Data Science Project

Voting reports in Switzerland's Parliament

Conceptual Design Report

31 October 2022

Lisa Asticher, Kim Lan Vu
Lisa.asticher@kpm.unibe.ch, kmln.vu@gmail.com

Abstract

Our project aims to analyze how the members of the Swiss National Council, the first chamber of the Swiss parliament, cast their vote, with regards to their parliamentary group affiliation and personal attributes, such as gender and age. More specifically, we take a closer look at the "deviant votes", i.e. Council members casting their vote against the majority of their own parliamentary group.

In this report, we present the objectives of the project as well as our methodology during the various stages. In particular, we present our data structure and architecture and explain how we collect data and process it throughout the project. We discuss the risks and hazards associated with our data and its protection in this project, both at the conceptual and physical levels, and explain the strategies we have implemented to ensure the integrity of our data and analytical work. We further discuss preliminary studies about the voting behavior in the Swiss parliament and present some first impressions based on a subset of the data.

Table of Contents

1 Project Objectives

Switzerland's parliament has two chambers: the National Council, which represents the people, and the Council of States, which represents the cantons. The National Council is composed of 200 members and the seats are allotted to the 26 cantons on the basis of their resident population (system of proportional representation). The Council of States is composed of 46 members, each canton having two representatives while the half-cantons have one. Each Council member belongs to one of the six parliamentary groups, representing the major parties in Switzerland. Parliament holds four sessions a year – in March, June, September and December – each lasting three weeks. During those sessions, several items of business are debated and voted upon. Considering the votes during the sessions over the current legislative period (2020-22), our objective is to determine if we can find out any correlation or pattern about National Council members casting their vote against the majority of their parliamentary group, as regards to their political orientation, gender, age and the items of business voted upon. In particular, we aim to verify if this behavior of "deviant voting" is more likely to happen with Council members belonging to the "Center" group.

To achieve this, we will look at the public voting reports. These voting reports inform the public about how Council members cast their votes on each item of business in the National Council and in the Council of States. Detailed voting reports of the National Council are available on the website of the Swiss Parliament[1]. Detailed voting reports of the Council of States are not available, which is why we will focus our analysis on the National Council.

We will first analyze the political orientation, gender and age distribution of the Council members. Then, for each debated item of business, we will assess how the different parliamentary group casted their vote and if the Council members voted in accordance with their own group or not, with the aim of identifying which type of Council member is more likely to vote against their group and in which context.

2 Methods

Our main working environment is Google Colab, which provides an interactive environment (called a notebook) that allows us to combine text and executable code in the browser. We are using the Chrome browser and Python code. The data files are stored in Google Drive, with which Colab is integrated. The data pre-processing and cleaning will be done in Microsoft Excel.

To support our analysis, we will need different Python libraries and packages, namely:

---

[1] https://www.parlament.ch

- Pandas: a package providing fast, flexible, and expressive data structures that allows to work with "relational" or "labeled" data[2].
- NumPy: a library that contains multidimensional array and matrix data structures[3].
- Matplotlib: a comprehensive library for creating static, animated, and interactive visualizations[4].
- SciPy: a library that provides algorithms for optimization, integration, interpolation, algebraic equations, differential equations and statistics[5].

3 Data

We will base our analysis on the data available in the voting reports. These reports are public and available for download in xlsx format on the website of the Swiss Parliament. There is a voting report for each session.

A voting report contains:

- the vote of each Council member on every items of business (yes / no / abstention / absent)
- attributes of each Council member: ID, name, parliamentary group, canton, date of birth
- attributes of each item of business: ID, title, department

We will first need to pre-process the data from those reports and convert them into a csv dataset. For each item of business, we will evaluate how the majority of each parliamentary group voted, compare it with the individual vote and add the corresponding information in the dataset. We will also group the parliamentary groups in 3 categories: Left (Social Democrats, Green Group), Center (Center Group), Right (Green Liberal Group, Liberals Group, Swiss People's Party). To perform our analysis, we will also need to add data about the age and gender of the Councilors. In the finalized dataset, each row will correspond to the vote of one Council member on one item of business (Table 1: first rows from the May 2022 National Council session dataset)

