The Pipeline of Data Visualization
illustrated with Swiss parliament data

Joachim Muth, Gael Lederrey & Jonas Racine

École Polytechnique Fédérale de Lausanne

Introduction

This project represent the final point of **Applied Data Analysis** cours, taught by Michele Catasta in 2016. We put into practice the analysis of real-world data, following the whole pipeline, from **scraping** to **interactive visualization**. We experiment what are the **real-world challenges** encountered in such a project. Our case study focus on the data available on `hwww.parliament.ch` which describe all publicly available activities of Swiss parliamentary deputy.

Motivation

Since the **The Freedom of Information Act** entered into force (on 1 July 2006) a huge amount of data related to political activities are freely available to the public. However, it represent a big amount of documents hardly understandable. We suggest an interface giving **light overview** of Swiss parliamentaries.

1. Data Acquisition

Parliament IT service provides an API, which allows direct access to database tables. However, it presents some restrictions:

- **Limited** returned items per request
- Too short **timeout**, forbidden the use of `skip` to access last elements of big tables.

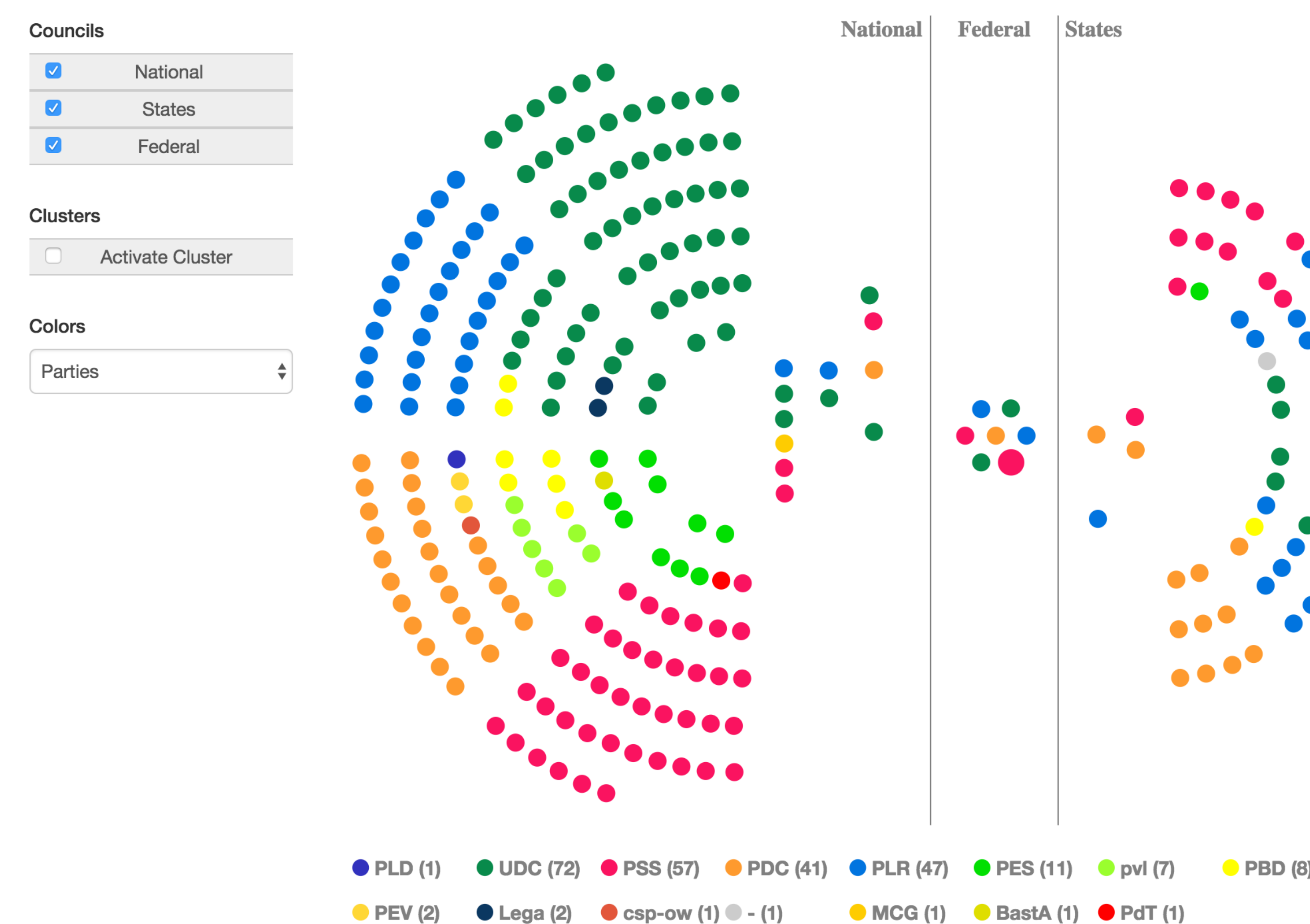We developed a python scrapper that bypass these restrictions.

2. Data Preparation & Cleaning

The database tables collected contains numerous problems:

- Hashed primary key (ID)
- Multiple entries for same item in different languages
- 3 different ID's to designate a parliamentary

3. Constraints choices

As working on data involving public peoples, we forbid ourself for any **subjective interprepretations**. Instead, we focus on a visualization tool from which other people (e.g. journalist) can draw conclusions. Hence, it needed to be transparent enough to be trusted.

4. Data Interpretation

Under reasonable assumptions:

- **Involvement:** number of speeches and initiatives
- **Interests:** distribution of wrote and signed initiatives between themes.
- **Partnerships:** cosigned initiatives.

PLD (1)   UDC (72)   PSS (57)   PDC (41)   PLR (47)   PES (11)   pvl (7)   PBD (8)
PEV (2)   Lega (2)   csp-ow (1)   - (1)   MCG (1)   BastA (1)   PdT (1)

5. Visualization

We want the visualization to be:

- **Intuitive** enough to be used without explaination.
- **Enjoyable**, so people have fun playing with it.
- **One page sized**, to lower complexity
- View per **deputy** or per **agglomerated group**.

6. Assessment from Domain Expert

Meeting with **Alain Rebetez**, political journalist, and **Philippe Nantermod**, PLR deputy.

- Expertise about Political system.
- Is the visualization intuitive enought?
- What informations would they like to see?

7. ... and next ?

Go back to point 3 and iterate until reaching the deadline :)

Acknowledgements

Conclusion

This project gives an **overview of some challenges** that could be encountered at each step of the process, when trying to analyse real-world data. It puts in evidence the interdisciplinarity of data analysis, requiring the full range of skills: **codding** to master the numerous IT tools; **analytical** to take on what to focus on; **curiosity** to understand the field where data come from; **social** to interact with domain experts and understand their point of view.

Involved IT tools

- Scrapping and data analysis with **Python**, enhanced with Pandas, NumPy and SciPy libraries.
- Visualization with **JavaScript**, enhanced with D3JS and JQuery
- Website with **Bootstrap**.

Contact Information

- github.com/jmuth/parliament-viz.ch
- www.parliament-viz.ch
- @ joachim.muth@epfl.ch
  gael.lederrey@epfl.ch
  jonas.racine@epfl.ch