

The Pipeline of Data Visualization illustrated with Swiss parliament data

Joachim Muth, Gael Lederrey & Jonas Racine

École Polytechnique Fédérale de Lausanne

Introduction

This project represents the final point of **Applied Data Analysis** course, 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**. By doing so we experienced the **real-world challenges** that arise in such a project. Our case study focuses on the data available on hwww.parliament.ch.

Motivation

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

1. Data Acquisition

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

- **Limited** number of returned items per request
- A **timeout** so short it makes it impossible to use **skip** to access last elements of big tables.

To overcome these issues we developed a Python scraper that bypasses these restrictions.

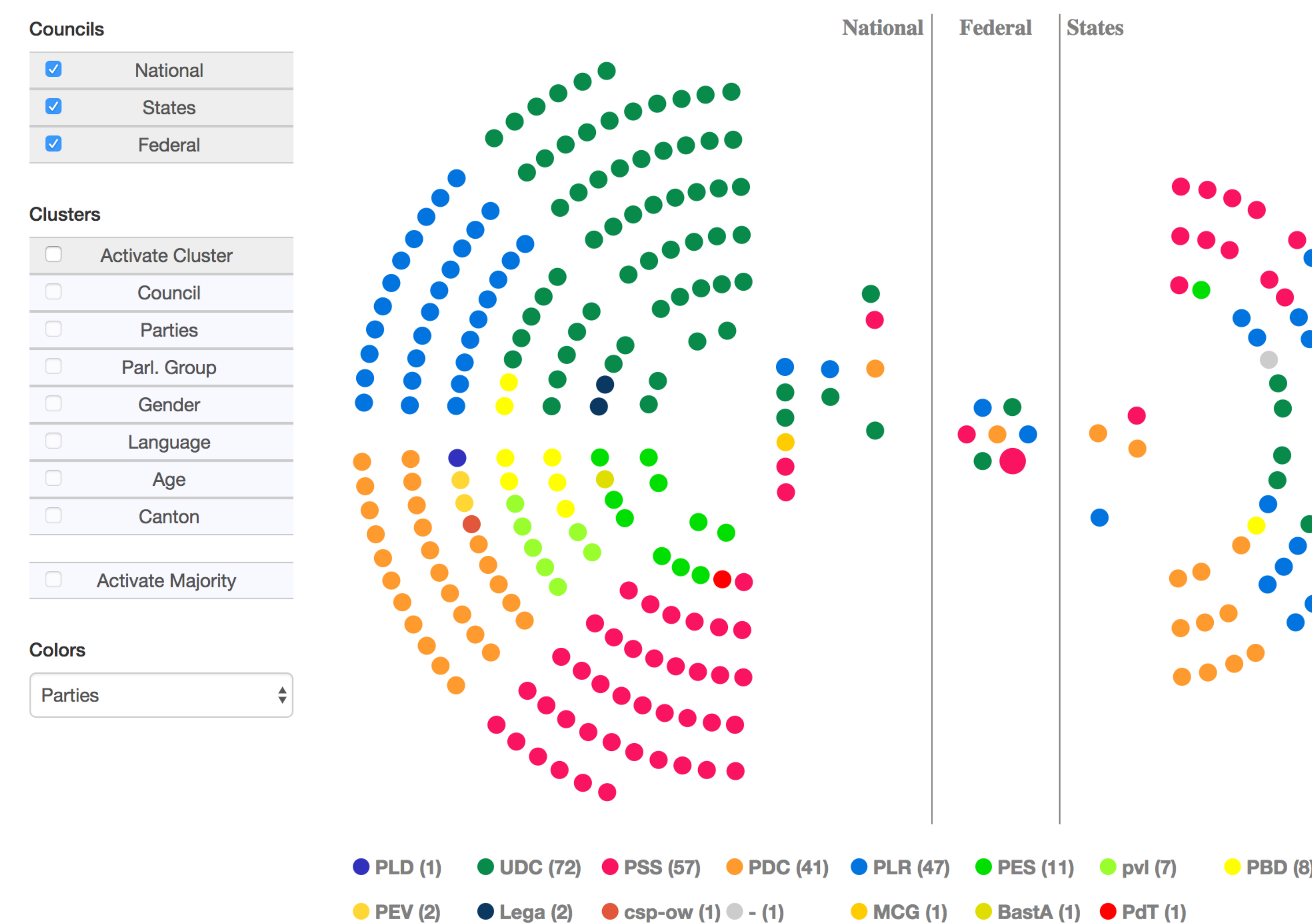
2. Data Preparation & Cleaning

The database tables collected contain numerous problems:

- Hashed primary key (ID)
- Multiple entries for same item in different languages
- 3 different IDs to designate a member of parliament

3. Constraints

As we are working on data involving public people, we refrain from any **subjective interpretation**. Instead, we focus on a visualization tool that other people (e.g. journalists) can use to draw conclusions. Hence, it needed to be transparent enough to be trusted.



Conclusion

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

4. Data Interpretation

Under reasonable assumptions:

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

Involved IT tools

- Scraping and data analysis with **Python**, enhanced with Pandas, NumPy and SciPy libraries.
- Visualization with **JavaScript**, enhanced with D3.js and JQuery
- Website design with **Bootstrap**.

5. Visualization

We want the visualization to be:

- **Intuitive** enough to be used without explanation.
- **Enjoyable**, so people have fun exploring it.
- **Single-Page**, to lower complexity
- Focused on the **individuals** or **agglomerated groups**.

6. Assessment from Domain Experts

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

- Expertise about political system.
- Is the visualization intuitive enough?
- What information would they like to see?

7. ... and next ?

Go back to point 4 and iterate until reaching the deadline :)

Acknowledgements

A warm thank you to **Alain Rebetez**, political journalist, **Philippe Nantermod**, PLR deputy, and **Mathias Reynard**, PS deputy, for their expertise in the Swiss parliamentary system, their advices and their external point of view on the project.

Contact Information

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