UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

FRANCISCO GERDAU DE BORJA

CivisAnalysis: Exploring Representatives' Voting Behaviour

Thesis presented in partial fulfillment
of the requirements for the degree of
Master of Computer Science

Advisor: Carla Maria Dal Sasso Freitas

Porto Alegre
April 2017

CIP – CATALOGING-IN-PUBLICATION

Gerdau de Borja, Francisco

CivisAnalysis: Exploring Representatives' Voting Behaviour / Francisco Gerdau de Borja. – Porto Alegre: PPGC da UFRGS, 2017.

69 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2017. Advisor: Carla Maria Dal Sasso Freitas.

1. Visualization, Political Data, Roll Call Analysis. I. Dal Sasso Freitas, Carla Maria. II. Título.

AGRADECIMENTOS

Gostaria de agradecer imensamente a ajuda, otimismo e paciência que minha orientadora Carla Maria Dal Sasso Freitas teve durante esses anos de trabalho, Rodrigo Moni por ter me ajudado com as avaliações e aos grandes amigos que encontrei e me acompanharam durante a pós-graduação no Graphics, Visualization and Interaction Lab no Instituto de Informática da UFRGS.

Agradeço em especial a todos que cooperaram e defendem o programa de concessão de bolsas de pós-graduação do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pois sem este programa seria muito difícil a realização deste trabalho.

ABSTRACT

In this work, we present CivisAnalysis, an open-source web-based system for the visualization of roll calls in the Brazil's Chamber of Deputies. Using roll calls of six legislatures as well as six presidential elections, CivisAnalysis provides a unique view of the political history of our country. It combines classical roll calls visualization techniques with techniques for the visualization of temporal data.

Votes of each representative are interpreted as a set of multidimensional data, which are represented in a n-dimensional space mapped to a bi-dimensional representation. The Votes of each representative determine their position in the political spectrum of deputies, and the votes of all deputies determine the political spectrum of roll calls. The work also shows how exploring subsets of deputies allows the discovery of voting patterns in the roll call spectrum, as well as exploring the roll calls spectrum allows the discovery of behavioral voting patterns in the political spectrum of deputies.

For supporting a long-term political analysis, CivisAnalysis provides a time line visualization integrated with election data (election results and political alliances). The tool implements textual and visual filtering and includes auxiliary visualizations that provide an overview of the political scenario regarding deputies, parties, coalitions and their behavior along time.

Keywords: Visualization, Political Data, Roll Call Analysis.

CivisAnalysis: Explorando o Comportamento de Congressistas em Votações

RESUMO

Este trabalho apresenta CivisAnalysis, uma aplicação web de código aberto para visualização das votações ocorridas na Câmara dos Deputados do Brasil. Com dados abrangendo seis legislaturas e seis eleições presidenciais, CivisAnalysis provê uma visão inovadora da história política do Congresso Brasileiro combinando técnicas clássicas de visualização de votações e técnicas de apresentação de dados temporais.

As votações de cada deputado podem ser interpretadas como dados multidimensionais e são representadas em um espectro n-dimensional mapeado para uma representação bidimensional. Os votos de cada deputado determinam suas posições numa representação conhecida como espectro político. O trabalho mostra como a exploração de conjuntos de Deputados permite a descoberta de padrões de votos no espectro de votações, assim como explorar conjuntos de votações permite a descoberta de padrões de comportamento no espectro de Deputados. Para uma análise política de longo termo, foi criada uma linha do tempo apresentando as distâncias relativas dos partidos no espectro político durante seis legislaturas e eleições presidenciais. Na linha de tempo é possível perceber os alinhamentos políticos e mudanças comportamentais de partidos conforme o progresso de legislaturas, alianças eleitorais e coalizões de governo.

A interface de CivisAnalysis oferece filtros textuais e visuais e inclui visualizações auxiliares para destacar cenários políticos de acordo com Deputados, partidos, unidades federativas, alianças eleitorais e suas posições no espectro político

Palavras-chave: Visualização de Dados Políticos, Espectro Político, Análise de Votações.

LIST OF ABBREVIATIONS AND ACRONYMS

nD	n-dimensional
PCA	Principal Component Analysis
C2C	Connect 2 Congress
DRM	Dimensionality Reduction Method
SVD	Singular Value Decomposition
t-SNE	t-Distributed Stochastic Neighbor Embedding
CT	Citizen Task

LIST OF FIGURES

CONTENTS

1 INTRODUCTION

Active and informed citizens are essential for a healthy democracy, but being informed about public affairs has become a challenging task due to the immense size of modern representative democracies. In representative democracies, the most important tool of citizenry is the vote, which is how citizens can provide feedback that may improve the democratic system. To be able to vote effectively, a citizen needs consistent political information, especially regarding the actions of elected representatives. Political data about representatives can be found in every part of their public lives: speeches, interviews, meetings, etc. However, their votes in congress are especially important as they have a direct impact on society. How representatives behave in roll calls is thus an invaluable source of data that can be used to keep citizens informed.

1.1 Motivation

During roll calls, each representative explicitly chooses an option regarding a matter that is being decided ("Yes", "No" or "Abstention"). Beyond this individual choice, each option in a roll call can reveal the behavior of groups in congress and in society at large. Each representative vote is directly associated to an electoral district and a party; indirectly, it is associated to personal characteristics such as religion, gender, wealth, etc. This makes roll call votes a reliable data set for understanding representatives' policies as well as parties' policies in several aspects. However, the large number of votes each year makes it impractical for common citizens to track representatives' voting records.

A simple way of understanding representatives' behavior is to observe how they are distributed in a n-dimensional space where similar behavior corresponds to closer positions in that space. Similar behavior means similar votes in the same roll calls. A geometric representation of such space is called "political spectrum", which also provides visual information about the alignment of representatives with government or opposition.

Open-data government initiatives give access to different political data sets, allowing for the easier dissemination of information. On the other hand, the huge volume of unorganized and highly biased information spread across the Internet may cause a negative impact on the assimilation of information due to the effect commonly known as information overload. It is therefore challenging to retrieve cohesive information from the political big data and communicate it to the common citizen.

The inspection of a single roll call without assistance of a tool is feasible, but usually one may also need/want to obtain information concerning larger periods of time—e.g., retrieving votes of different sets of deputies (e.g., parties, alliances and districts) linked to sets of roll calls (possibly about a specific subject) within a specific time interval. An exploration tool to quickly test hypotheses is therefore necessary to be able to draw conclusions about larger and/or different political contexts.

1.2 Objectives and Contribution

This work focuses on roll calls that have taken place in the Brazilian Chamber of Deputies, the lower house. The chamber is formed by 513 deputies elected by proportional representation in electoral districts to serve four-year terms. Brazil has 27 electoral districts, each with a minimum of 8 and a maximum of 70 seats; the exact number being proportional to each district's population. Proportional representation and flexible electoral alliance laws make it possible for Brazil's multiparty system to have a large number of parties. The last legislature (2011-2014), with 21 parties, voted in 352 roll calls (a total of 114,994 votes).

Using the open data from the Brazilian Chamber of Deputies, we developed CivisAnalysis, a visualization tool aimed at making it easier and quicker to discover and understand congressional voting patterns. This tool integrates state-of-the-art and new visualization techniques for roll calls analysis, using a data set that covers voting records of six legislatures (24 years)—comprising 914 motions voted through 2,458 roll calls (853,952 votes)—as well as the information of six presidential elections, including election results and alliances made for the elections.

By providing several exploration features, the tool allows for the discovery of the voting pattern of any arbitrary number of roll calls in the deputies data set as well as the voting pattern of any arbitrary number of deputies in the roll calls data sets, along different time frames across legislatures.

The main contributions of the present work are:

1. a visualization of the outcome of roll calls as a projection of a n-dimensional (nD) space, linking the votes of a set of roll calls with the political spectrum

of deputies and—vice versa—the votes of a set of deputies to the roll calls political spectrum;

2. the integration of long-term political and electoral data (result and alliances) in the roll call analysis.

1.3 Structure of the Dissertation

The remaining text of this dissertation is organized as follows. Next chapter (Chapter 2) gives a brief background on Brazilian politics and roll call analysis. In Chapter 3, we review literature concerning political data visualization techniques. Then, in Chapter 4, we present our tool, which is called CivisAnalysis, describing its visualization techniques and several features based on how it can be used to fulfill different information needs that typical users might have. We report evaluation of the use of CivisAnalysis by people with different levels of expertise in Chapter 5 and discuss the importance of such data for the analysis of the Brazilian Chamber of Deputies, from a user's point of view. Finally, in Chapter 6, we comment on the improvements we foresee as future work.

2 BACKGROUND

2.1 Brazilian Politics

Brazil is a democratic republic, with a presidential system. The president is both head of state and head of government of the Union and is elected for a four-year term, with the possibility of re-election for a second successive term. Legislative houses in each political entity are the main source of law in Brazil. The National Congress is the Federation's bicameral legislature, consisting of the Chamber of Deputies and the Federal Senate. Judiciary authorities exercise jurisdictional duties almost exclusively.

All members of the executive and legislative branches are directly elected. Judges and other judicial officials are appointed after passing entry exams. For most of its democratic history, Brazil has had a multi-party system, proportional representation. Voting is compulsory for the literate between 18 and 70 years old, and optional for illiterates and those between 16 and 18 or older than 70.

The Constitution organizes the country as a Federate Republic, formed by the indissoluble union of the states and municipalities, and the Federal District. The 26 federate states have powers to adopt their own Constitutions and laws. Their autonomy, however, is limited by the principles established in the Federal Constitution.

2.1.1 General Elections

A general election is held every four years. Brazil elects on the national level a head of state – the president – and a legislature. The president is elected for a four-year term by absolute majority of votes through a two-round system. The federate states elect governorship and state legislatures.

The National Congress (*Congresso Nacional*) has two chambers. The Federal Senate (*Senado Federal*) has 81 members, elected for an eight-year term, with elections every four years for alternatively one-third and two-third of the seats. The

Chamber of Deputies (*Câmara dos Deputados*) has 513 members, elected for a four-year term by proportional representation.

The Chamber of Deputies members are elected by proportional representation within its electoral district. Each electoral district elects a number o deputies proportional to its inhabitants at minimum of 8 and maximum of 70. In the 2014 general election the most populous electoral district in Brazil, São Paulo, elected 70 deputies; eleven states elected the minimum quota of deputies. The maximum and minimum quota of deputies and the high number of districts electing the minimum quota create a distortion in the proportional representation. Seats in the Chamber of Deputies represent distinct number of citizens; for example, a deputy from São Paulo district represents 591.000 inhabitants, while from one from Roraima district, only 52.000.

Within each electoral district deputies are elected to the (Federal) Chamber of Deputies using a form of party-list proportional representation known as the open list. The party chooses the candidates, and citizens vote in a candidate or party. The candidate list of each party is sorted by the nominal votes. The number of seats for each party are distributed by a variation of d'Hondt method (D'HONDT, 1878). The variation consist of: allied parties at district level share their votes in the allocation of seats. Two distortions can occur in this case: (1) candidates of small parties are elected by the votes of an allied big party. The citizen vote in the party and its respective proportional representation is diffused to the district alliance, which is independent from the federal alliance; (2) a candidate with a huge number of votes "elect" allied candidates who would not have enough votes to be elected.

Electoral alliances (*Coligações*) are made at federal level and at the 27 electoral districts (26 federal states plus Federal District). Alliances at federal level and at electoral districts can occur independently. The federal electoral alliance tend to become the coalition government, to form a majority in next legislature. The "alliance verticalization", where alliances in electoral districts should follow the federal alliances, was extinct in 2006.

After elections, in order to create a governable state, or approve the programs promised in the electoral campaign, multi-party systems tend to form a **coalition alliance**. Coalitions are a multi-party pro-government majority as well as the opposition group. The ideological spectrum of parties in these legislatures is partially reduced to a government-opposition dimension. Studies on Brazilian legislatures and coalitions cite the important role of presidential elections and electoral alliances on the patterns of legislative behavior (ZUCCO, 2009)(KRAUSE; SCHMITT, 2005)(MORGENSTERN, 2004). As such, having electoral alliances and coalitions as parameters for the visualization may be an important part of the

process of analyzing the data extracted from the behavior of representatives during roll calls.

2.1.2 Roll Calls in the Chamber of Deputies

As mentioned before, the Chamber of Deputies has 513 members, elected to a four-year term by proportional representation, the current legislature have 28 parties. Since 2012 the Brazilian Chamber of Deputies provides an open web service to access data from the legislatures. The data includes legislative motions (proposed laws, procedures, etc.), roll calls, deputies information, budget, speeches in plenary session and committees processes.

A roll call is the stage of the legislative process that completes the regimental turn of the discussion of a proposition. The validity of a roll call depends on the quorum check. There must be at least 257 deputies in the plenary.

Votes required to pass a motion varies according to the type of proposal. To approve a bill, a quorum of 257 deputies is necessary. Obtained the quorum, the project can be approved by simple majority, i.e. the majority of those deputies present in the plenary session. Proposed amendments to the Constitution (*Projeto de Emenda Constitucional*) must be approved by the plenary in two rounds, with the votes of 3/5 of the deputies (308 votes). The "complementary bills" (*Projeto de Lei Complementar*) also pass by two rounds of voting. For approval, 257 favorable votes are needed, i.e., absolute majority of deputies.

All roll calls are recorded and can be accessed through the open web service, including roll calls without quorum. The data consist in an array of nominal votes, i.e., the name of deputy and its vote value. The vote values are:

Yes Approve the motion.

No Reject the motion.

Abstention No political option; counted for quorum purposes.

Obstruction Did not vote, left plenary or directly expressed "in obstruction"; do not count for quorum purposes.

Null Deputies who were not present are not registered; do not count for quorum purposes.

White Depreciated value of Abstention.

Art. 17 Chamber's president neutral value, the vote of the Chamber of Deputies president is only requested to break a tie.

The web service does not provide concise information of the deputy history, for example, change of party or electoral district. It is possible to have a coarse estimation of the deputy party or electoral district at a point in time using the party defined in his last votes at the Congress, since the roll call data define the party and the district of each deputy's vote. Another problem that the developer may encounter using the Chamber's web service is finding an unique identifier to track the deputies. There is an id (*idcadastro*), but it can not be used to track the deputies on long term because different deputies may have the same id, since deputies change over time. We decided to track deputies by their names, there is some misspelling, which we detected with Levenshtein Distance (LEVENSHTEIN, 1966), also known as Edit Distance, and fixed manually.

2.2 Roll Call Analysis

Today, most theories of campaigns, legislatures and elections are based, explicitly or implicitly, on the spatial model of politics. Studies aim at positioning each political player at its ideal point on a spectrum (ENELOW; HINICH, 1984). This produces techniques with results that can be interpreted both in probabilistic and in geometric terms.

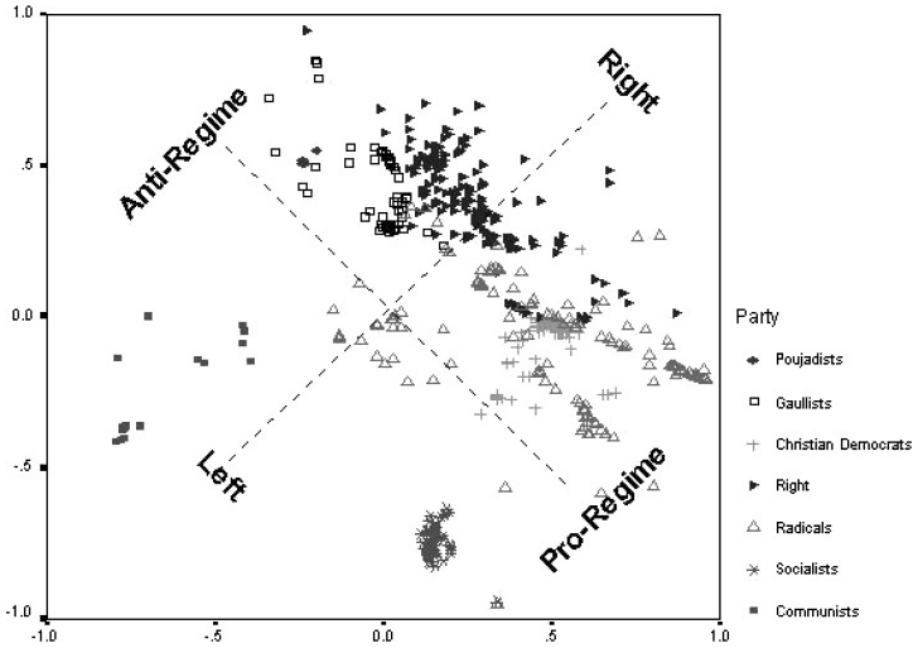
Researchers estimate congressional spaces from transcripts of hearings, bill cosponsorships, public opinion surveys, committees seats, campaign contributions and roll call votes (FISHER; LEIFELD; IWAKI, 2013)(DESPOSATO, 2006).