	CouncillorId	CouncillorName	Party	Position	Canton	Region	Date_de_naissance	Age	Sexe	Voting_Object	Vote	Party_Vote	VoteAgainstParty
0	2513	Maillard Pierre-Yves	Socialists	Left	VD	swiss french	16/03/1968	54	M	01. Augmentation progressive des depenses de l...	No	No	0
1	2513	Maillard Pierre-Yves	Socialists	Left	VD	swiss french	16/03/1968	54	M	02. Transparence sur les cas de discours haine...	Yes	Yes	0
2	2513	Maillard Pierre-Yves	Socialists	Left	VD	swiss french	16/03/1968	54	M	03. Prevenir le gaspillage d'energie d? aux ap...	Yes	Yes	0
3	2513	Maillard Pierre-Yves	Socialists	Left	VD	swiss french	16/03/1968	54	M	04. Amelioration de la securite routiere gr?ce...	Yes	Yes	0
4	2513	Maillard Pierre-Yves	Socialists	Left	VD	swiss french	16/03/1968	54	M	05. Agir pour une plus grande neutralite des a...	Yes	Yes	0

Table 1: first rows from the May 2022 National Council session dataset

---

[2] https://pandas.pydata.org
[3] https://numpy.org
[4] https://matplotlib.org
[5] https://scipy.org

4 Metadata

The metadata is stored in a readme file and a codebook, both accessible in a metadata directory situated in the root directory of the project, which is located on Google Drive. All project members have read-access on the file, however only the project manager can edit it to avoid mistakes, misunderstandings and inconsistencies.

The readme file contains necessary and basic information about the project and the data, such as the authors, number of data files, time range of data collection, sharing and access information about the data (license), as well as a short data overview. It also describes the global architecture of the project directory.

The codebook file contains detailed information regarding each variable and measurement, as well as a first statistical description of the dataset. For each variable we describe the content of the variable, the data type and the coding method. For example, looking at Table 1, we describe the variables "Party", "Position" and "VoteagainstParty" in the following way:

Name	Description	Type	Coding
Party	Corresponds to the parliamentary group to which the Council member belongs. A parliamentary group is made up of members of the same party or parties sharing the same political affinities.	Categorical (6)	V: "Swiss People's Party" S: "Social Democrats" M-E: "The Center Group" RL: "FDP. The Liberal Group" G: "Green Group" GL: "Green Liberal Group"
Position	Derived from the variable "Party", it groups the parliamentary groups into three categories according to their general political orientation and voting behavior.	Categorical (3)	Left: Combination of "S" and "G" Center : "M-E" Right : Combination of "V", "RL", "GL".
VoteagainstParty	Derived from the two variables "Vote" and "Party_Vote". Indicates whether the MP voted the same as the majority of the MP that share the same position.	Dummy (1/0)	1: MP voted against majority 0: MP did not vote against majority

Table 2 Extract from the codebook

Because the voting data are systematically collected and processed by the Swiss Parliamentary Services, we have no missing values, hence the absence of indication for missing data code.

