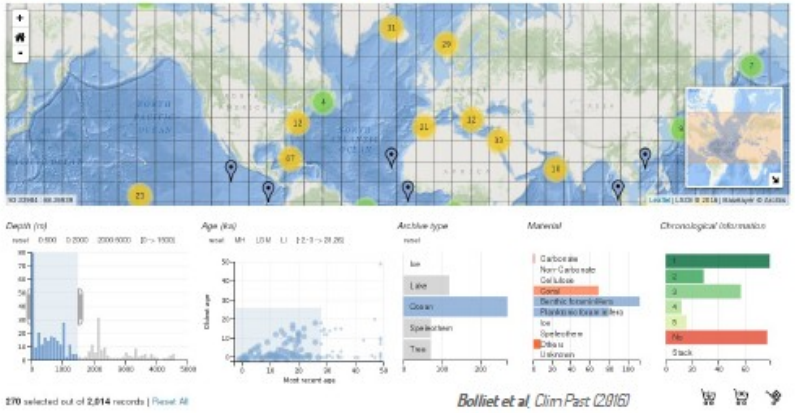
Our mission : create data visualizations for the interactive exploration, discovery and communication of data.

Our work

https://github.com/LSCE-DataVisGroup/

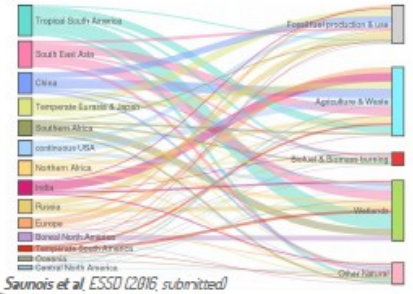
We have made exploratory tools for climate proxies, analogues of circulation, models of climate indices, and bibliographic data, as well as interactive graphics for coastal CO2 fluxes, CO2 and methane emissions.

Climate Proxies Finder – linked charts and map filterable by ocean proxy attributes



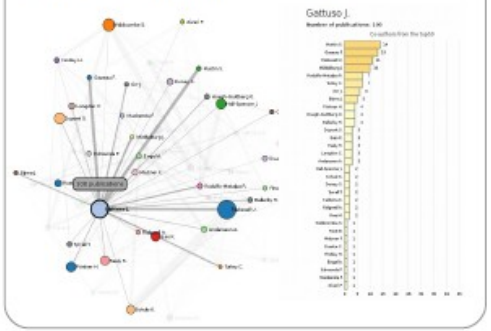
Bolliet et al, Clim Past (2016)

CH4 Emissions – Sankey diagram for different sources

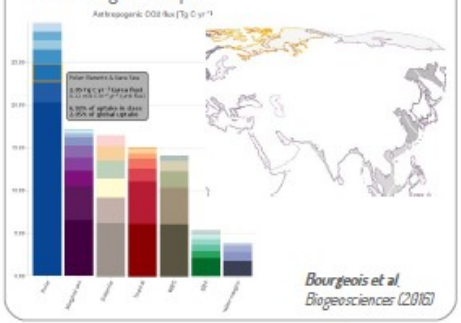


Saunois et al, ESSD (2016, submitted)

Bibliograph – searchable network graph for the ocean acidification community (OA-ICC database)

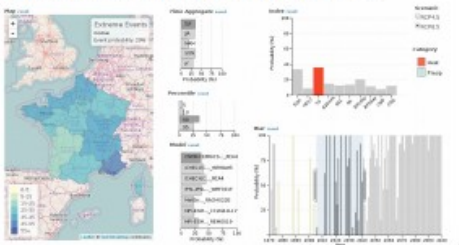


Coastal CO2 flux – linked stacked bar chart with coastal regions map



Bourgeois et al Biogeosciences (2016)

Extremoscope – linked dasboard for extreme event probabilities based on different climate indices



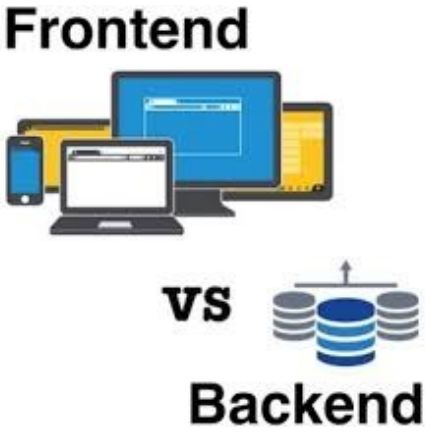
Demo time !