Roll calls are commonly used in spatial models because of their nature of reliably expressing the policy of each representative. The Spatial Theory of Voting put into practice with roll call analysis was concisely highlighted by Clinton (CLINTON; JACKMAN; RIVERS, 2004): "In short, roll call analysis makes conjectures about legislative behavior amenable to quantitative analysis, helping make the study of legislative politics an empirically grounded, cumulative body of scientific knowledge."

The Poole and Rosenthal algorithms (POOLE; ROSENTHAL, 1985), known as the NOMINATE family, are widely used in roll call analysis and compare favorably to more modern algorithms (CARROLL et al., 2009). NOMINATE is a spatial model to describe the parliamentary roll call votings. The voting decision of the legislators are modeled as normally distributed utility functions. The ideal points of legislators in the spatial model are estimated by maximum-likelihood estimation (MLE) methods.

Later, Poole developed the Optimal Classification method (POOLE, 2000) for roll call analysis. Which is best suited for multi-party political systems where the legislature have the following characteristics: "variation in discipline across parties, unstable party memberships, proxy voting, near perfect two-dimensional spatial

voting, and parliamentary institutions that provide incentives for strategic voting." (ROSENTHAL; VOETEN, 2004). The Optimal Classification method is probably best suited to analysis of the Brazilian legislatures since have similar characteristics to the French Fourth Republic (1946–58) analyzed in the Rosenthal and Voeten work; the deputies ideal points plot can be seen in Figure 2.1.



Source: (ROSENTHAL; VOETEN, 2004)

Figure 2.1: Optimal Classification method of the National Assembly of France (1946–1958).

In contrast to the complexity of previous methods, fast and less accurate (for roll call analysis) dimensionality reduction methods are used, such as Principal Component Analysis (PCA). PCA or Karhunen-Loeve transformation (JOLLIFFE, 1986)(DONY, 2001) reduce the number of dimensions while retaining the variance of the data. These dimension reduction techniques try not to position different points together, but remove correlations. The remaining subset of dimensions are a compact summary of the variation in the original data. Despite the relative inaccuracy of PCA to find ideal points on the roll call space, the voting patterns formed are quite similar to those obtained with more complex methods (JAKULIN; BUNTINE, 2004).

2.2.1 Principal Component Analysis

Principal component analysis is a widely used unsupervised technique that reduces high dimensional data to a more manageable set of new variables which sim-

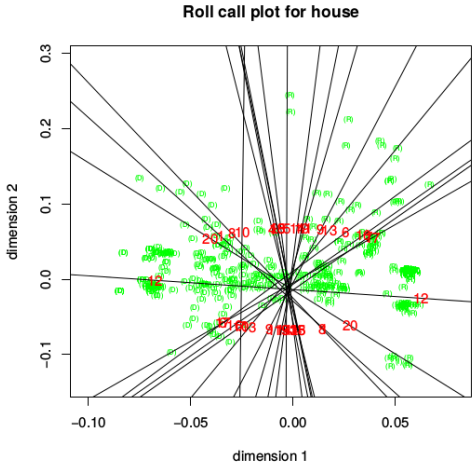
plifies the visualization of complex data sets for exploratory analysis. In PCA, the new variables (principal components, PCs) are linear combinations of the original variables extracted in order of the amount of data variance that they explain. The contribution that each variable makes to a PC is called the loading, and the amount of the PC present in a particular sample is known as the score. The results of PCA are visualized as projections of multidimensional scores and loadings in two- or three-dimensional plots. Understanding the sample information (scores) can be straightforward if there are few samples and any expected separation is observed, but is more difficult if the classes are unknown or unexpected clusters are detected. Interpreting the variable behavior (loadings) and understanding the significance of the combinations of variables can be challenging, especially if there are many significant PCs, but is important and can help explain unexpected groups in the sample (IVOSEV; BURTON; BONNER, 2008).

2.3 Final Comments

As mentioned before the main contribution of this work is the visualization of the roll Call space and the linked, synchronized visualization, with the deputies space to best understand the legislature. The PCA suits this task since it is a simple method to obtain both projections in a direct relationship. The results of a roll Call Analysis based on PCA are projections of the deputies space, the scores; and the roll Call space, the loadings.

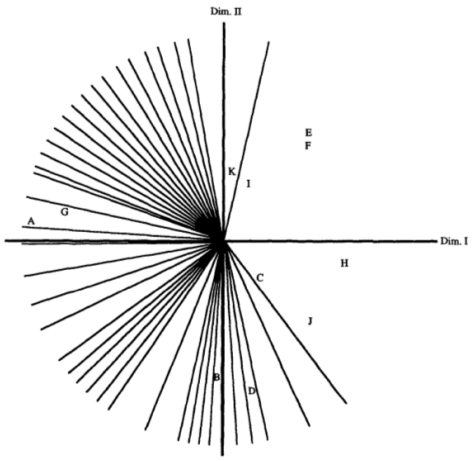
The link between the two spaces is clarified by Poole: "If the individuals are treated as roll calls and the roll calls are treated as individuals, then the individual becomes a cutting plane through the space, and the point where the cutting plane passes through the normal (individual) vector is the individual's threshold. That is, the individual approves (accepts) the stimuli on one side of the plane. Pick-any-N data, widely used in marketing applications (DESARBO; CHO, 1989), has this form. For example, respondents are given a list of soda pops are then displayed as points in a space, and the individuals as cutting lines that divide the soda pops into drink and do not drink." (POOLE, 2005) - see Figure 2.2(b) for the soda pop preferences plot and cutting plane of individuals.

In fact, many works depicted the political actors ideal points and the cutting planes of roll calls - see Figure 2.2. A central idea of CivisAnalysis is to provide the user the interaction in the both spaces, deputies and roll calls, and on demand, display the cutting plane and the values of votes of the selected deputy or roll call in the roll calls space or deputies space, respectively.



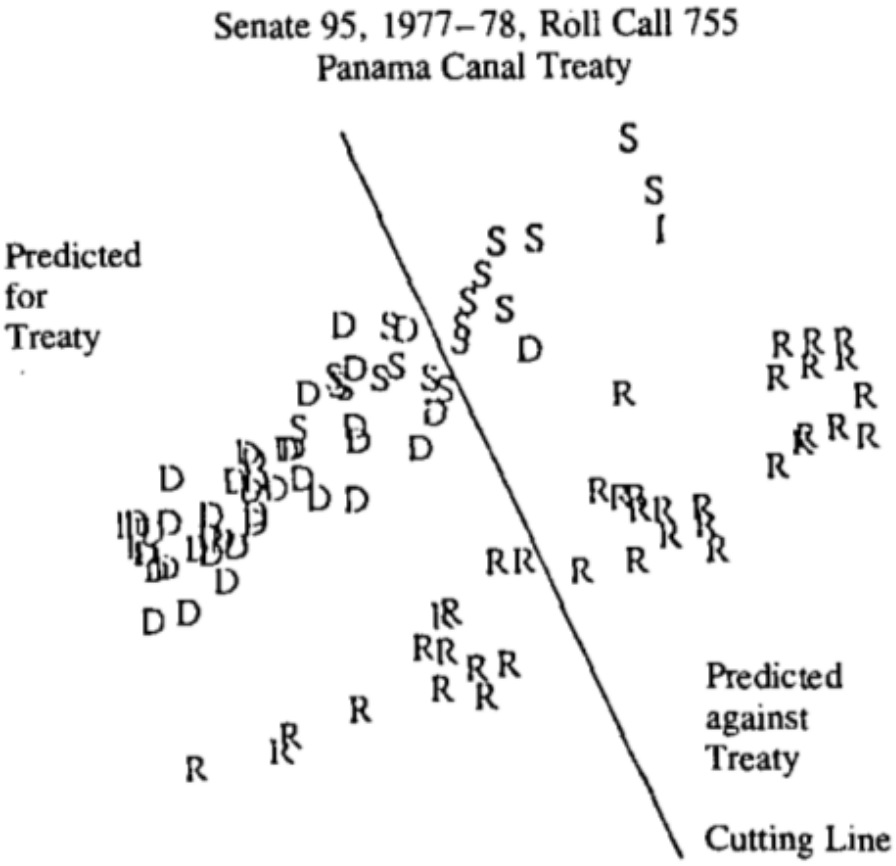
(a) Binary-PCA, plot of the 2001 U.S. Senators ideal points, and the roll calls cutting planes labeled by numbers, in red.

Source: (LEEUW, 2003)



(b) Pick any/J data method, soda pop ideal points, individuals are depicted by the cutting planes.

Source: (JEDIDI; DESARBO, 1991)



(c) D-NOMINATE plot of th U.S. Senate and the cutting plane of the Panama Canal Treaty roll call.

Source: (POOLE; ROSENTHAL, 1991)

Figure 2.2: Ideal-points spaces and cutting planes.

3 VISUALIZATION OF POLITICAL DATA

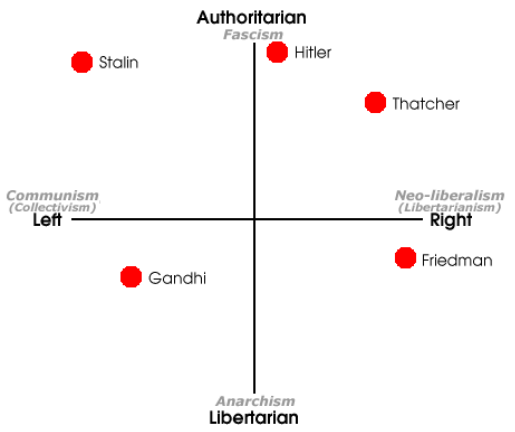
In this chapter we briefly review the related work on visualization of political data. Some of the works are results from research projects, but others are practical applications used mainly for communication in news media.

3.1 Political Spectra

Political spectrum visualizations are widely used to represent different political positions according to one or more geometric axes that represent independent political dimensions. Systems based on such visualizations try to provide ways of solving the problem of how to describe the political variation in a given space. Early systems were usually built around questionnaires and regression analysis, and are often considered highly biased because constant adaptation was needed to represent different political contexts (LESTER, 1994).

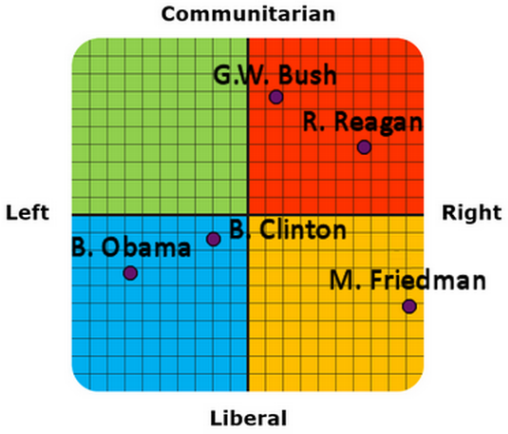
As one of the popular political spectra on social media that are created using questionnaires we find Political Compass (COMPASS, 2015) and Political Coordinates (CELEBRITYTYPES, 2015) (Figure 3.1). In such tools, the user can create its own entry to compare himself with friends or historical personalities. Besides political spectrum both applications provide spectra of different parties by country and specific elections.

In order to reduce bias, modern systems tend to use only quantitative information about representatives, especially from recorded votes. In this case, the political spectrum is built based on the results from roll call analysis techniques. Spatial models created by roll call analysis techniques are traditionally represented as scatterplots, where representatives are displayed as points whose positions represent their values associated to the variable axes. As such, the closer the points are, the more their corresponding representatives' policies agree. To help identify representatives, a table can be used. Nowadays, it is possible to apply interactive techniques over the spectrum to identify neighbor points and calculate distances between representatives.



(a) Political Compass

Source: (COMPASS, 2015)



(b) Political Coordinates

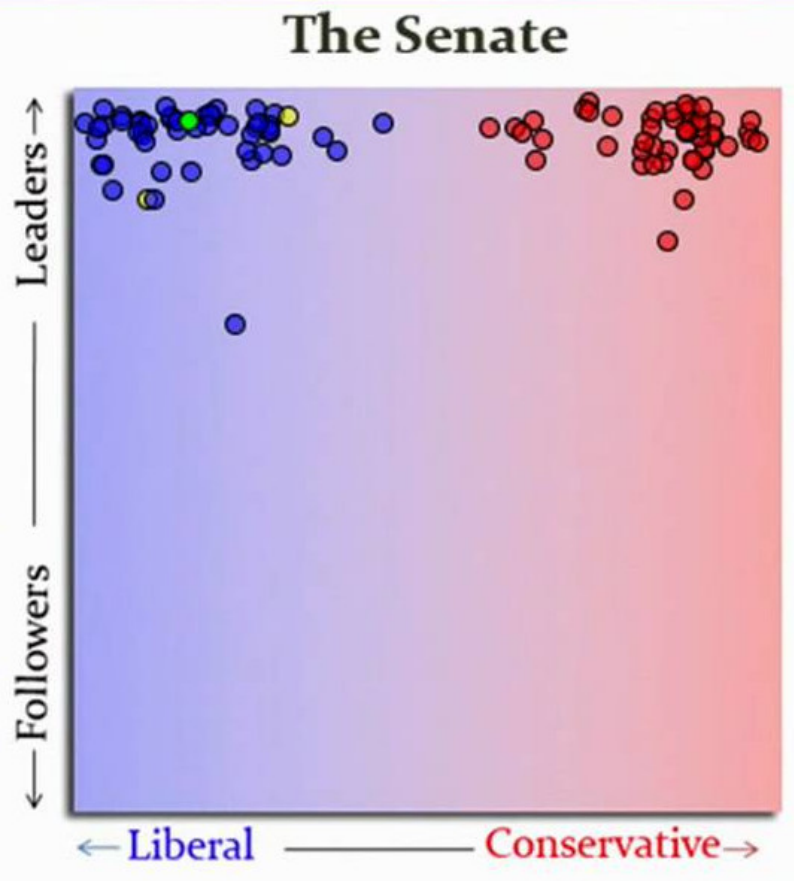
Source: (CELEBRITYTYPES, 2015)

Figure 3.1: Political spectra generated by questionnaires: the political variation of historical personalities are represented in two dimensional scatterplots. Both spectra use the most long-standing political dimension of the modern late period, left-right wing.

Connect 2 Congress (C2C) (KINNAIRD; ROMERO; ABOWD, 2010) is an analytic web application which creates a two-dimensional political spectrum (through NOMINATE scores and Leaders-Followers) of the U.S. Congress for arbitrary time frames within two years (2007-2008) (Figure 3.2). Roll calls can be inspected as tabular data and representatives can be filtered and highlighted (by name, state, party, religion and gender). The time frame can be dynamically modified, resulting in an animation where representatives are continuously repositioned in the spectrum according to their behavior. C2C does not have inter-legislature analysis as it comprises only two years. Along the political spectrum of representatives C2C aimed at providing a feed of relevant news so the user could keep track of votes and political debate. C2C is not updated since 2010.

For the Brazilian Congress, Marino (MARINO, 2014), Basômetro (PAULO, 2013) and Radar Parlamentar (RADAR..., 2015) created two-dimensional spectra of deputies. Marino and Basômetro display the political spectrum of Deputies as a scatterplot, where the diagonal is according to government coalition and opposition, in a way similar to Rosenthal's analysis of french legislatures (anti/pro-regime axis) (ROSENTHAL; VOETEN, 2004) using the Optimal Classification method (refer to Figure 2.1). Marino's result can be observed in Figure 3.3.

In Radar Parlamentar, the spectrum is displayed in a radial format, and only distances can be analyzed since the positions of parties change significantly through short periods of time. Marino and Radar Parlamentar are based on PCA analysis. Marino's work is a video, Radar Parlamentar is a web application that provides a



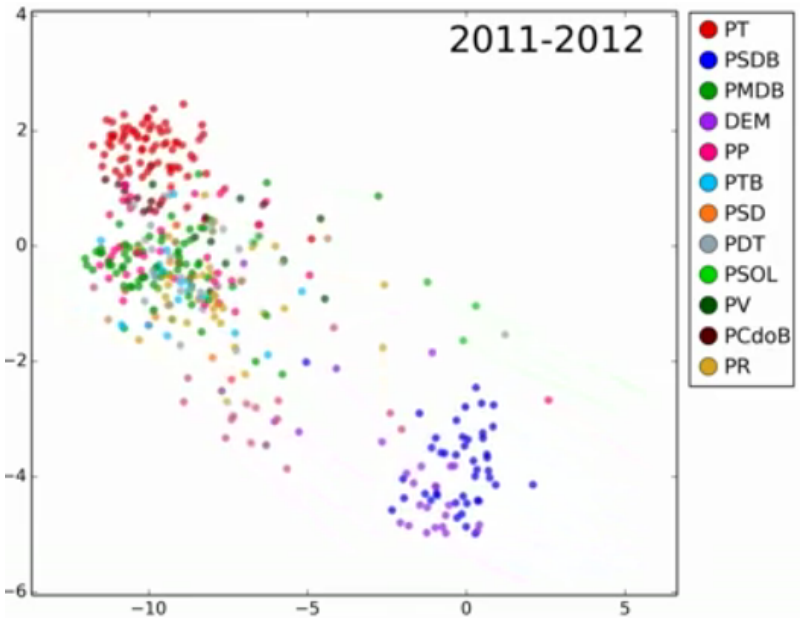
Source: (KINNAIRD; ROMERO; ABOWD, 2010)

Figure 3.2: Connect to Congress depicting the U.S. Senate in the period of 2007-2008. The Liberal-Conservative axis represents the variation of NOMINATE values of U.S. senators, the Democrats-Republican division. Followers-Leaders represents the correlation of bill support and authorship.

predefined sliding window of time (Figure 3.4). Basometro, from the Estadão newspaper, uses two criteria of "government support", and implements deputy tracking animation by a growing time window, similar do C2C.