5 Data Quality

The raw data comes from the official website of the Swiss Parliament, which provides .xlsx tables with all the voting data of the National Councilors for each parliamentary session. The votes are recorded in a systematic and consistent way over the sessions and years. This is adequate for the successful realization of our analysis. However, the original data goes through several stages of cleaning and remodeling before it can be used for each analysis. To ensure the integrity of the data at each stage, we systematically implement the "four eyes" principle, meaning that all data manipulations are reviewed by another member of the project team.
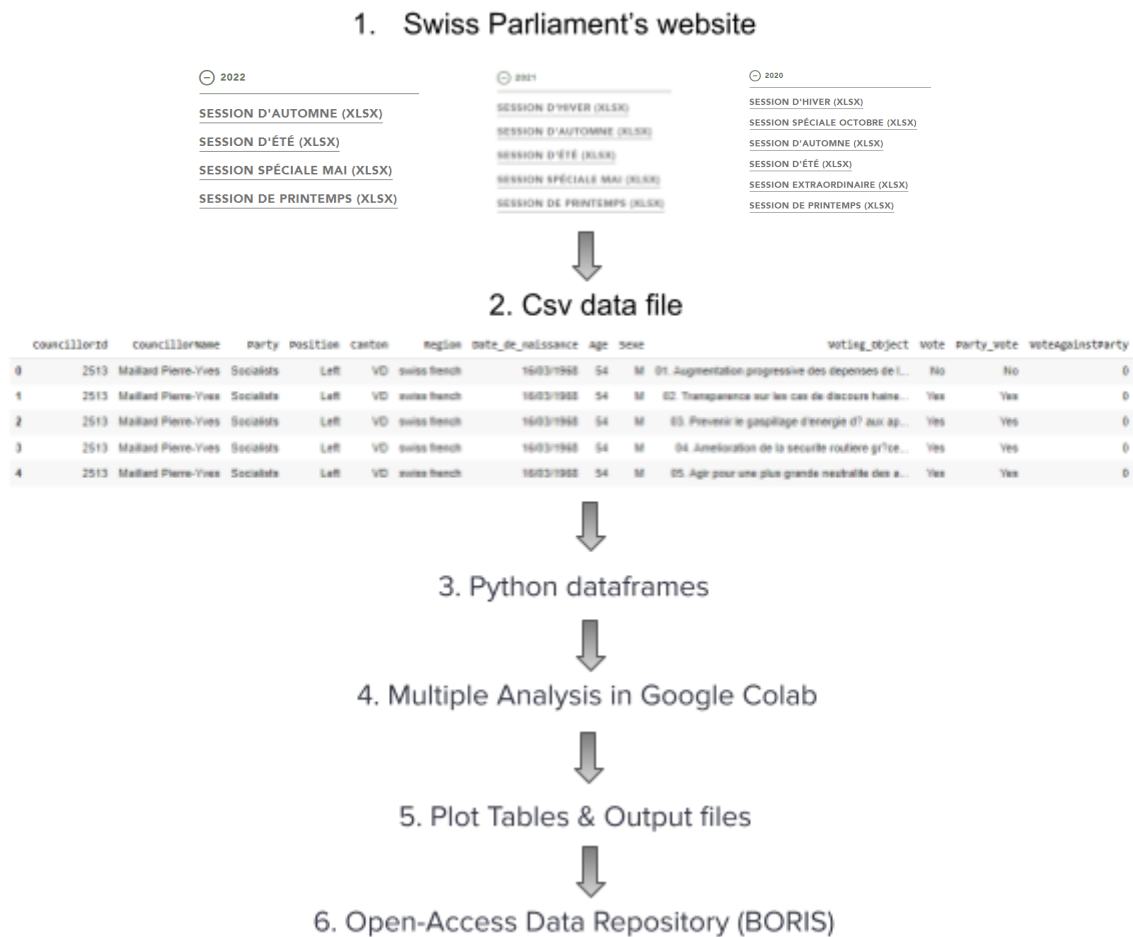
6 Data Flow

Figure 1: Data Flow

Figure 1 summarizes the dataflow of the project. The raw data is extracted from the official Swiss Parliament's website. At this stage, it consists of several excel tables listing all votes of each member of the National Council per session. In the next step, we consolidate these different tables into a single csv table, listing all the votes of all elected representatives for all sessions. We also add the age and the gender of each Councilor, which are not available in the primary datasets. We convert this new dataset into a main Python Dataframe. This main dataframe is then copied and remodeled to go through the analysis in Google Colab. Beyond the descriptive statistics, analysis consist of paired t-test to test the difference of voting deviation (i.e. votes against group majority) and compare parliamentary groups, and a regression analysis to identify explanatory variables with a significant effect on voting (non-) compliance. We then store our plots, tables and others outputs in identifiable files in the shared Google Drive directory. At the end of the project, the data and the metadata will be published on the non-profit, open-access data repository of the University of Bern (BORIS).