"Now! ... *That* should clear up a few things around here!"

http://climateproxiesfinder.ipsl.fr/screencast.mp4

Architecture
front-end versus *back-end*

Ici tout est du côté navigateur !



dc.js - Dimensional Charting Javascript Library

dc.js is a javascript charting library with native crossfilter support and allowing highly efficient exploration on large multi-dimensional dataset (inspired by crossfilter's demo). It leverages d3 engine to render charts in css friendly svg format. Charts rendered using dc.js are naturally data driven and reactive therefore providing instant feedback on user's interaction. The main objective of this project is to provide an easy yet powerful javascript library which can be utilized to perform data visualization and analysis in browser as well as on mobile device.

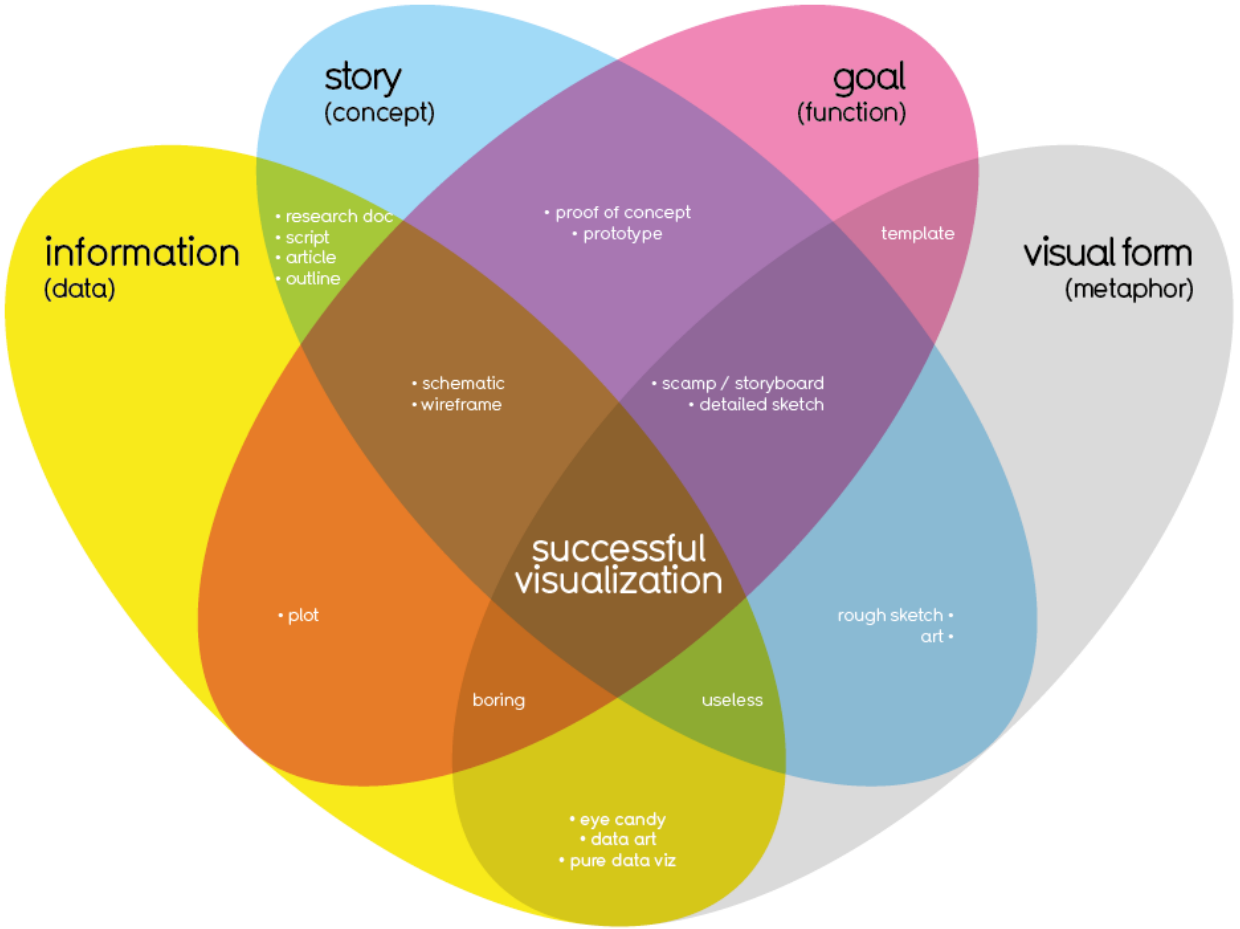
Pre-processing

Gestion de code

rollover for more detail

What Makes a Good Visualization?