3.2 Political Networks

Social networks are used in many fields of science; anthropology, biology, communication studies, economics, geography, information science, organizational studies, social psychology, sociology, and sociolinguistics. A social network is a social structure made up of a set of social actors (such as individuals or organizations) and a set of the ties between these actors. In politics, social networks analysis are applied to describe political actors, such as: congressman, congress committees, political issues, etc; and its relationships based on roll call votes, speeches, committees seats,



Source: (MARINO, 2014)

Figure 3.3: Screenshot of Marino's video representing the Deputies of the Brazilian Chamber in the period of 2011-2012. The axes are components from PCA.

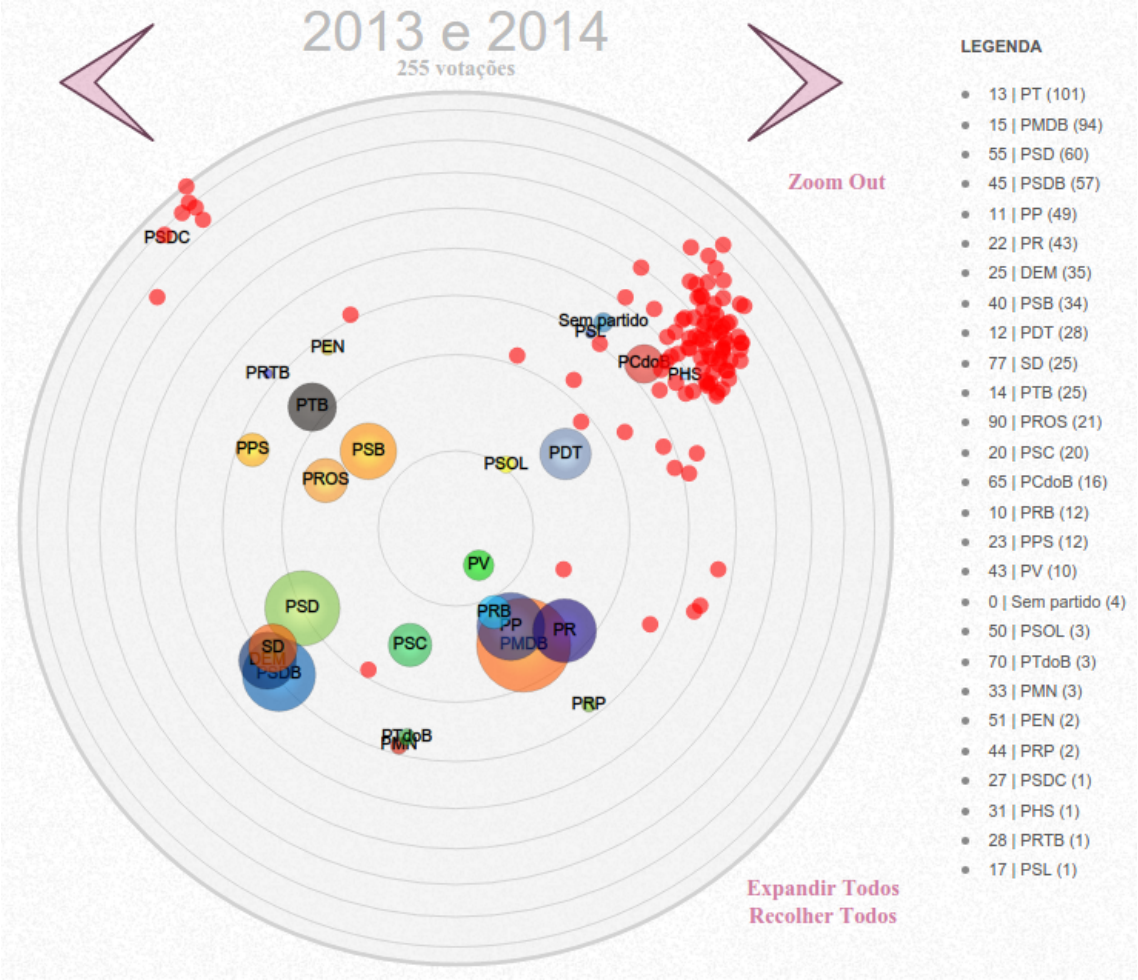
etc.

Social Action (PERER; SHNEIDERMAN, 2009) takes an alternative approach to spectrum visualization, representing the correlation of votes between U.S. senators as a social network displayed using force-directed layout algorithm. Social Action also lets users interactively apply filters and statistical tools to uncover patterns of voting groups at single points in time.

Fisher et al. (FISHER; LEIFELD; IWAKI, 2013) made a series of studies on the U.S. Congress hearings of climate changes politics. Transcripts of speeches were analyzed through labeling, clustering and network techniques, employing the Discourse Network Analyzer (LEIFELD; HAUNSS, 2012). The information is displayed as trees and networks, speakers being depicted as actors in the networks (see Figure 3.6). It is a newsworthy political analysis paradigm, since it relate as equals in the data analysis members of the Congress, scientists, environmentalists and business actors, all of them political actors in the climate change debate.

3.3 Political Timelines

To address the temporal variation of political positions in a single visualization, some techniques provide a long-term overview using political timelines, which are compact visualizations of political trajectories of individuals or parties over time.

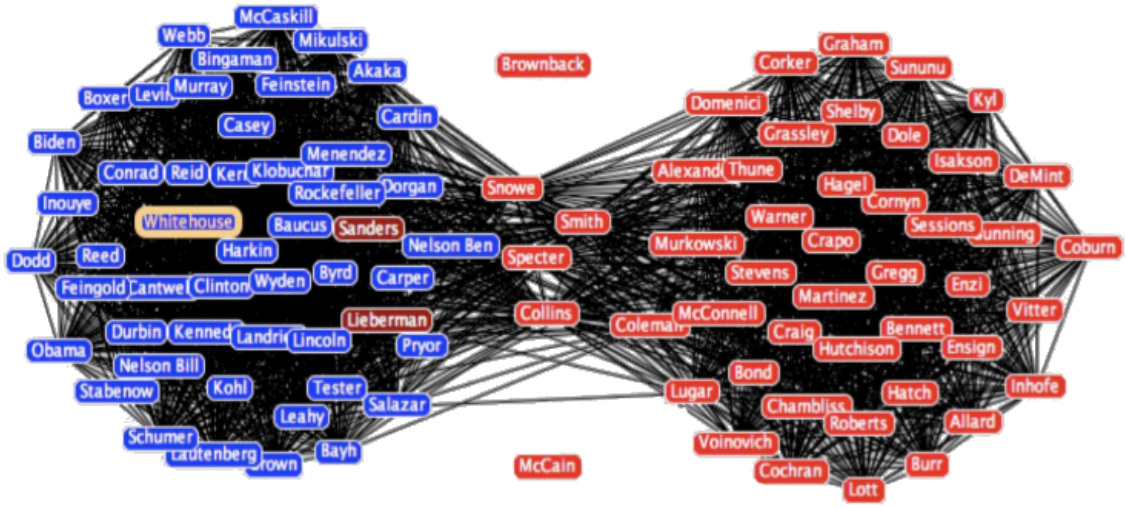


Source: (RADAR. . . , 2015)

Figure 3.4: Radar Parlamentar: the PCA radial representation of Chamber of Deputies political spectra for the 2013-2014 biennium. Parties are shown in clusters, the radius of the party circle is proportional to the the number of deputies and its position is the average position of deputies. Parties can be expanded to show the original positions of deputies, in the Figure the Worker's Party (*Partido dos Trabalhadores*, in red, is expanded.

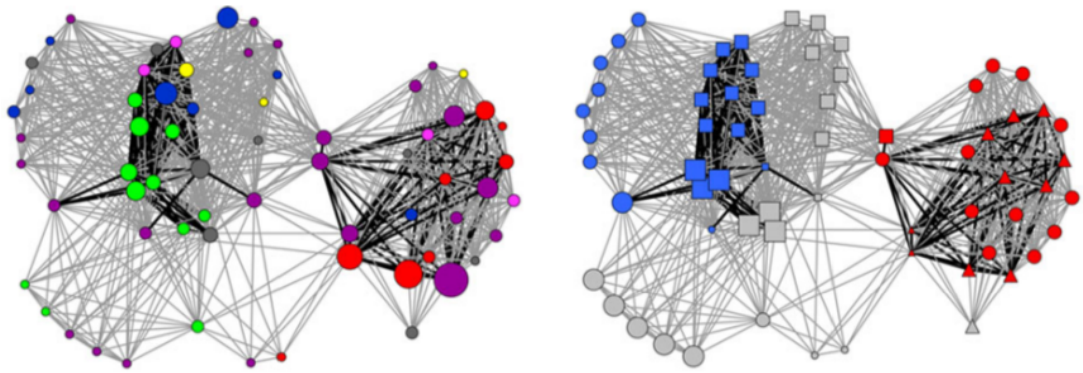
Political timelines are popular depicted as infocharts. An early form of political timeline is "A New Chart of History" (PRIESTLEY; FRANKLIN, 1769), were historical empires and cultures are represented in a time per territory axes. Later, we find the "Conspectus of the History of Political Parties and the Federal Governmet" (HOUGHTON, 1880) (Figure 3.7), which shows the U.S. parties political position as a flowchart in the period of ∼1700 to 1880, the x-axis representing the time and y-axis representing which party has the president seat. It also has annotations of historical facts and main seats in executive, legislative and judiciary.

The webcomic Xkcd (MUNROE, 2012) presented a large poster where DW-



Source: (PERER; SHNEIDERMAN, 2009)

Figure 3.5: Social Action: The social network of the U.S. Senators voting patterns in 2007. Republicans are colored red, Democrats blue and Independents maroon. The partisanship of Democrats and Republicans appear as two well separated clusters.



(a) Visualization of actor types. Blue indicates Democrats, Red indicates Republicans (of the 109 th Congress), Green indicates environmental groups, Purple indicates businesses and business and trade organizations, Yellow indicates scientists, and Gray indicates policy actors that fall into the "other" category. Dark and bold edges indicate agreement on more than one issue stance

(b) Visualization of issue stances. Political actors are labeled by its support in three stances. Actors are depicted with three styles variants - shape, color, and size - each representing a sentence. For example the stance "Legislation that regulates carbon dioxide emissions will not hurt the economy" label the color of political actors (blue: yes; red: no; gray: no statement available).

Source: (FISHER; LEIFELD; IWAKI, 2013)

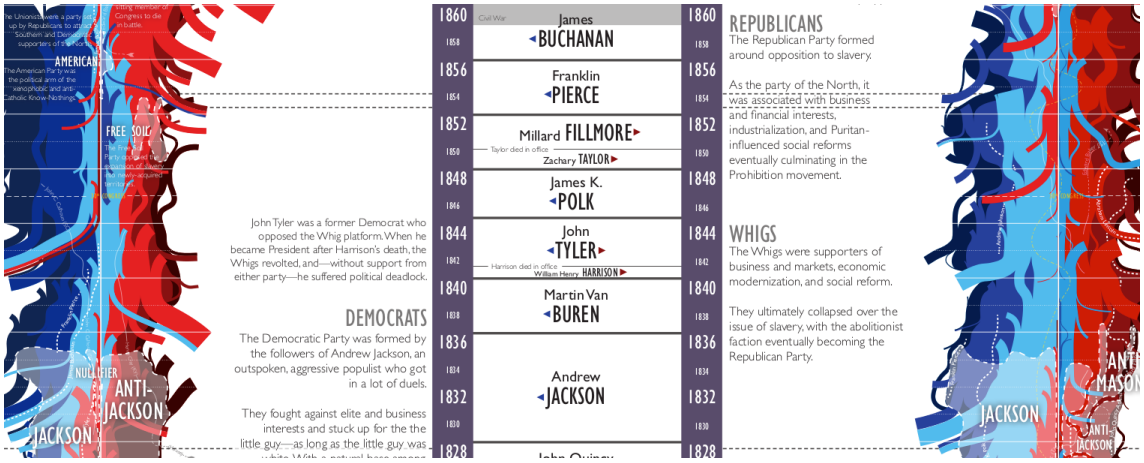
Figure 3.6: Mapping the ideological networks of American climate politics

Source: (HOUGHTON, 1880)

Figure 3.7: A fragment from 1829 to 1860 of the infochart Conspectus of the History of Political Parties and the Federal Government. In the first line, it depicts the president, followed by the ministers, judiciary presidents and legislative presidents. The parties are flows over time, y-axis aligned to the presidential government.

NOMINATE scores were used to create a political timeline of the U.S. Congress since 1788, low and upper house. Xkcd's visualization also shows annotations of historical events, presidents, control of the chamber, new and leaving members, and noteworthy members represented apart from their party. Position of parties in the x-axis are defined by the quantitative roll call analysis using DW-NOMINATE. Xkcd assigns only the color of the two major parties, red and blue, to boost the idea of U.S. two-party system. Minor parties are undefined clouds above the major parties. We can observe the differences of Xkcd's infochart (2012) in Figure 3.8, and Houghton (1880) in Figure 3.7, for the same historical period.

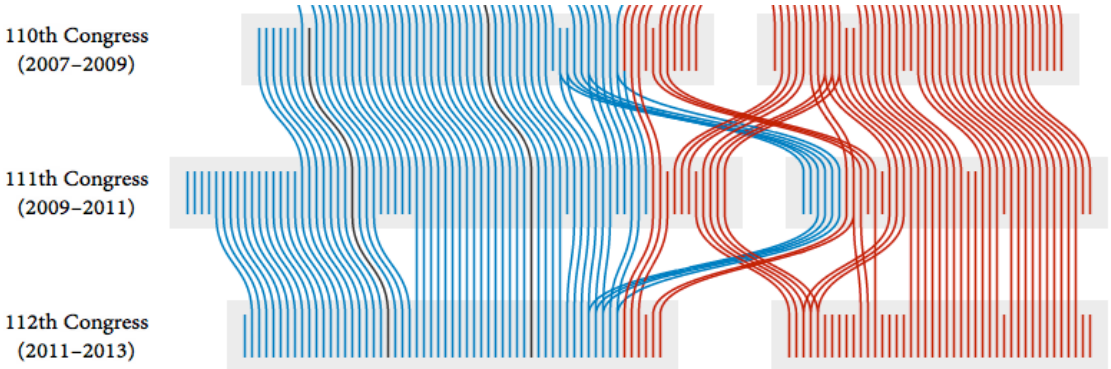
Friggeri and Fleury's visualization (FRIGGERI; FLEURY, 2012) shows the paths of U.S. senators through agreement groups for eight Congresses using a custom clustering algorithm applied to roll call data, with all Senators depicted. It is the only political timeline visualization that we found during our research based on a



Source: (MUNROE, 2012)

Figure 3.8: A fragment from 1829 to 1860 of The History of the United States Congress: Partisan and Ideological Makeup.

web-application, which explores user interactions, allowing the selection of a single senator to track his cluster alignment (Figure 3.9).