7 Data Model

Conceptual

We identify two entities: (1) National Councilors and (2) the voting topics. Voting represents the relationship between both entities. National Councilors have various attributes, such as their parliamentary group, age, gender and canton. Within the scope of this project, we only consider one attribute for the voting topic: the federal department responsible for the object. Entities and their attributes are summarized below (Figure 2: Conceptual Data Model).
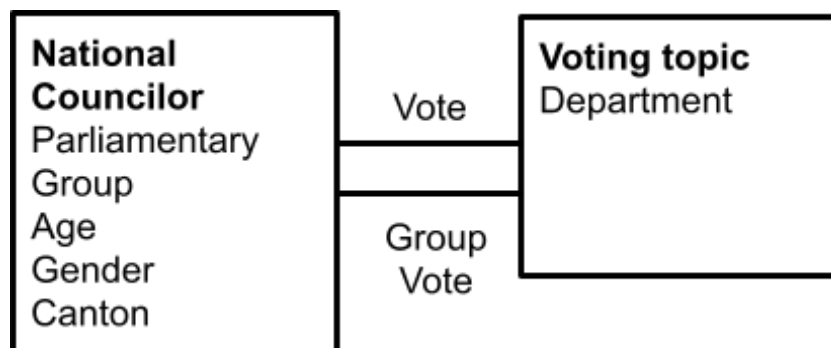

Figure 2: Conceptual Data Model

Logical

1)

Primary keys

Councilor ID	Councilor Name	Group	Age	Gender	Canton
12345	John Doe	V	55	M	VD
...	...	...	...	...	
12346	Jane Does	S	38	F	BE

2)

Primary key

Voting Object n°	Department	Vote John Doe	...	Vote Jane Does
1	EDA	No	...	No
...	...	...	...	
i	DETEC	Yes	...	No

Primary key

Voting Object n°	Department	Vote Group V	...	Vote Group S
1	EDA	Yes	...	No
...	...	...	...	
i	DETEC	Yes	...	No

3)

Primary keys

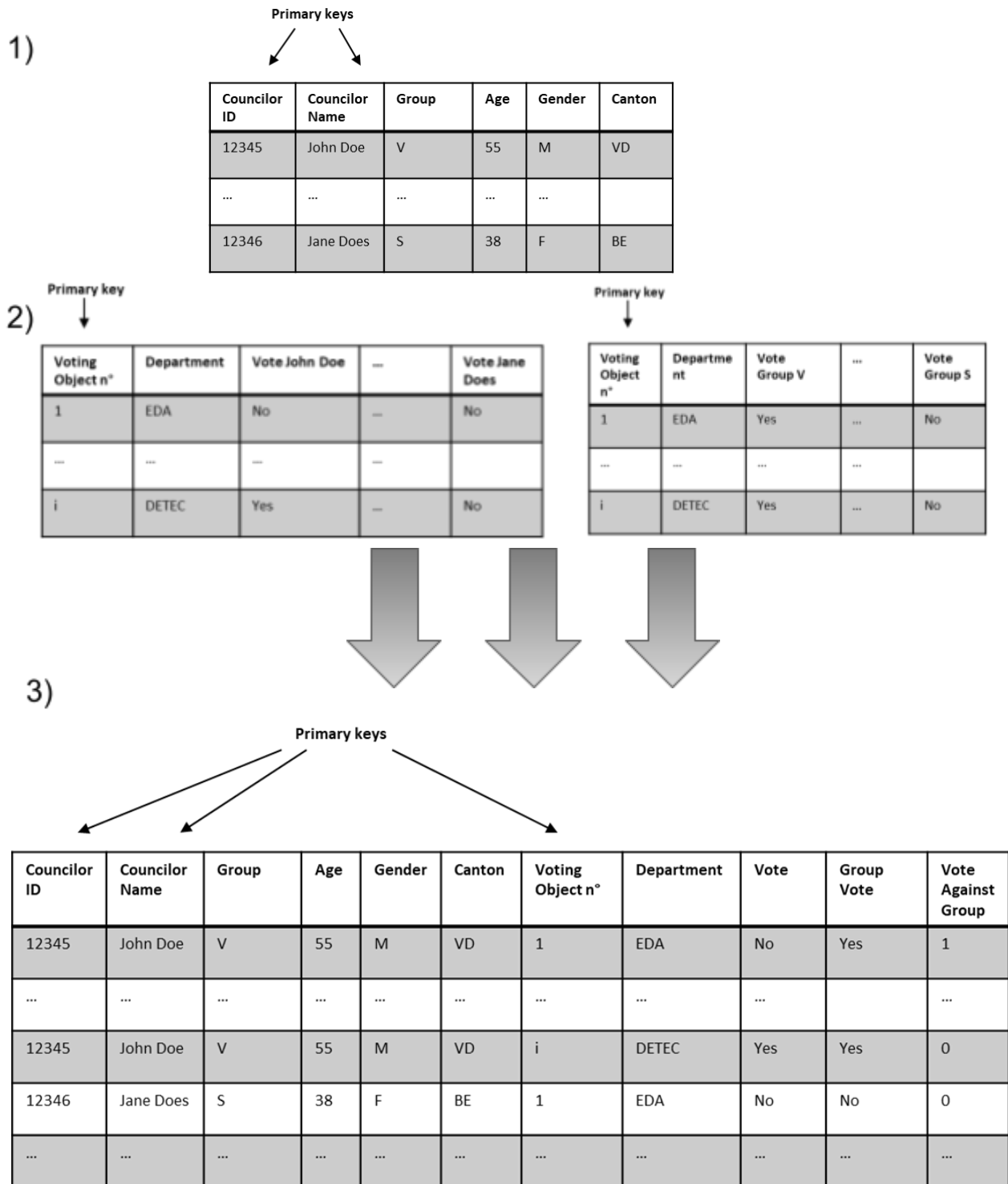| Councilor ID	Councilor Name	Group	Age	Gender	Canton	Voting Object n°	Department	Vote	Group Vote	Vote Against Group
12345	John Doe	V	55	M	VD	1	EDA	No	Yes	1
...	...	...	...	...	...	...	...	...		...
12345	John Doe	V	55	M	VD	i	DETEC	Yes	Yes	0
12346	Jane Does	S	38	F	BE	1	EDA	No	No	0
...	...	...	...	...	...	...	...	...	...	...