explicit (implicit)



David McCandless
InformationisBeautiful.net

taken from new book
Knowledge is Beautiful

find out more
bit.ly/KIB_Books

Profils

dichotomy?

generalist x specialist

the creative developer

by @almirfilho

developer

↓

web developer

↓

front-end developer

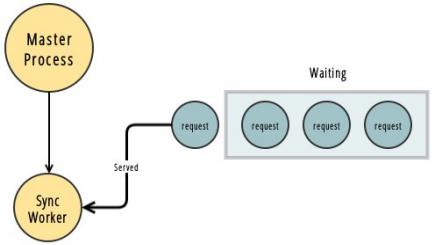
↓

ui developer

http://www.slideshare.net/almirfilh0/the-creative-developer
Almir Filho, Web Developer

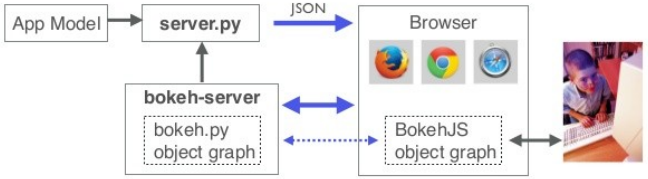
Différentes solutions pour le back-end

gunicorn



bokeh.py & bokeh.js

CONTINUUM ANALYTICS

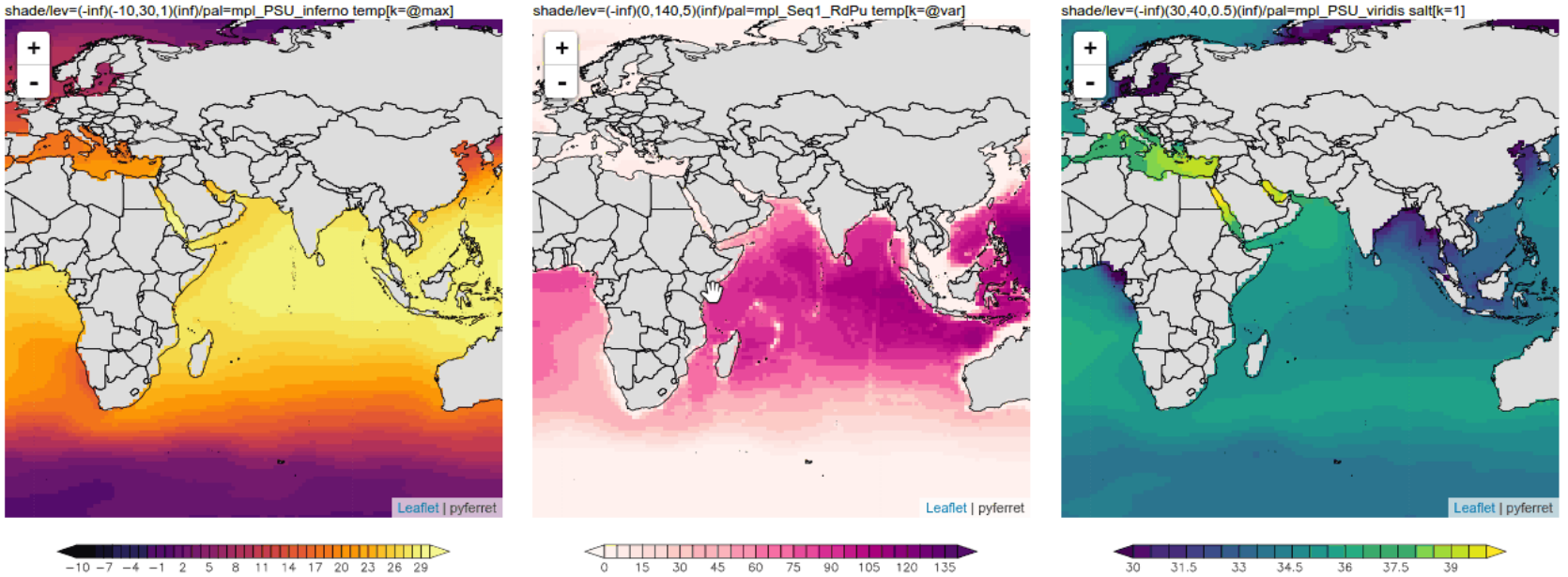


Flask
web development, one drop at a time

mongoDB

Slippy synchronous maps

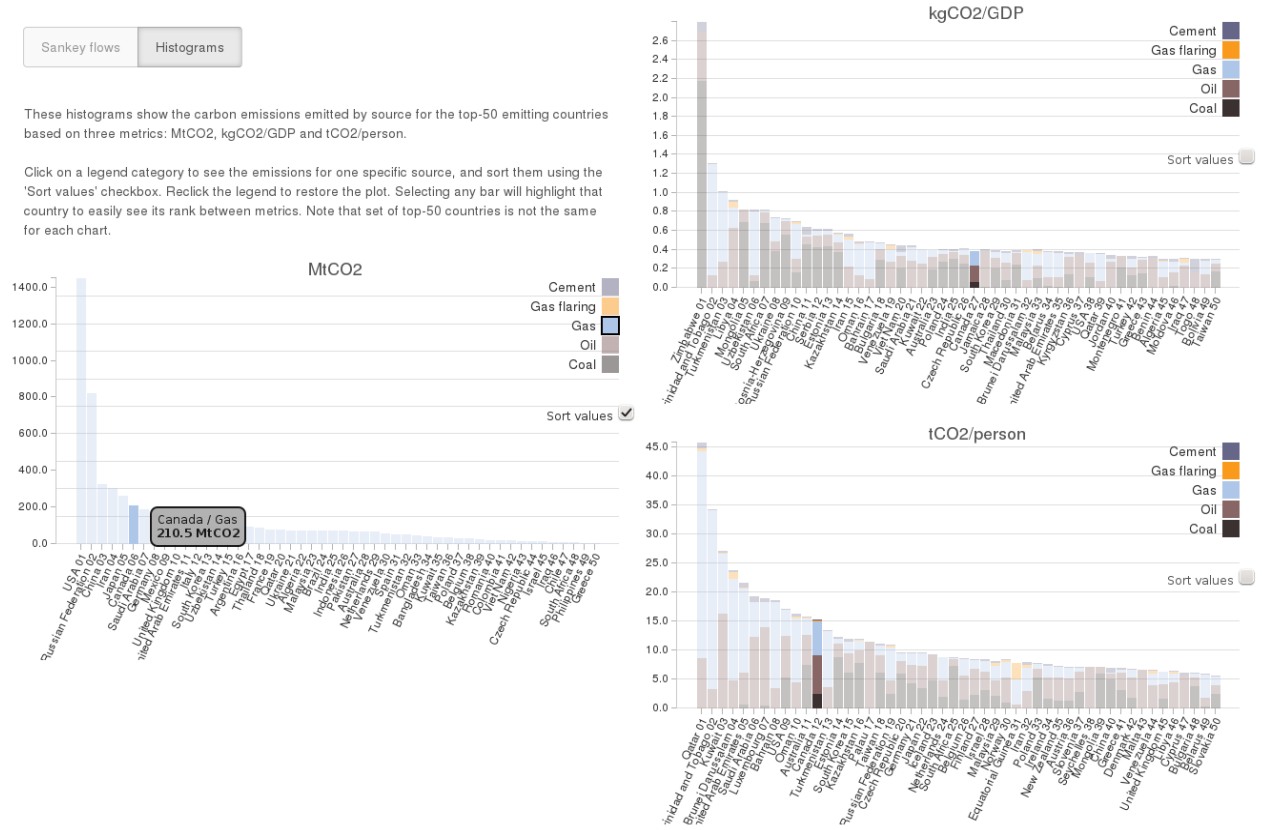
- gunicorn
- pyferret



https://github.com/PBrockmann/wms-pyferret

Data visualisation basée sur d3.js

- rendre interactif des graphiques statiques
- offrir un mode exploratoire



http://lsce-datavisgroup.github.io/CO2emissions/