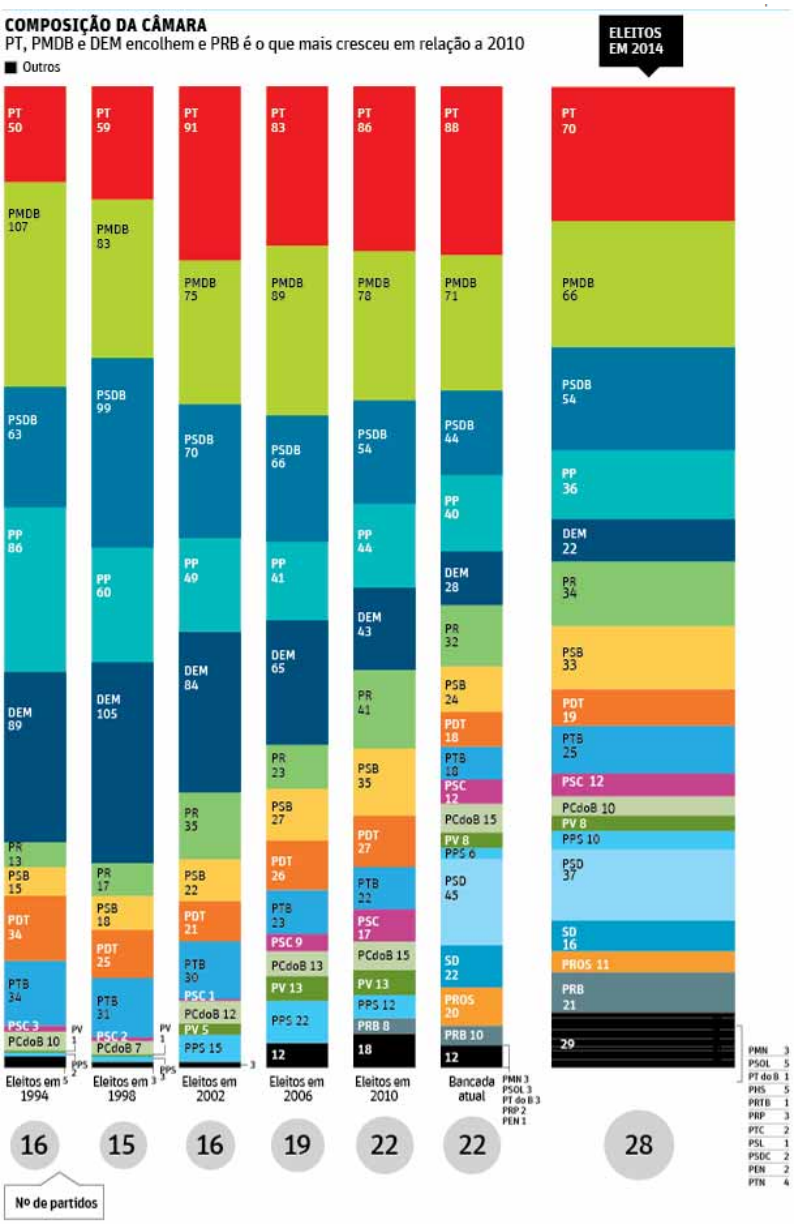


Source: (FRIGGERI; FLEURY, 2012)

Figure 3.9: A fragment of three legislatures of Agreement groups in the United States Senate. Each trace represents a Senator.

Moody *et al.* (MOODY; MUCHA, 2013) represented the U.S. Senate's political polarization using a dynamic coalition network. In each period, there are two "party loyalist" node-cluster positions, anchored on the y-axis proportional to a modularity score, representing the polarization. Other nodes, small clusters or individuals, are based on the balance of their votes relative to these anchors.

An article from the newspaper Folha de São Paulo (FOLHAPRESS, 2014) shows the proportion of parties seats in the Brazilian Chamber of Deputies (Figure 3.10). It shows six legislatures (1994-2018), each legislature is represented by a stacked bar



Source: (FOLHAPRESS, 2014)

Figure 3.10: Brazilian Chamber of Deputies Party composition. Each column represents one legislature, exception of the sixth representing the 54th legislature before the 2014 general elections.

chart, each column displaying the relationship of party seats (different column parts) to the whole, 513 seats of the Chamber of Deputies. To obtain a clear visualization parties who were renamed in the past are shown under the current party name; for example, PP and DEM were called PPB and PFL in 1994. The parties are sorted in the stack, partially, by their size in the last legislature.

3.4 Alternative Tools

Another representation of multiparty competition is the Nagayama Diagrams and Simplex Representations (GROFMAN et al., 2004) that use ternary plot, ratios of the three variables as positions in an equilateral triangle, all designed for three-party systems.

VoteEasy (REES; CITRARO, 2011) and MeRepresenta (MEREPRESENTA.ORG, 2015) assist the citizen by comparing questionnaires answered by users and representatives, indicating the politicians with similar answers, linking opinions of citizens with representatives. The Brazilian VoteNaWeb (WEBCITIZEN, 2009) offers summaries of proposed laws under consideration in the legislature. The users can record their votes and compare them with votes made by representatives in the system. VoteNaWeb, VoteEasy and MeRepresenta do not reflect roll call votes made in the legislature.

3.5 Final Comments

C2C is the main reference to CivisAnalysis. It is an analytic web application with which users can explore arbitrary periods of time in the political spectra based on roll call analysis. C2C stores the run of W-NOMINATE for each selectable period of time, but this solution is too costly for large periods. C2C has only two years of data, and in CivisAnalysis we wanted to provide all the recorded roll calls in the Brazilian Chamber of Deputies along 24 years. In C2C 'future works' it is stated that "improving the speed and efficiency of our analytic algorithms could allow real-time computation, enabling both simple categorical analysis (how does my Senator score on Gun Control?)...". In CivisAnalysis we decided to implement a light-weight roll call analysis, enabling real-time computation of any period of time in 24 years and any set of roll calls.

The political spectra created by dimensionality reduction methods (DRM) are snapshots of the interaction of political actors in a selected period. The political spectrum of a large period is a coarse representation of the political context of smaller periods, i.e.,how each smaller period started and developed is hidden, the representation showing only the average context in the period. To show small changes along

time, political spectra tools as C2C, Basômetro and Radar Parlamentar use animations. Political timelines can be best suited to represent these variations since it shows small snapshots side by side.

CivisAnalysis idea was to integrate the analytical tool of inspection of roll calls with a political timeline, fully interactive, allowing users to inspect the roll calls and major events which defined the political variations along the legislatures.

4 CIVISANALYSIS

In this chapter we present the design and implementation of CivisAnalysis, a tool aimed at helping citizens to explore the data from the Brazilian Chamber of Deputies's open web services (www.camara.gov.br). The software is named from Latin *civis* (citizen) and *analysis*. CivisAnalysis is an open source project (https://github.com/fgborja/CivisAnalysis) with extensive use of D3.js library.

4.1 Motivating User's Tasks

The design of CivisAnalysis was driven by questions that we understand common citizens, interested in politics, usually ask about the behavior of deputies and results from roll calls. These questions can be interpreted as citizen tasks supported by our tool. They imply analysis of roll calls, analysis of voting patterns of deputies and parties, and more complex questions about legislative behavior along time. The citizen tasks (CTs) are:

1. What were the roll call's results?
   (a) How did the deputies vote?
   (b) How did the parties vote?
   (c) How did the districts vote?
   (d) Was there a dispute between government and opposition?
2. Which deputies voted alike?
3. Which parties voted alike?
4. Which deputy is politically divergent from his or her party?
5. How can we identify roll calls with similar disputes?
6. In which roll calls a party's members did not vote alike?

7. Which parties make up electoral alliances? And coalition governments?

8. How did parties and deputies behave across different legislatures?

   (a) Is their behavior aligned to ideology?

   (b) Or is it aligned to government coalitions?

In CivisAnalysis, besides calculating and displaying the political spectrum of deputies for a given time interval, we also calculate and display the political spectrum of roll calls within the same time interval. The former shows deputies arranged by their similarity in voting behavior, i.e., deputies that give the same votes in the same roll calls are closer. The later shows the roll calls arranged according to their voting similarity, i.e, roll calls that received the same pattern of votes by the deputies are closer. Tasks CT1 to CT6 can be supported by scatterplots (Figure 4.1) showing two political spectra, one for deputies and other for roll calls.

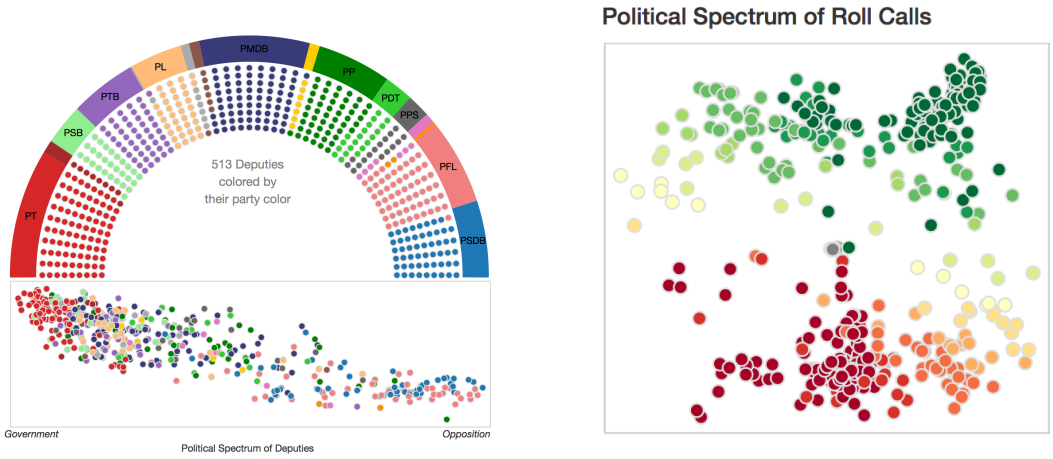


Figure 4.1: Political spectra of deputies (left) and roll calls (right) in the period of the 52 legislature, from 01/01/2003 to 01/01/2007. The political spectrum of deputies shows each deputy as a circle, and is aligned with the infographic that shows the composition of the Chamber in the same period. One can observe the alignment of deputies with government or opposition depending on their parties. The political spectrum of roll calls shows each roll call as a circle in that period. The circle is colored according to the result of the roll call: green=approved, red=not approved, shades of yellow meaning dispute.

For supporting tasks CT7 and CT8, CivisAnalysis provides a timeline (Figures 4.2 and 4.3) that shows the flow of parties approaching government or opposition along time. Different color palettes allow to distinguish parties (Figure 4.2) and ideology (Figure 4.3), as we will discuss later.

In the next sections we describe our tool using the citizen tasks as arguments in the discussion of each visualization component.

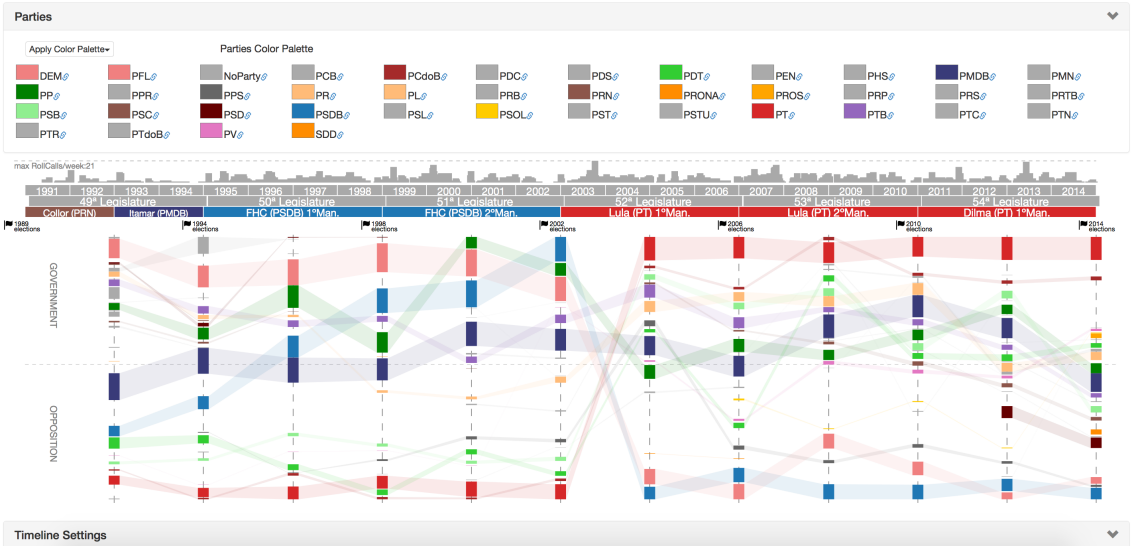


Figure 4.2: CivisAnalysis's Summary visualization, showing the political spectrum of parties as a flow along 24 years and six legislatures. Parties are colored according to the *Parties Color Palette*, where each party has a different color, parties in gray are the ones that do not exist anymore.

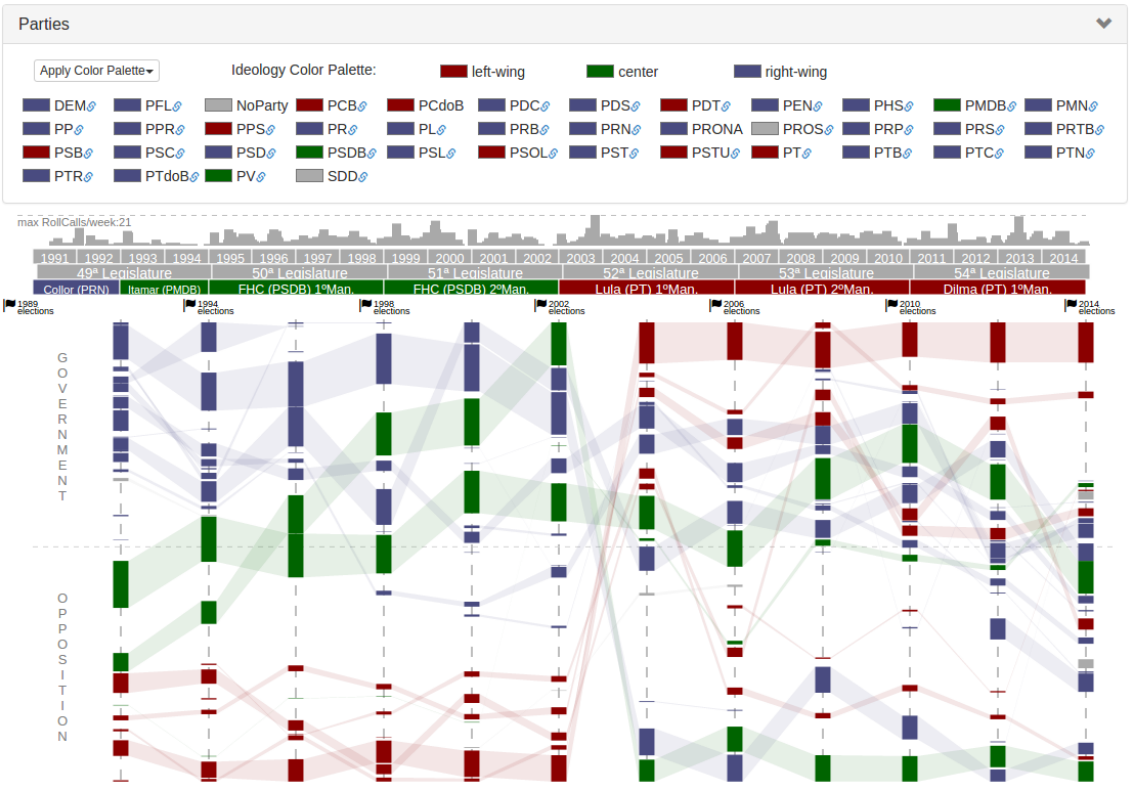


Figure 4.3: CivisAnalysis's Summary visualization, showing the political spectrum of parties as a flow along 24 years and six legislatures. Parties are colored according to the *Ideology Color Palette*. The 49th legislature—the first elected by direct vote after the military regime—is still marked by the polarization of left-wing and right-wing parties (socialism versus capitalism). The polarization weakens along the legislatures due to the need of establishing a coalition government.

4.2 Political Spectrum Computation

Before giving details of the visualizations, we explain how the political spectrum is created and how we measure the political variation of deputies and roll calls.

The spectrum is relative to a period in time, its start and end dates are provided by users as parameters through selection in the timeline (Figure 4.2). Given a time period, we find N roll calls voted by M deputies and build the political spectra of both deputies and roll calls, which are shown separately in space (Figure 4.1) but logically linked for interaction, as we will discuss later in this chapter.

For both deputies and roll calls the spectrum is created using the same algorithm, PCA by Singular Value Decomposition (SVD) (GOLUB; REINSCH, 1970). Despite political scientists' preference for NOMINATE analysis, PCA analysis has advantages: it is non-parametric, deterministic, fast, and allows real-time computation on the client side.

The use of PCA to create the political spectra in CivisAnalysis is not a requirement. Our visualization techniques can be applied to NOMINATE spectra or other dimensionality reduction methods (DRM). But, since we are working with data of political content we need to maximize the unbiased value of the spectra. For this objective we suggest the following criteria: (1) choose a non-parametric DRM, (2) prefer a deterministic DRM, and (3) use widely known/referenced algorithms. PCA is a general purpose DRM that fits to these criteria.

Modern DRMs like t-SNE (t-distributed stochastic neighbor embedding) (MAATEN; HINTON, 2008) or LSP (Least square projection) (PAULOVICH et al., 2008) use random samples and local optimization making them non-deterministic, the visualization creator or user would have to reproduce the method multiple times and choose the "best" result.