Figure 3: Logical Data Model

The logical data model is summarized in Figure 3 : Logical Data Model. The first table lists the elements of the first entity, the National Councilors, and their attributes. The two tables list the second entity, as well as the vote of each National Councilor (left table), and the vote of each parliamentary group (right table). The table at step 3) brings together both entities and the voting behavior, and adds a variable derived from the voting behavior, i.e. whether the National Councilor voted against its group or not.

Physical

The final dataset is structured as shown on the logical model in Figure 3. The physical infrastructures required for this project are quite basic. To access the common files, each member needs a computer and a Google account. The size of the datasets is at all times well within the range of what can be stored on Google Drive without the need to pay for storage. Analysis can be done easily using Google Colab. Because of the reasonable size of the dataset and number of variables, we forego the creation of a more complex database, and do not implement a relational database management system.

8 Risks - What can go wrong?

As mentioned previously, the original data goes through several stages of processing and remodeling. The first risk therefore lies in a human error during data processing, which would then distort the results of our analysis. To ensure the integrity of the data at each stage, we therefore systematically implement the "four eyes" principle, which means that all data manipulations are proofed by another member of the project team. The validation of data manipulations by a second team member obviously requires personal and time resources. However, we understand code proofing and the "four eyes" principle as standard procedures, recognized as good practice in data science and guaranteeing the integrity of the data and the legitimacy of the results. We also plan on publishing our data as well as our codebooks in a digital open-access data repository, making our research reproducible. This, although it does not directly prevent inconsistencies or errors in data manipulation, allows the scientific community to verify our processes and gives more legitimacy and transparency to our research. The work required to prepare templates that are publishable on a data repository is also considered as part of standard procedure, as best practices in data science and research require that the data and the coding be documented at all times in a way that is accessible by third parties.

We also identify a risk in the use of Google Drive as our working platform. Not because the platform is unreliable in itself, but because, as everywhere, there is the risk of a cyber-attack leading to the loss of all the data and the analysis work done, which would obviously cause a

huge setback and a loss of resources. To prevent this issue, we are considering regular backups on the local server of the University of Bern. The IT department of the University of Bern also performs backups, which should protect our data and our work in case of a cyber-attack.

9 Preliminary Studies

A study conducted by the RTS, a Swiss French speaking public owned broadcasting company, found that, in the first half of 2020 at the beginning of the current legislature, women were found to swing one in ten votes in the National Council. Authors argue that the increased presence of women in parliament after the national 2019 elections is leading to changes in decision-making within the National Council, and that in particular women from the center-right tend to go against their party on specific societal issues, including equal pay, and more generally on social policies [1]. Among other things, we want to test whether these figures from the very first moments of the new legislation are confirmed throughout the legislation. Moreover, we also want to identify other factors that may lead to a vote that does not comply with the parliamentary group.