In addition, CivisAnalysis implements political spectrum computation by t-SNE as an alternative to PCA: this feature is explained in Section 4.5.7. But, since t-SNE optimizes local similarities, the political spectra of deputies and roll calls lose their direct relation. The deputies spectrum will represent the local similarities of deputies, and the roll calls spectrum the local similarities of roll calls.

4.2.1 Deputy Set and Recorded Votes Matrix

The first step for creating the political spectra is to build the deputy set and corresponding votes in the selected time period. Elected deputies that are called for positions in the Executive or that run for municipal elections are not part of the legislature for the entirety of its term. As such, the number of deputies participating in roll calls that took place during the selected periods is larger than the expected

513 seats of the Chamber of Deputies. To approximate the best representation of the Brazilian Chamber of Deputies during the period we select only the 513 deputies who were most present in the roll calls. This way, negligent deputies will be excluded, resulting in a slightly different number of seats than the elected by parties and states.

The second step to compute the political spectrum is to build a matrix of recorded votes. Consider the matrix $\boldsymbol{A}$ *(M x N)*, with $M = 513$ deputies and N roll calls and where each cell $\boldsymbol{X}$*(m,n)* represents the vote of the *mth* deputy in the *nth* roll call. The value of each vote is defined as: 1 for *Yes*, -1 for *No*. In case of missing data, vote abstention or obstruction, the cell's value is set to 0. There are different techniques to fill the missing data. It is possible to infer the probability of vote in the motion, analyzing the historical data and similar deputies. The obstruction usually are attempts to block the roll call by lack of quorum, against the majority of votes or it is just a call for further discussion in plenary.

4.2.2 Computing the Political Spectra

We apply the SVD algorithm on the recorded votes matrix $\boldsymbol{A}$, obtaining the following matrices:

$$\underbrace{A}_{513\times N} = \underbrace{U}_{513\times 513} \times \underbrace{\Sigma}_{513\times N} \times \underbrace{V^T}_{N\times N} \tag{4.1}$$

To reduce the votes matrix into a two-dimensional spectrum we set the smallest of the singular values, found in Σ, to zero. If we set the n smallest singular values to 0, then we can also eliminate the corresponding n columns of $\boldsymbol{U}$ and $\boldsymbol{V}$. We obtain following composition of a 2-rank reduction:

$$\underbrace{Ur}_{513\times 2} \times \underbrace{\Sigma r}_{2\times 2} \times \underbrace{V^T r}_{2\times N} \tag{4.2}$$

The two-dimensional spectrum of deputies *(513 x 2)* is calculated by multiplying the two largest singular values found in Σr by the left-singular vectors of $\boldsymbol{Ur}$. Figure 4.1 shows on the left the political spectrum of deputies.

The two-dimensional spectrum of roll calls *(N x 2)* is calculated by multiplying the two largest singular values found in Σr by the right-singular vectors of $\boldsymbol{Vr}$. Figure 4.1 shows on the right the political spectrum of roll calls.

4.2.3 Implementation Decisions

In a period of time, a deputy can assume more than one party or district. Then, we set deputy's party and district by their last vote in the period. As mentioned before, when we have more than 513 voting deputies in the period we select only

the first 513 deputies more present in the roll calls. The chamber has 513 deputies, the surplus being substitute deputies. We remove them because deputies with few votes gave the scaled visualization a focus on the outliers, and there would be no correlation of votes with the majority.

NOMINATE computation keeps track of parties to maintain a consisten view in different periods. PCA is deterministic for each period dataset; however, for each different period, parties can assume different quadrants in the spectrum. The cause is not an abrupt political change, since one can observe that distances between deputies in the spectrum stay almost unchanged across short periods. The cause is reflections in the axes, which are expected to occur since axes represent the principal components, which encode the variation in the data [+1 Yes, -1 No]. For visualization purposes, such reflections are not desirable because they demand more cognitive effort from the user to analyze patterns, track deputies and parties. To keep the axis coherent along the legislatures, we labeled the first component as *Government versus Opposition*, and choose to force this point of view, since these political forces are present in all legislatures and represent the major cause of variance in the data. When necessary, a reflection transformation is applied to both spectra (principal component multiplied by -1) to keep the *Government versus Opposition* view. Note that the scale proportion and distances between deputies and roll calls are preserved. The result can be observed in Fig. 4.2, where the government coalition parties keep their position at the top of the spectrum and opposition parties at the bottom.

4.3 Overview of CivisAnalysis

CivisAnalysis is an open source project implemented as a web application (JavaScript) capable of running in any modern browser. All roll call analysis techniques and spectrum creation, search and filtering functions run on the client side, so the web server is free from CPU/memory intensive processes. The roll calls are loaded on demand using XMLHttpRequest. A user's inspection of all legislatures requires the client's browser to load the entire data set (currently around 40MB). CivisAnalysis is scalable, not demanding powerful web servers or too much bandwidth.

The first screen that users have access in CivisAnalysis is the *Summary Module* (Fig. 4.2), which presents a summary of 24 years of the Brazilian Chamber of Deputies's history, displaying in chronological order presidents, legislatures, general elections and the placements of parties in the spectrum dimension government-opposition along time. To obtain details on roll calls and deputies, users can select a period of time. When a period is selected, the *Inspection Module* is displayed.

The *Inspection Module* has multiple views of the roll call data set (Fig. 4.4). In the *Chamber infographic* (Fig. 4.4.C), users can inspect the deputies sorted by

parties. Deputies can be analyzed individually in the spectrum of deputies (Fig. 4.4.D). The spectrum of roll calls shows how roll calls relate to one another (Fig. 4.4.F). To represent the geographic properties of the data set, we display electoral districts on a map (Fig. 4.4.B), with users being able to click on a district to select it. Since roll calls, deputies, parties and districts are strongly bonded, the *Inspection Module* uses techniques of coordinated views extensively. User actions in one view are thus reflected in all the others. The user has also some options for interacting with deputies (Fig. 4.4.A) and roll calls (Fig.r̃effig:inspectionModule.E, Fig.r̃effig:inspectionModule.G).

4.4 Summary Module

Following the approach of *overview first, and details on demand* (SHNEIDERMAN, 1996), the application starts with a visualization that is basically a timeline of the basic information about the Brazilian Chamber of Deputies's history. The goal is to give users an overview of how parties behaved in the Chamber. It aims at a quantitative reasoning of *CT* 8a and *CT* 8b, i.e. the political behavior along consecutive legislatures.

Flag icons to query general elections are displayed below the respective election date. When selected, the leading four candidates are displayed associated with their electoral alliances, each deputy and party is colored according to the leading party of the electoral alliance it belongs.

To obtain specific details about a period, users can select arbitrary time frames on the timeline by clicking on specific presets (e.g., years or legislatures) or by using the brushing tool. Figure 4.4 shows the selection of the period from January of 2005 to December of 2006.

4.4.1 Party Timeline

The Party Timeline displays the political spectrum of parties as rectangular elements anchored on the y-axis for each two-year period. The biennial timeline was chosen because each two years Brazil has elections - so each step represents the political spectrum between elections. The party's position in the spectrum is related to the average of the position of the party's deputies in the political spectrum of deputies. The height of the party is relative to the number of deputies of the party in each period.

Since many parties have close roll call policies, the elements can be cluttered on the y-axis. To prevent cluttering, we included a simple *Unclutter* function, where parties' sizes and political distances can be perceived without any obstruction. This

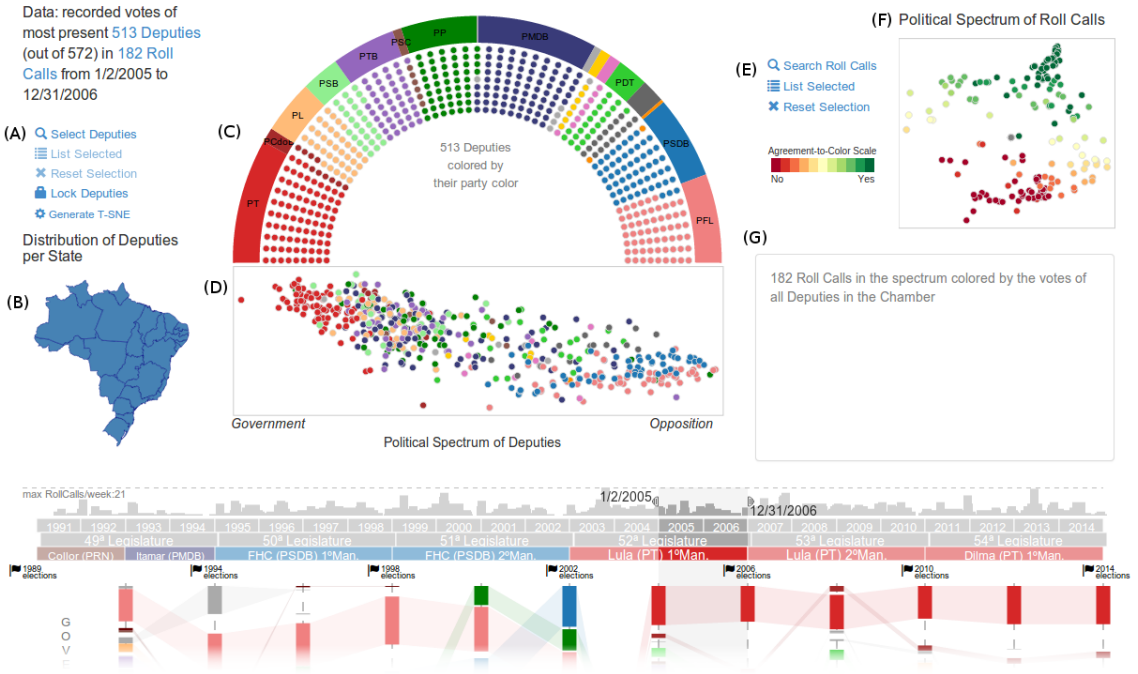


Figure 4.4: CivisAnalysis's Inspection module, depicting the Brazilian Chamber of Deputies in 2005-2006 (selected in the timeline of legislatures at the bottom). (A) is an advanced menu for search, filter and lock the selected deputies, (B) the Brazilian map shows the pattern of votes by federal districts, (C) shows the distribution of deputies per party, sorted by deputy positions in (D), which depicts the political spectrum of deputies. On the right, (E) is the advanced menu to search and filter the roll calls depicted in (F), the political spectrum of roll calls. (G) is a text window which shows information about the selected roll call. At the bottom, part of the timeline of the Summary Module.

function sorts the parties according to their spectrum positions, and uses a force-directed algorithm to arrange the parties in the y-axis closer to the original spectrum position. The height of the elements and the distances between them are calculated according to the percentage of pixels the users want to use to represent the 'political distance' or 'party size'.

The application provides a configuration settings menu below the timeline (Fig. 4.5) to allow users to change this percentage or display the parties' original positions in the spectrum. Figure 4.3 was generated with the use of the *Unclutter* function, with parties' size filling 45% of the y-axis. Figure 4.6 shows the trade-offs between showing the political size, number of deputies (votes), and the political distance in the spectrum, (dis)similarity of votes.

4.4.2 Parties Color Palette

Party colors are fully customizable. When first loaded, CivisAnalysis displays the summary module using the *Diff Major Parties* palette, which is an arbitrary selection of colors that makes the main parties visually distinguishable from each other (Fig. 4.2). The other option is the *Ideology Color Palette*, which categorizes the parties in left-wing, center and right-wing parties, as shown in Fig. 4.3, to assist users reasoning in *CT* 8a and *CT* 8b.

4.5 Inspection Module

The *Inspection Module* consists of different views of the roll call data set for a selected time period (Fig. 4.4). The views display important variables that should be considered in the roll call analysis: parties, deputies, roll calls (votes) and electoral districts. These variables relate to one another: a party is represented by a set of deputies; a district is represented by a set of deputies; a deputy has a set of roll call votes; a roll call has a set of deputy votes.

To reveal these relations, the views are coordinated in order to express subsets of the data as demanded by users. User actions in one view affect all the other

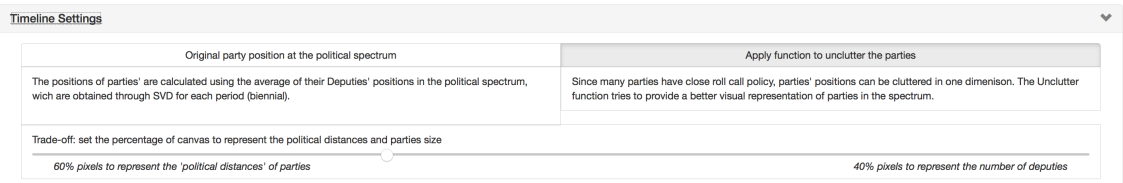
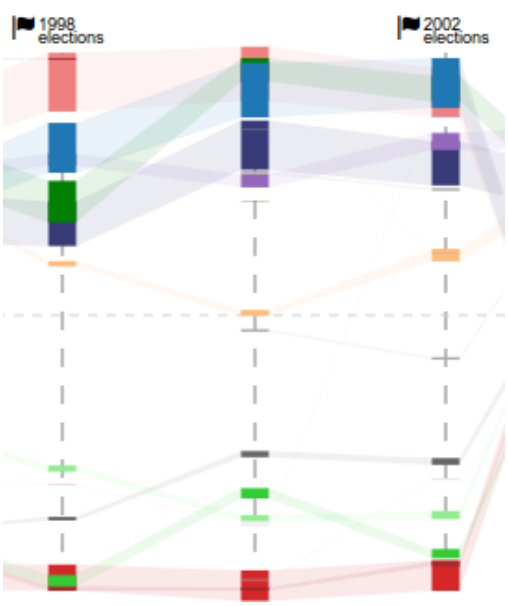
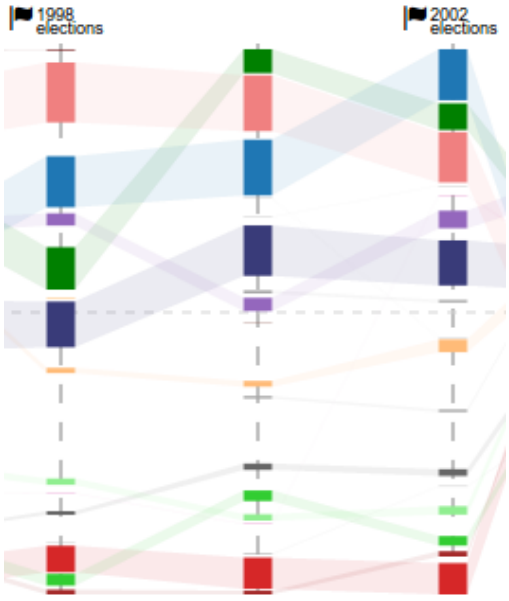


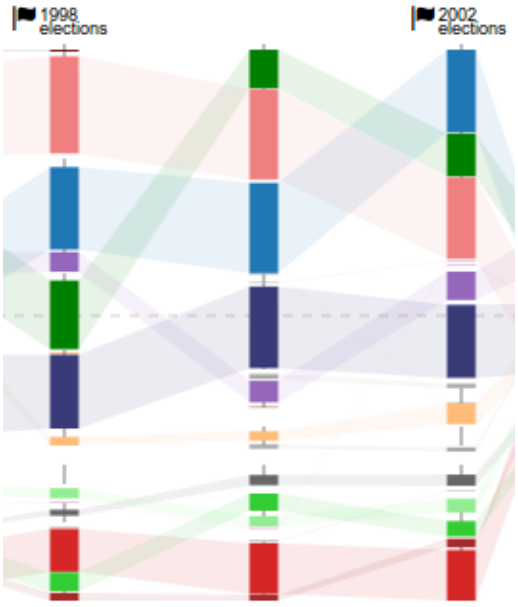
Figure 4.5: Timeline settings menu allowing the user to unclutter the parties positions in the timeline.



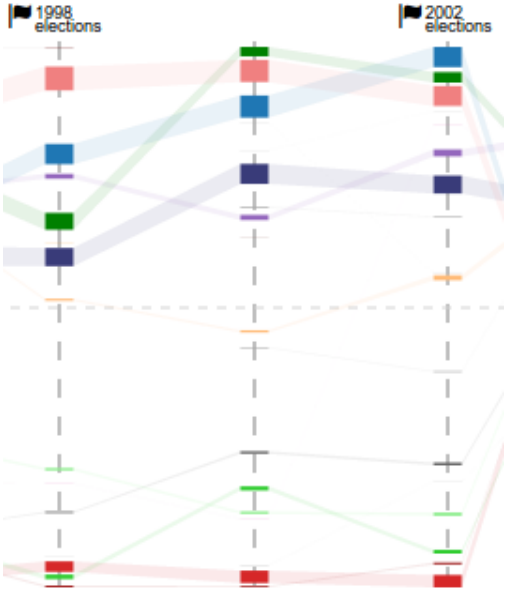
(a) Original parties' position in the spectrum, parties are cluttered in their coalitions.



(b) Unclutter function, using the y-Axis to represent 50% the parties' size and 50% the political distances.



(c) Unclutter function, using 80% of y-Axis to show the parties' sizes.



(d) Unclutter function, using 80% of y-Axis to show the parties' political distances.

Figure 4.6: Political spectrum of the 1997-2002 period. Each y-Axis represents the spectrum of the legislature biennial. The top squares represent the Government-loyal parties and, at bottom, the opposition. The images show how the user can use the trade-off between cluttered or uncluttered and party size vs parties' political distances to analyze the timeline. The larger the space to represent their number of votes, smaller is the space used to represent their political dissimilarities.

views. For example, selecting a single roll call in the political spectrum of roll calls scatterplot will highlight the data of this roll call in all the other views: deputies in

the political spectrum of deputies will have their colors changed to represent their vote in the roll call based on a *Vote-to-Color* map; parties and districts will have their color changed to represent the votes made by their respective set of deputies using an *Agreement-to-Color* scale (see Fig. 4.8, for an example).

The **Agreement-to-Color Scale** is a function that maps the vote agreement scale [(100% "Yes" votes) to (100% "No" votes)] to a color scale [green to red]. For example, a roll call with only "No" votes will be displayed in red; a roll call with 70% "Yes" votes and 30% "No" votes will be light green—the roll call had a moderate agreement favorable to "Yes."

The **Vote-To-Color Map** is a function that maps a single vote ["Yes", "No", "Obstruction", "Abstention", chamber president, absence] to a color [green, red, blue, purple, yellow, gray].

4.5.1 Chamber Infographic

To represent the role of parties and deputies in roll calls, supporting *CT* 1a and *CT* 1b, the Chamber Infographic displays the deputies and parties sorted by their position in the political spectrum. The parties are sorted according to the average position of their deputies in the spectrum of deputies. We adopted the semi-circle design because it is a common representation of assemblies, chambers or legislative houses. A half-donut chart is aligned with the chamber to represent the proportion of parties. Users can select party and deputy elements by clicking; hovering over them displays more information. Figure 4.4 depicts a selection of all the parties, while Figure 4.7 shows how divergent deputies can be identified using the Chamber Infographic, thus providing support for *CT* 4. If users select one or more roll calls (Fig. 4.8 and Fig. 4.9) in the roll calls spectrum, the color of the parties is changed to reflect how their deputies voted using the Agreement-to-Color scale.

4.5.2 Spectrum of Deputies

To analyze each deputy individually, the spectrum of deputies is presented as a scatterplot. Each deputy is represented as a circle positioned according to PCA. The more similar deputies vote, the closer their respective circles are, helping support *CT* 2. Users can select deputies by clicking or brushing, and can get more information by hovering over them with the cursor. By default, deputies are drawn in their respective party's color. When a single roll call is selected, deputies colors change—according to the Vote-to-Color map (Fig. 4.8)—to represent their vote in the roll call. If a set of roll calls is selected, deputies colors change—according to the Agreement-to-Color scale (Fig. 4.9)—to reflect their votes in the roll calls. For example, if a

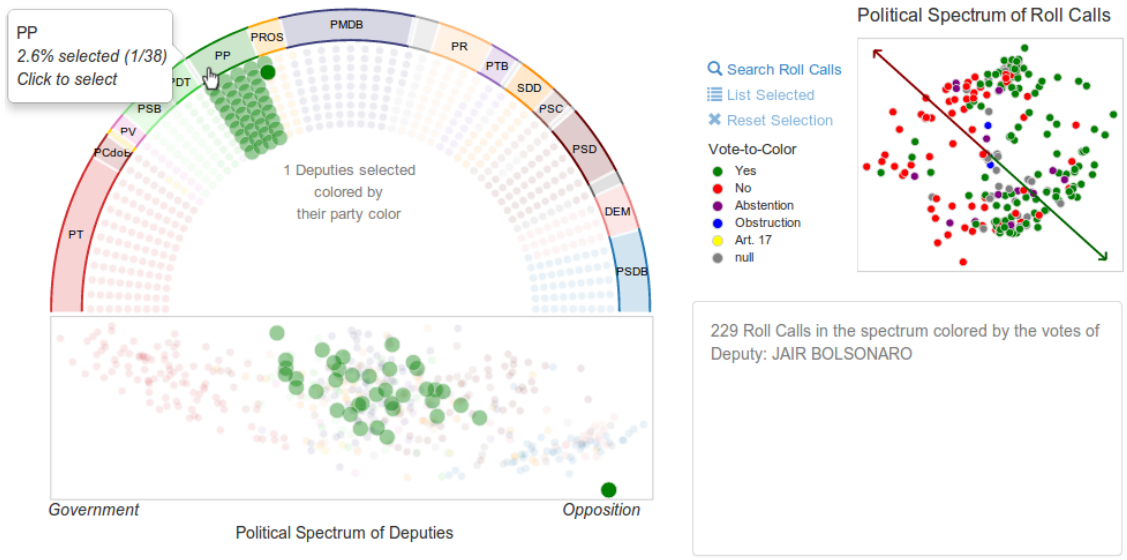


Figure 4.7: The cursor is hovering over the Progressist Party (PP), which has only one selected deputy, shown as the non-transparent circle: Jair Bolsonaro is an electoral success with radical political views. We can observe, in the spectrum of deputies, his distance from his party and his polarized position. The spectrum of roll calls represents the votes of the selected deputy using the Vote-to-Color map. Jair Bolsonaro's, cutting plane, positive and negative variance ("Yes"/"No" votes) is represented as a vector in the roll calls spectrum.

deputy voted "Yes" in 30% of the roll calls and "No" in 70%, he or she will be colored in light red since his/her votes in the selected roll calls are moderately favorable to "No". To reinforce the idea of coupling between the roll calls and deputies' spectra, arrows representing the vectors of positive and negative ("Yes"/"No" votes) variance are shown in the spectrum of roll calls (see Fig. 4.7).

4.5.3 Spectrum of Roll Calls

To let users directly inspect roll calls and make it easier for them to find similar roll calls, we designed the spectrum of roll calls as a scatterplot where each roll call is represented as a circle. Users can select roll calls by clicking or brushing and can hover over them with the cursor to display the amendments summary. The spectrum axes represent the same components as the ones in the spectrum of deputies, so the distribution of roll calls is closely related to the distribution of deputies in their spectrum. A short distance between roll calls expresses their voting pattern similarity, which is adequate for *CT* 5.

To increase the representativity of the spectrum, roll calls circles are colored using the Agreement-to-Color scale. Roll calls in dark red are those that received only "No" votes, while those in dark green got only "Yes" votes. In the transition

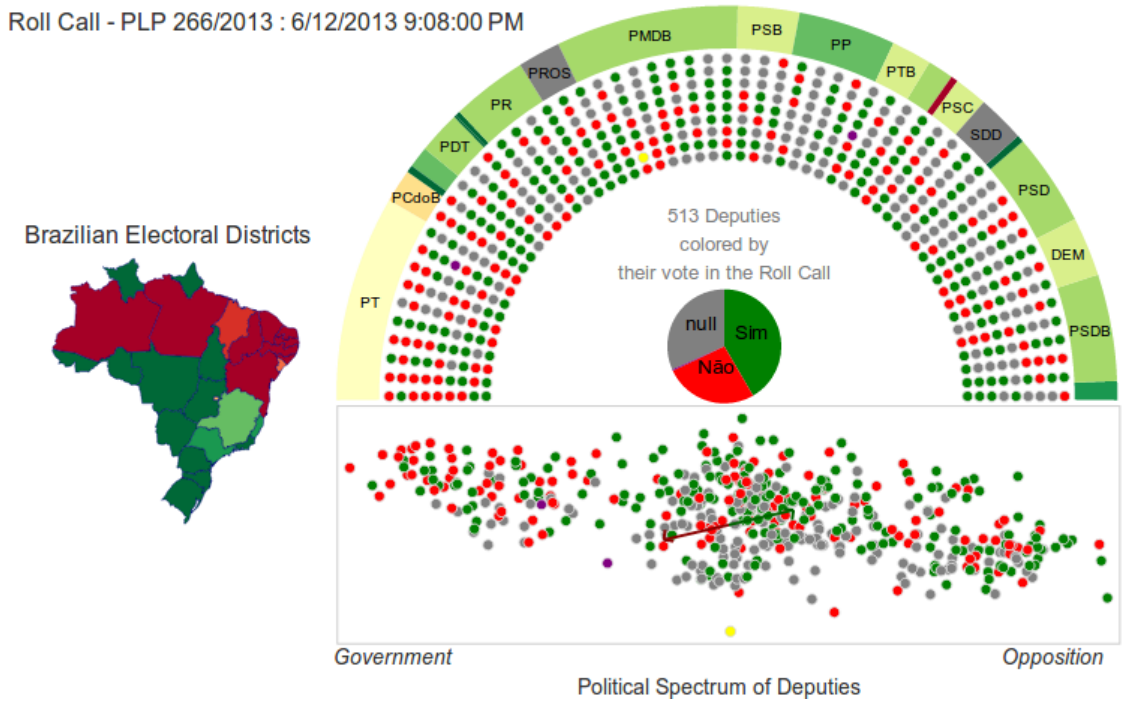


Figure 4.8: Selection of roll call PLP 288/2013, which comprised amendments to the Participation of States Fund law. The selection links the roll call votes to deputies, parties and districts. Deputies are colored using the Vote-to-Color map; parties and districts with the Agreement-to-Color scale. The central pie chart shows the proportion of votes. We can see the disagreement of parties that resulted of a dispute between northern and southern districts.

from red to green, it is possible to identify where impasses occurred: they are colored with a neutral color (light yellow). By default, roll call colors represent the votes of all the deputies of the Chamber (see Fig. 4.4), although users can select a subset. For example, by selecting one of the parties, the roll call colors will change to show only the votes of its members (see Fig. 4.10), thereby making explicit which roll calls the party's deputies had agreed or disagreed (*CT* 6). The same logic applies when selecting districts, in the map, or any arbitrary set of deputies.

4.5.4 Visual and Textual Filters

CivisAnalysis provides interactive filtering, letting users create selections and compare them using the coordinated views. Parties, states, deputies and roll calls can be selected by left-clicking: this operation will select only the corresponding elements (i.e., left-click on a state element select all deputies from that state) and deselect others. The left-click can be combined with the *Ctrl* key to add elements or with the *Shift* key to exclude elements from the current selection. Each filtering operation modifies the views on-the-fly: for example, selecting states and parties

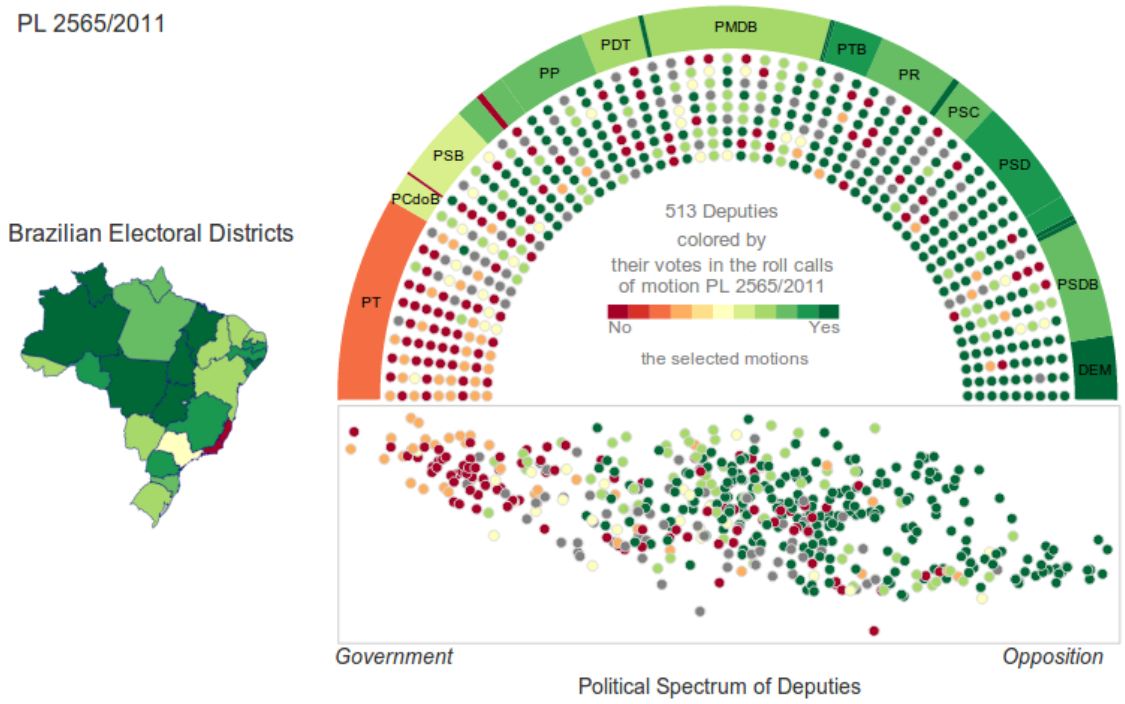
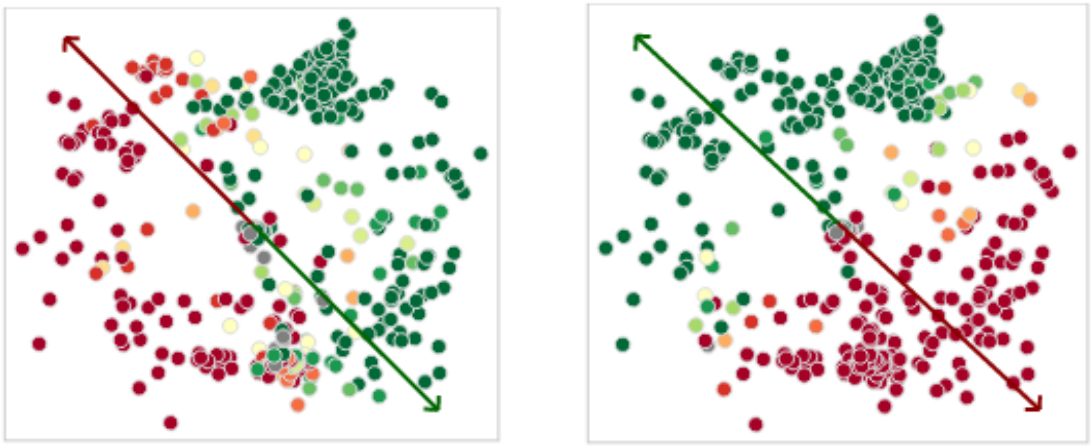


Figure 4.9: The proposed bill PL 2565/2011 aimed at an equal distribution of oil royalties to all the states and the use of these royalties for education only. The visualization depicts the votes of this bill's three roll calls. Reddish elements gave more "No" votes. We can notice at a glance that the dispute was between districts that are oil producers (in red), those that are not, and the government party (PT).



(a) Selection of PSDB

(b) Selection of PT

Figure 4.10: Antagonist voting patterns formed in the spectrum of roll calls after selection of opposition and government parties. The yellowish circles are roll calls where members of the same party disagreed.

causes only votes in which related deputies participated to be displayed on the spectrum of roll calls.

Deputies and roll calls can be searched for and filtered using textual keywords.

Deputies are indexed by name, party and state. Figure 4.11 shows the text box used for selecting one or more deputies by their names. Roll calls are indexed by the identifier, the motion summary and tags that are provided by the Brazilian Chamber of Deputies. Roll calls texts are in Brazilian Portuguese (Figure 4.12).

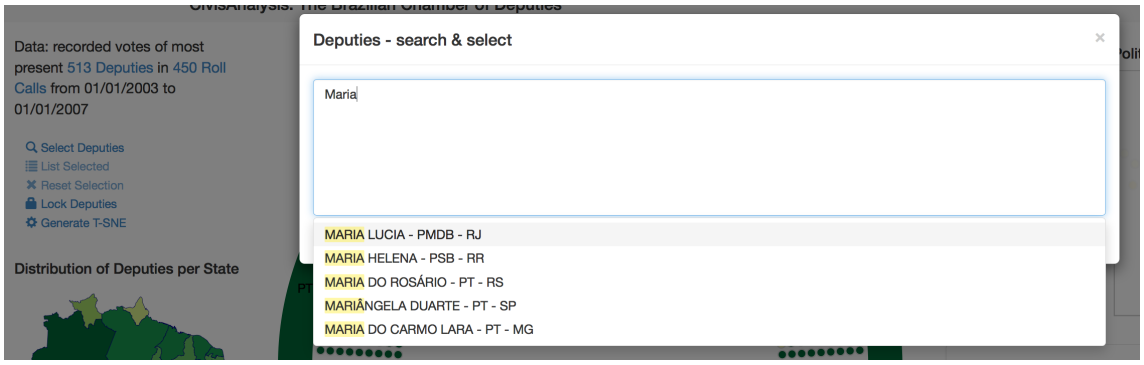


Figure 4.11: Selecting one or more deputies by their name. Clicking on "Select Deputies" on the options above the map causes the search box popping-up and the user can type part of the deputy's name, and select one or more from the list.