As part of the module 2 of the CAS in Applied data science at the University of Bern, we conducted a preliminary analysis on the data of one session, the May 2022 special session. We formulated three hypothesis with regard to the voting behavior of the Swiss National Councilors:

1. Based on the RTS article of 2020, we expect women to be more likely to vote against their parliamentary group.
2. Based on that same source, we expect Councilors located in the center of the political spectrum to be more disposed to vote against their parliamentary group.
3. We also expect older members of the National Council to vote more often against their group. This because older Councilors usually have been longer in office and are likely to rely less on coalition majority.

First, we performed some descriptive statistics to analyze the composition of the National Council, with regards to the distribution of the political orientation, the age and gender of the Council members.

Our first descriptive statistics, that can be verified in the official parliamentary statistics [2], show that:

- Men are more represented than women (Figure 4: Gender distribution in the National Council, May 2022).

- The political orientations "Right" and "Left" are the most represented. Only ~15% of the National Council members belong to the "Center" (Figure 5: Political orientations distribution in the National Council, May 2022).
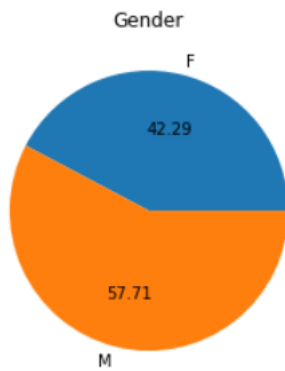

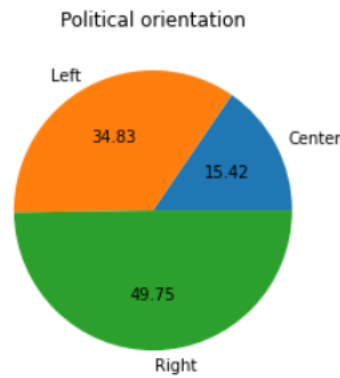Figure 4: Gender distribution in the National Council, May 2022

Figure 5: Political orientations distribution in the National Council, May 2022

We then proceeded to analyze the vote of councilors casting their vote against the majority of their own parliamentary group. Further analysis regarding those "deviant voting" seem to indicate that:

- Men tend to disagree more with their group than women. Older people also seem to disagree more with their group (Figure 6: "Deviant voting" and gender distribution in the National Council, May 2022).
- Most people voting against the Party seem to belong to the center. Having less extreme views in politics, we can assume they might be more prone to have different opinions depending on the matter (Figure 7: "Deviant voting" and political orientation in the National Council, May 2022 and Figure 8: Vote deviation frequency and political orientation in the National Council, May 2022).
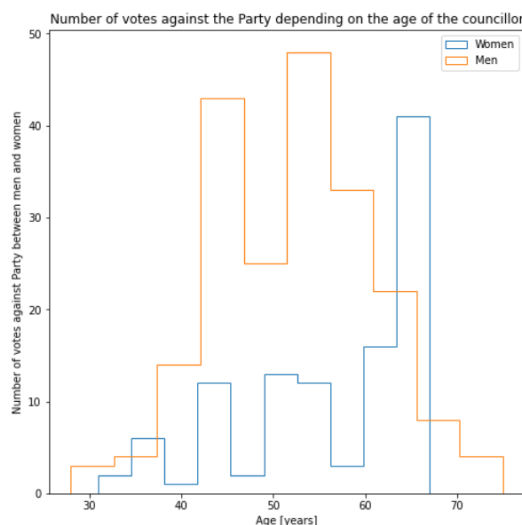

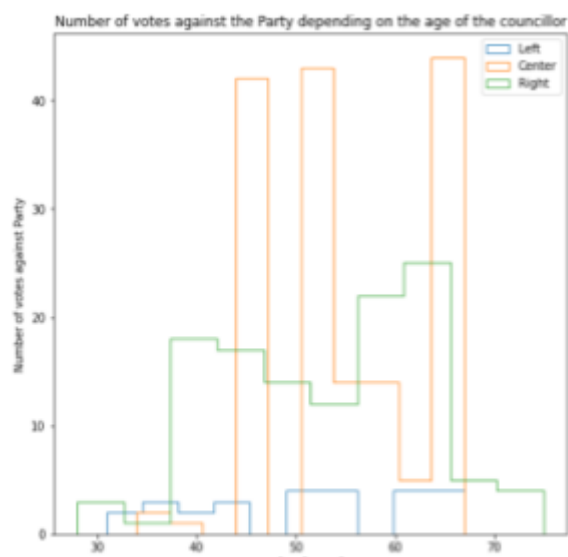Figure 6: "Deviant voting" and gender distribution in the National Council, May 2022