Figure 4.12: Filtering roll calls based on text. Clicking on "Select roll calls" on the options at the left of the political spectrum of roll calls causes the list of all roll calls popping-up, and the user can type any expression in the search ("Search") field, causing immediate filtering of roll calls that do not comply with the search criteria.

4.5.5 Toggle General Election Alliances

Electoral alliances in presidential elections are an important subject in order to analyze the legislatures. Before elections parties behavior in the legislature can be analyzed to answer how the electoral alliance was formed and after elections how the electoral alliances affected the next legislature. CivisAnalysis provide seven general elections for the user select. When one election is selected the parties are colored by the color of the president nominee's party in the alliance, see Figure 4.13. The Chamber Infographic sort the parties by its alliance and shows an alliance ribbon

(donut chart) over their parties. If the user click in the alliance ribbon only the deputies from the alliance is selected. Using this alliance, by selecting the spectrum of roll calls will represent only the votes of the deputies in the alliance, so, the user can identify the roll calls in which the alliance had agreement or disagreement.

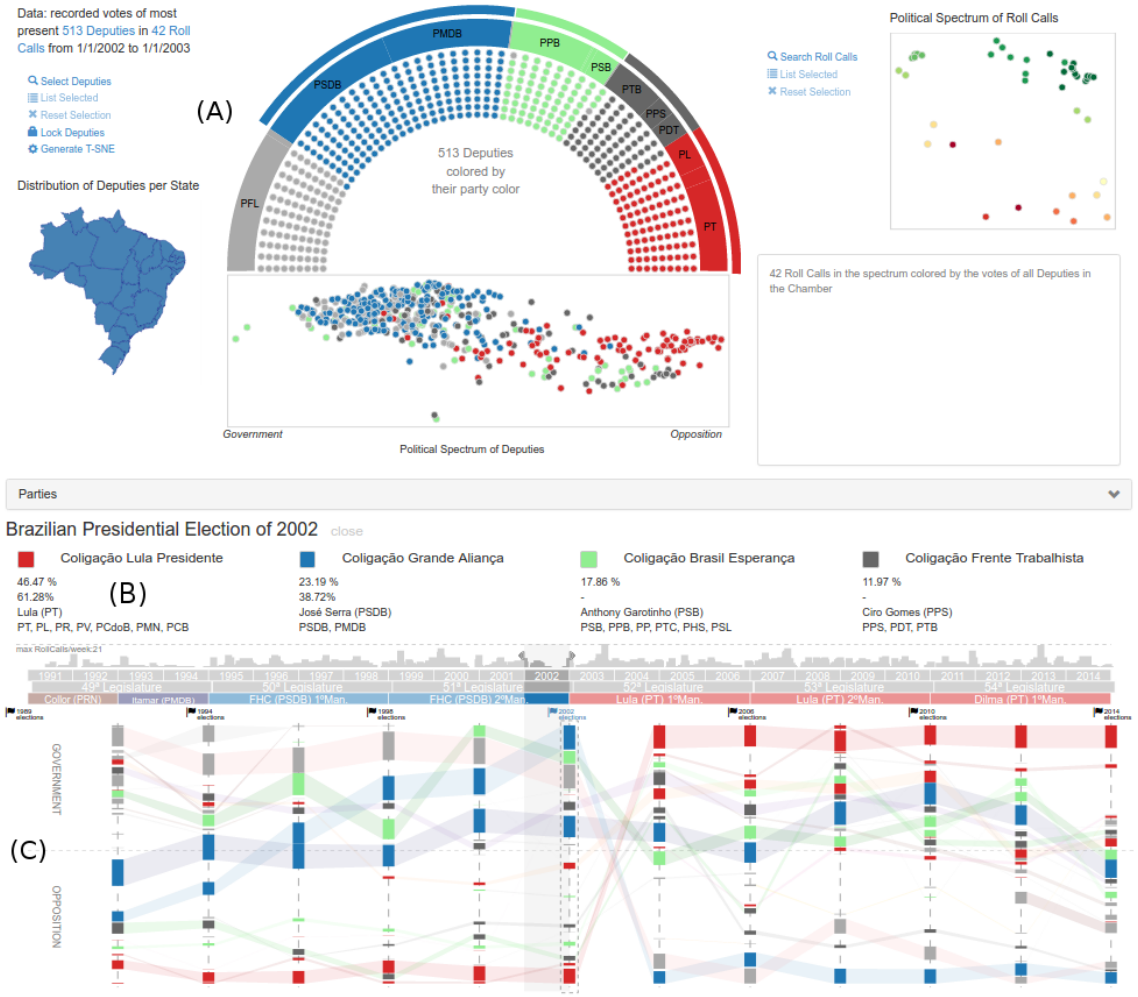


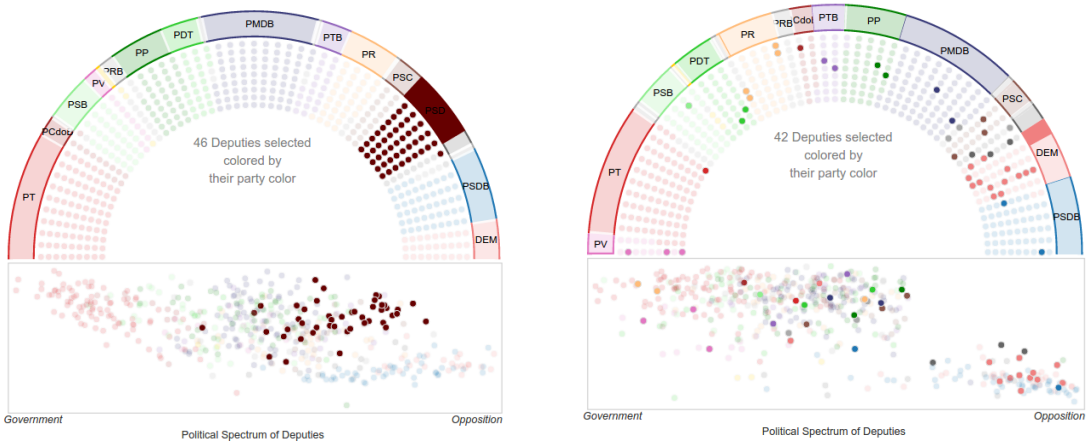
Figure 4.13: The 2012 year period selected and general elections of 2012 toggled. When a election is toggled CivisAnalysis change the parties' color according to its alliance, the color is set by the president nominee's party. (A) The Chamber Infographic shows the alliances ribbon, sorting, according to the deputies spectrum, first the alliances and later the parties. (B) Information about the four most voted nominees and alliances is shown: name of the alliance, first round vote results, second round votes result, presidential nominee, parties in the alliance. (C) The Party Timeline change the colors according to the 2012 alliance.

4.5.6 Lock Deputies Feature

CivisAnalysis reset the deputies and roll calls selections when the user change the period of time, because the deputies change. To avoid the deselection and keep

track of an arbitrary set of deputies along different periods of time we provide the *Lock Deputies* feature. When a group of deputies is selected, CivisAnalysis enables the *Lock Deputies* button, shown in Figure 4.4(A).

With this feature one can track deputies positions in the spectrum and observe deputy's party change, and party creation. The creation of a party during a legislature can be tracked as shown in Figure 4.14.



(a) Selection of PSD deputies in 2012. The position of the deputies in the spectrum shows a party close to the opposition with a relative point of convergence, Figure (b) show us this deputies in a different period of time.

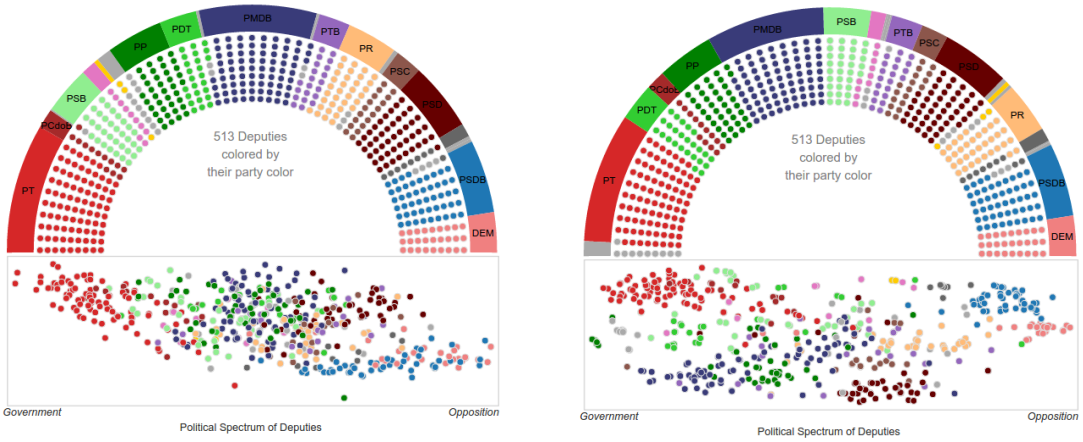
(b) Locking the selection of PSD in 2012 and changing the period for the first semester of 2011, we can observe that before the party changes, the PSD deputies were not aligned by their votes in roll calls.

Figure 4.14: Tracking the Social Democrat Party (PSD) creation.

4.5.7 Alternative method: t-SNE spectra

In the left menu of the *Inspection Module* (Figure 4.4(A)), there is a *Generate t-SNE* button. Clicking on this button will open a dialog to input the parameters of the t-SNE method (Figure 4.15). t-SNE will run in two instances in parallel to generate the two spectra, roll calls and deputies. After the spectra is generated, they are displayed in CivisAnalysis using the same methods for displaying the PCA spectra, the user being able to explore it with the same tools. Figure 4.16 shows the difference of the spectrum of Deputies generated by PCA and t-SNE, the latter with randomly chosen, parameters, as learning rate = 10, perplexity = 10, and 10 seconds of calculation in a laptop Intel CORE i5 8GB RAM.

Figure 4.15: To generate the t-SNE spectra the user must input the parameters of the method, perplexity, learning rate, and the number of seconds of calculation. When the user click on Calc t-SNE, CivisAnalysis will run in parallel two t-SNE instances to generate the spectrum of Deputies and the spectrum of roll calls.



(a) Spectrum of deputies generated by PCA (b) Spectrum of deputies generated by t-SNE

Figure 4.16: Spectrum of deputies generated for the 2013 year. (a) PCA is a global dimensionality reduction technique, the Government vs Opposition dimension becomes more evident. (b) t-SNE is a local dimensionality reduction technique, the parties partisanship becomes more clear in the spectrum.

4.5.8 Final comments

In this chapter we have presented CivisAnalysis with examples of all its features. CivisAnalysis was developed based on a cycle of design, prototyping and informal evaluation by people interested in politics, which acted as voluntary stakeholders. A final evaluation round is reported in the next Chapter.

5 EVALUATION AND DISCUSSION

As mentioned in Chapter 4, CivisAnalysis was developed following cycles of design, prototyping and informal evaluation. These informal evaluations were performed by different people throughout the development. We also had feedback from the presentation of a preliminary version of the application (BORJA; FREITAS, 2014).

Then, the current version was subject to a remote user study for addressing specific aspects related to the CivisAnalysis use by people that had not previous contact with the application. In section 5.1 we describe this evaluation and discuss the results, which have already been published (BORJA; FREITAS, 2015).

Another evaluation was performed with a different public, mainly high school students, to check the different perceptions. User experience issues were identified in both evaluations.

Finally, with a third group of people, we applied the standardized System Usability Scale questionnaire to obtain an overall usability score of our application.

This chapter ends with a discussion of the performed evaluations and proposed solutions for the identified users' difficulties.

5.1 Remote Users Study

The survey was focused on a population of users that were potentially interested in analyzing the data from the Brazilian Chamber of Deputies. We aimed at evaluating the intuitiveness of our design choices with respect to both visualization and interactive features, in abstract tasks like:

- Understanding the parties' flow and its selection possibilities;
- Understanding the political spectrum of deputies;
- Understanding the political spectrum of roll calls;
- Assessing the ease of selecting deputies and roll calls, and

- Assessing the filtering and textual selection controls.

5.1.1 Procedure

We set up CivisAnalysis website and a questionnaire designed to measure the above aspects in the context of free use of the application, without prior training or tutorial. We sent the survey to professors, students and researchers of different areas as well as designers via mailing lists and social networks. The volunteers were asked to freely play with CivisAnalysis and complete our questionnaire.

The questionnaire had eight statements with a five-point Likert scale of agreement (from 1 meaning strongly disagree to 5, strongly agree), as follows:

1. I can understand the meaning of the flow of different parties along time.
2. I can understand easily use the configuration controls in the Timeline Settings menu.
3. I can easily select legislatures, years and intervals for analyzing the Chamber of Deputies composition, the political spectrum of deputies and results from roll calls.
4. I can easily comprehend the political spectrum of deputies.
5. I can easily comprehend the political spectrum of roll calls.
6. I can easily select a single deputy.
7. I can easily select roll calls.
8. I can easily understand how to use the interaction controls with the different representations.

5.1.2 Results

In the short period of five days we received 15 complete responses. Most of our 15 subjects are male (80%, 12 out of 15). They are in average 33 years old, with their ages ranging from 21 to 49. Among participants, 5 are professionals (4 are software engineers and 1 is a designer), 5 are computer science (CS) graduate students and 5 are professors (4 CS and 1 human sciences professor). We asked how often they read news about politics and obtained the following profile: 5 (33.3%) read daily; 4 (26.7 %) read 3 to 4 times a week; 2 (13.3%) read 1 to 2 times a week, and 4 (26.7%) rarely read this type of news. As to whether they had ever seen visualizations of

political data, 50 % of our subjects reported that they had seen them before, mostly infographics about the 2014 elections.

In general, we had a positive feedback on both *Summary* and *Inspection* modules. The main roll call analysis functions were performed and understood by users. Tasks CT 1 to CT 4 could be easily executed in the *Inspection Module*; CT 7 and CT 8, in the *Summary Module*. As for CT 5 and CT 6, users found difficulties, with their main doubt being about the roll call spectrum calculation. As expected, the main complaints were the lack of information about the techniques and the lack of a glossary (about Brazilian legislative procedures) and online help.

The *Party Timeline*, in the *Summary Module*, was the best understood component: 60% marked "Strongly agree" and 40% marked "Agree", with a total of 100% of agreement.

Regarding the interactive timeline controls, the level of understanding was also high both for the configuration controls (80% of agreement) and the selection of time intervals (93.3% agreement). One user reported difficulties to select or follow small parties because they have a small height in the visualization due to their small number of deputies. Party spectrum positions after application of the *Unclutter* function was preferred over their original, but there were concerns because these positions can change according to the function's parameters.

As for the *Inspection Module*, the political spectrum of deputies was well understood by the subjects (80% of agreement) and they also found it easy to select deputies (86.7% of agreement). Although less understood, the political spectrum of roll calls got 73.3% of agreement in understandability. They also considered easy to select roll calls (86.7% of agreement). However, interactive selection tools in the Chamber of Deputies representation, the map and both scatterplots, as well as the text-based search, were not very well understood: 66.7% of users found them understandable, while 20% were undecided and 13.3% did not understand them.

Five out of 15 subjects found difficult to understand the political spectrum of roll calls, especially how the two spectra relate to one another. Most users reported that they missed legends, hints and clear explanations about some of the components in the visualizations.

One of our subjects said that CivisAnalysis is "an application with a huge potential of going viral on social media networks". It should be noticed that we did not collect information about the computational platform the subjects used neither how much time they spent playing with CivisAnalysis.

5.2 Local Users Study

The second evaluation was performed locally, during the UFRGS Portas Abertas 2016 event (the annual Open House event for high school students). We also aimed at evaluating the intuitiveness of our design choices with respect to both visualization and interactive features. Moreover, we would like to discover interest points for further development with a different kind of public.

5.2.1 Procedure

People taking part in the event were asked to volunteer, by using the application always in the same room (a laboratory), same computer (running Windows 7 and Google Chrome). The event took place precisely at a troubled time for politics in Brazil - the country was going through the impeachment process of President Dilma Rousseff. That for sure have directly influenced the interest for politics of our visitors.

Visitors were invited to use the system, and perform three sets of small tasks. The first set of tasks was related to the evaluation of the Timeline, while the second set of tasks aimed at evaluating the use of the Chamber of Deputies' Infographic, the spectrum of Deputies and the map of the districts of Brazil. Finally, the last set of tasks was related to the spectrum of roll calls.

After each set of tasks, the volunteer was asked if he/she wanted to continue the experiment to the next set of tasks. Only users that had shown interest in a specific part of the interface throughout the experiment were invited to the third set of tasks.