Figure 7: "Deviant voting" and political orientation in the National Council, May 2022
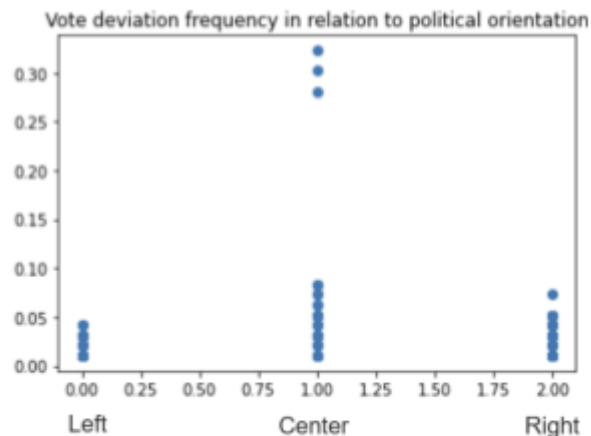

Figure 8: Vote deviation frequency and political orientation in the National Council, May 2022

This preliminary analysis gives us a good overview of the composition of the National Council. There is an indication that the behavior of "deviant voting" might be more present in the center. However, the hypothesis that women disagree more with their group cannot be validated so far. The dataset being quite limited in terms of size, it is difficult to draw clear conclusions at this point.

The next step is therefore, in line with the aim of this project, to aggregate the data over the whole legislation to take a deeper look at the content of the voting topics, with the ambition to conduct comprehensive analyses and deliver results that allow an in-depth understanding of voting behavior and moreover voting non-compliance in the Swiss National Council.

10 Conclusions

The aim of this project is to identify which type of Swiss National Council members are more likely to cast their vote against the majority of their own parliamentary group, based on individual attributes such as gender, age and political orientation and the voting topic for the current legislative period (2020-2022). Based on preliminary analysis on a smaller dataset, we formulated the hypothesis that Council members with a Center political orientation, having less clear-cut positions, could be more prone to this behavior of "deviant voting". With the larger dataset considered in this project, we expect to obtain firmer results to test this hypothesis and to potentially identify other correlations between the different attributes. In a later step, these results

could be verified by reiterating the same analysis on the voting reports of the next legislative period.

We considered several attributes related to the Council members in order to establish a "typical" profile but, due to lack of available data at this stage, only one attribute regarding the voting topic, which is the federal department responsible for the matter in question. Although the voting behavior is without a doubt a personal matter, the topic of the vote as well as external parameters can cast a strong influence. The scope and limitation of the attributes we considered for this project lies in the data source that was used, the public voting reports. Although the quality of the data is very high, it has limitations to conduct a more comprehensive analysis. For further analyses, it will be relevant to build a more complex data model and consider additional data source to include information related to the matter of vote, such as the theme (e.g. social, economics, healthcare, etc.), the media coverage, the political campaigns and the lobbying activities in the analysis.

Acknowledgements

This report is written as part of Module 1 of the CAS in Applied Data Science of the University of Bern. We would like to acknowledge and thank the program management and the lecturers of Module 1.

References and Bibliography

[1] Cabré, Gabriela (2020). *Les femmes font basculer un vote sur dix au Conseil national*. RTS Info, consulted on October 1, 2022. Last modified on September 1, 2020 https://www.rts.ch/info/suisse/11546799-les-femmes-font-basculer-un-vote-sur-dix-au-conseil-national.html

[2] Swiss Parliamentary Website. *PARLEMENTAIRES*. Swiss Parliamentary Website, consulted on October 1, 2022. https://www.parlament.ch/fr/%C3%BCber-das-parlament/faits-donnees-chifrees/chiffres-parlementaires