During the test notes were taken from the observation of volunteer's behaviour. At the end of the experiment with each volunteer, he/she was interviewed for answering specific questions and giving free comments. It is important to say that, for avoiding bias, the observer was not the author of CivisAnalysis, it was another M.Sc. student with experience in evaluation. We chose to describe the results in the next section while describing each task proposed in the evaluation.

5.2.2 Tasks and Results

This evaluation round was performed with 12 volunteers (average age of 19, standard deviation 2,97), mostly high school students. Some of them had some basic knowledge of computer programming. Despite their youth, the volunteers were very interested in the main focus of the application, politics. Some were just interested to see how an application for political data visualization works.

Regarding the first set of tasks (evaluation of the Timeline), firstly, we asked the volunteers what they could understand just watching the Timeline chart without performing any interaction. Two people could not extract any information from the chart, while most of the others reached incorrect conclusions, the most common mistake being misinterpreting the height of rows as the number of deputies. Four people have succesfully related the Timeline with the behavior of parties over the years, understanding the change in the party's position according to each president. A very common behaviour among the users in this task was to relate the colors of the parties of the elected Presidents with the colors in the Timeline, so they were able to identify certain parties in the Timeline without interacting directly with the representation.

The second task in this phase was to analyze the behavior of significant parties over time by interacting with the Timeline chart. Most of the volunteers could see the changes in position of the parties, relating the President's party with the position of other parties representing the opposition or the government at a given time. The main difficulty encountered in this phase was to find a particular party: few realized that it was necessary to move the mouse over the small vertical bar to trigger a tooltip with the party's name. About the line width of each party, which represents the number of deputies in the Chamber, only 3 volunteers found its real meaning, some have questioned and other did not realize that at all.

When volunteers finished these two tasks, they were asked about the difficulty level to interpret the Timeline. We had the following answers: Very easy, 2 participants; Easy, 3 participants; Medium, 6 participants; Hard, 1 participant; no one classified it as Very hard.

The second set of tasks started with the volunteer being asked to choose in the Timeline chart which time period he/she wanted to analyze. Users were instructed to choose a presidential mandate, a legislature, a year or an arbitrary period using the slider. As mentined before, these tasks had as objective to evaluate the use of the Chamber Infographic, the spectrum of Deputies and the map of Brazil with the districts.

Having selected the time period in the Timeline and started the analysis of the infographic, which caught more the volunteers' attention was the size of the parties. They clearly noticed which were the predominant parties (i.e., those who had the largest number of seats in the Chamber). Another feature quite perceived was the identification of each member (e.g., name, party and district). A volunteer noticed the arrangement of parties in the Infographic according to the ideology (i.e. left, right). However, most participants did not notice that it was possible to see the exact number of deputies of each party through the tooltip.

Concerning the analysis of the map, interaction was performed without difficulty

and all subjects managed to realize the number of members in each state and relate them to the infographic, noticing the proportion of each party in each district.

Regarding the political spectrum of deputies, which drew most the users' attention was the position of each member, if the deputy tended to the government or the opposition. Some volunteers realized the number of members on each side, concluding, for example, that at given time the opposition was stronger (i.e., more members) than the government. The main difficulty was in understanding the axes of the scatter plot: the horizontal axis has a title (Government/Opposition), though the vertical axis did not. This is because the chart was generated using PCA, and unfortunately it ended up not natural to lay users.

Finally, concerning to the the third set of tasks, for evaluating the spectrum of roll calls, only 3 volunteers performed them. What caught most of their attention were the actually the laws: they observed the name of the roll call, year and other details. However, they failed to understand the votes in the spectrum, what was the meaning of the distance between votes and the meaning of the axes. Again, the scatter plot was generated through SVD, and subjects did not understand it. When asked about what were the relations between the spectrum of roll calls and the infographic of the Chamber, it was evident the difficulty of users to perform this task. There were two conclusions: (1) Each bill is proposed by a party. (2) Each roll call in the spectrum shows the behavior of deputies regarding it. It was observed that they were unaware or gave little importance to the color of deputies in the infographic, which identify how each deputy voted in a given roll call.

In the end of the experiment we asked volunteers if they would use such an application to become acquainted and follow the deputies and votes. The answers were: Yes, 10 participants; and Maybe, 2 participants.

With these results it was evident the importance of applications such CivisAnalysis, since 83.3% of the volunteers would use an application like this, while only 16.7% of them said that "maybe" they would use it.

5.3 Preliminary usability evaluation

From both evaluations of CivisAnalysis, reported in this chapter, we identified user experience issues. So, to provide an initial idea about usability issues of CivisAnalyss, we decided to use the System Usability Scale (SUS) questionnaire in a different group of users. SUS (BROOKE et al., 1996) is a industry standard "quick and dirty" scale for application after usability tests. It is composed by 10 questions, with possible answers following a 5-point Likert scale.

As in the first evaluation, CivisAnalysis was presented by e-mail to 41 people. These were students from a Usability and Accessibility course in the context of

a specialization program at our Institute. They proactively accessed the website, played freely with CivisAnalysis and answered the SUS questionnaire.

CivisAnalysis scored 56.2 in SUS (standarf deviation = 22), which translates to "OK" (BANGOR; KORTUM; MILLER, 2009). This is a marginal score, regarding a data visualization application. Since the subjects' group was considered as having the same knowledge usability, we did not perform correlation test between the subjects' profile and the respective SUS scores.

5.4 Final Comments and Future Work

In times of political crises and corruption scandals, the population actually have more interest in politicians, wanting to know how they behave, and if they actually represent their electorate. Information on roll calls and deputies are available on the internet, however it is not well organized and hard to interpret. CivilAnalysis aims to allow anyone to use the application and monitor deputies. Therefore, all the effort invested in the creation, maintenance and extension of such application is fully justifiable and valid.

In general, the results of the evaluation of the Timeline chart and the Political Spectrum with the Chamber's Infograhic and districts views were interesting and allowed us to think about improvements.

Selecting a legislature in the Timeline and the biennial snapshot of the Brazilian politics in the congress shown in the scatterplot did not receive critics, suspicion, or dubious arguments concerning the expression of the reality. In fact, people feel amazed, after they comprehended the spectrum, since they discovered what they had already imagined but could not create in their minds or express in natural language.

It is totally understandable that the first and main critiques were about the missing axes titles, and more explanations. CivisAnalysis was created as a laboratory for testing, but ended up as an interesting prototype and can be extended to be a tool designed to end users. We should start by translating into portuguese, modifying graphic design so user experience can be restructured.

Not available in the evaluations, the unclutter function now has a force algorithm to arrange the parties, which minimizes the difference from the original positions in the spectrum and gives a better aspect to the Timeline. Small parties do not have many seats in the Congress, and this translates to few votes and low political power in the legislative. Therefore, their small height in the timeline, almost invisible, is justifiable. However, since Dilma Roussef impeachment, small parties like PSOL (left-wing liberals) and PSC (right-wing conservatives) have captured the political debate causing people to search for these parties in the Timeline. Next version of

CivisAnalysis legends with number of seats and highlight of parties should be made available to users. Another timeline feature planned and requested by users are the trace of one Deputy in the timeline and major events described along the time.

The less understood of our visual representations was the political spectrum of roll calls and its axes generated by SVD. It is created by the same technique and computation of the spectrum of deputies. Both spectra have full similarities and a direct relationship concerning votes. Explanation and information about the techniques would aid users to understand them, although doubts about the meaning of axes are a common problem of feature extraction techniques and recurrent case of evaluation discussion in roll call analysis studies that use dimensionality reduction.

Our first approach in solving this problem, found in the first evaluations, was to label the first principal component by observed correlation as "Government vs Opposition". However it was not enough since the second component now generated more doubts as we could not find or express a specific label that crosses many legislatures. The spectrum of roll calls showed us interesting and different patterns: when the coalition government is strong the principal component expresses only the division between the accepted and rejected roll calls - in a strong and cohesive coalition they have full control of which bills will be accepted or rejected. In a weak coalition government the principal and second component act like the second component in a strong coalition - a full spectrum of dispute.

Previous works on roll call analysis removed roll call with major agreement from the calculations. The idea is to enhance variance between the political actors since political differences are made by disagreement. This method can be implemented but with caution, to not conceal roll calls. CivisAnalysis does not implement this method since the idea was to make all roll calls available to inspection.

We are studying the change of axes' labels to their original meaning: principal and second components (or main components of variance in the data) and label the government as the average or median position of deputies in the spectrum. For example, in the Timeline besides the line marking the middle point between government and opposition, a line in function of the average and median will follow the position of the majority. Consequently, we could avoid abrupt changes in positions of parties caused by this government versus opposition view, like the changes that occurred in 2003, and in 2016, when the opposition became the government.

The roll call spectrum (or the idea of dimensionality reduction of the roll calls) was not fully explored in this work. Maybe the key to this idea is to use it in a reduced number of roll calls to tell a story about a political debate or context. It can be a powerful tool to train machine learning and automated storytelling systems. Using our method, the position of the roll call in the spectrum is a function of the position of deputies by a direct relationship with their votes. This interesting

property can be used to display the roll calls coherently in the timeline aligned to the parties political spectrum.

6 CONCLUSIONS

Civis Analysis offers a unique view of the Brazilian Chamber of Deputies's political history. Common "citizens tasks" are supported. One can explore and compare voting patterns of different legislatures, observe party behavior over time, investigate party or electoral district patterns in roll calls, relate the pre- and post-election alliances with government and opposition, coalitions, etc.

The overview provided by the party timeline gives the citizen a powerful tool to comprehend and express its feelings about politics. The inspection module, with its coordinated views and several selection and filtering features on raw data, allows users to investigate deputies, parties, and roll calls, improving their ability to understand, interpret, create, and personalize politically relevant information.

Features could be added to our application to allow for a deeper investigation: the tracking of individual behavioral over time, notes of events in the timeline, navigation over tabular data, etc. The roll call spectrum idea is likely to be a powerful tool, which its potential has not been fully explored in this work. The symmetry between the spectrum of deputies and the spectrum of roll calls can be used to create stories of disputes over legislative motions and subjects.

Roll call data have multiple possible perspectives, and each visualization design decision in CivisAnalysis was strongly thought to preserve all these political visions.

REFERENCES

BANGOR, A.; KORTUM, P.; MILLER, J. Determining what individual sus scores mean: Adding an adjective rating scale. **J. Usability Studies**, Usability Professionals' Association, Bloomingdale, IL, v. 4, n. 3, p. 114–123, may 2009. ISSN 1931-3357. Available from Internet: <http://dl.acm.org/citation.cfm?id=2835587.2835589>. One citation in page 57.

BORJA, F.; FREITAS, C. Civisanalysis: Interactive visualization for exploring roll call data and representatives' voting behaviour. In: **Graphics, Patterns and Images (SIBGRAPI), 2015 28th SIBGRAPI Conference on**. [S.l.: s.n.], 2015. p. 257–264. One citation in page 51.

BORJA, F. G.; FREITAS, C. M. Civisanalysis: Unveiling representatives' policies through roll call data visualization. In: **WVIS 2014 - 5th Workshop on Visual Analytics, Information Visualization and Scientific Visualization**. Fundação Getulio Vargas, 2014. Workshop co-located with SIBGRAPI 2014. Available from Internet: <http://hdl.handle.net/10438/11954>. One citation in page 51.

BROOKE, J. et al. Sus-a quick and dirty usability scale. **Usability evaluation in industry**, London, v. 189, n. 194, p. 4–7, 1996. One citation in page 56.

CARROLL, R. et al. Comparing nominate and ideal: Points of difference and monte carlo tests. **Legislative Studies Quarterly**, Wiley Online Library, v. 34, n. 4, p. 555–591, 2009. One citation in page 18.

CELEBRITYTYPES. **Political Coordinates Test**. 2015. Available from Internet: <http://www.celebritytypes.com/political-coordinates/test.php>. 2 citations in pages 22 and 23.

CLINTON, J.; JACKMAN, S.; RIVERS, D. The statistical analysis of roll call data. **American Political Science Review**, Cambridge Univ Press, v. 98, n. 02, p. 355–370, 2004. One citation in page 18.

COMPASS, P. **Political Compass**. 2015. Available from Internet: <http://www.politicalcompass.org/>. 2 citations in pages 22 and 23.

DESARBO, W. S.; CHO, J. A stochastic multidimensional scaling vector threshold model for the spatial representation of "pick any/n" data. **Psychometrika**, Springer, v. 54, n. 1, p. 105–129, 1989. One citation in page 20.

DESPOSATO, S. W. Parties for rent? ambition, ideology, and party switching in brazil's chamber of deputies. **American Journal of Political Science**, Blackwell Publishing, Inc., v. 50, n. 1, p. 62–80, 2006. ISSN 1540-5907. One citation in page 18.

D'HONDT, V. La représentation proportionnelle des partis par un électeur. **Ghent.(1882). Systdme pratique et raisonng de repre'-sentation proportionnelle. Brussels: Muquardt**, 1878. One citation in page 16.

DONY, R. Karhunen-loeve transform. **The transform and data compression handbook**, Boca Raton: CRC Press LLC, 2001. One citation in page 19.

ENELOW, J. M.; HINICH, M. J. **The spatial theory of voting: An introduction**. [S.l.]: CUP Archive, 1984. One citation in page 18.

FISHER, D. R.; LEIFELD, P.; IWAKI, Y. Mapping the ideological networks of american climate politics. **Climatic change**, Springer, v. 116, n. 3-4, p. 523–545, 2013. 3 citations in pages 18, 25, and 27.

FOLHAPRESS. **PT perde cadeiras, e Câmara dos Deputados se fragmenta**. 2014. Accessed: 2015-03-31. Available from Internet: <http://www1.folha.uol.com.br/poder/2014/10/1528021-pt-perde-cadeiras-e-camara-dos-deputados-se-fragmenta.shtml>. 2 citations in pages 29 and 30.

FRIGGERI, A.; FLEURY, E. Agreement Groups in the United States Senate. In: **IOGDC - International Open Government Data Conference**. [s.n.], 2012. Available from Internet: <http://hal.inria.fr/hal-00705682>. 2 citations in pages 28 and 29.

GOLUB, G. H.; REINSCH, C. Singular value decomposition and least squares solutions. **Numerische Mathematik**, Springer, v. 14, n. 5, p. 403–420, 1970. One citation in page 36.

GROFMAN, B. et al. Comparing and contrasting the uses of two graphical tools for displaying patterns of multiparty competition: Nagayama diagrams and simplex representations. **Party Politics**, v. 10, n. 3, p. 273–299, 2004. One citation in page 31.

HOUGHTON, W. R. **Conspectus of the History of Political Parties and the Federal Government**. [S.l.]: Granger, Davis & Company, 1880. 2 citations in pages 26 and 28.

IVOSEV, G.; BURTON, L.; BONNER, R. Dimensionality reduction and visualization in principal component analysis. **Analytical Chemistry**, v. 80, n. 13, p. 4933–4944, 2008. PMID: 18537272. Available from Internet: <http://dx.doi.org/10.1021/ac800110w>. One citation in page 20.

JAKULIN, A.; BUNTINE, W. Analyzing the us senate in 2003: Similarities, networks, clusters and blocs. 2004. One citation in page 19.

JEDIDI, K.; DESARBO, W. S. A stochastic multidimensional scaling procedure for the spatial representation of three-mode, three-way pick any/j data. **Psychometrika**, Springer, v. 56, n. 3, p. 471–494, 1991. One citation in page 21.

JOLLIFFE, I. T. **Principal Component Analysis**. [S.l.]: Springer-Verlag, 1986. One citation in page 19.

KINNAIRD, P.; ROMERO, M.; ABOWD, G. Connect 2 congress: visual analytics for civic oversight. In: ACM. **CHI'10 Extended Abstracts on Human Factors in Computing Systems**. [S.l.], 2010. p. 2853–2862. 2 citations in pages 23 and 24.

KRAUSE, S.; SCHMITT, R. **Partidos e coligações eleitorais no Brasil**. [S.l.]: Unesp, 2005. One citation in page 16.

LEEUW, J. de. Principal component analysis of binary data. applications to roll-call-analysis. **Department of Statistics, UCLA**, 2003. One citation in page 21.

LEIFELD, P.; HAUNSS, S. Political discourse networks and the conflict over software patents in europe. **European Journal of Political Research**, Wiley Online Library, v. 51, n. 3, p. 382–409, 2012. One citation in page 25.

LESTER, J. The evolution of the political compass (and why libertarianism is not right-wing). **Journal of Social and Evolutionary Systems**, Elsevier, v. 17, n. 3, p. 231–241, 1994. One citation in page 22.

LEVENSHTEIN, V. I. Binary codes capable of correcting deletions, insertions, and reversals. In: **Soviet physics doklady**. [S.l.: s.n.], 1966. v. 10, n. 8, p. 707–710. One citation in page 18.

MAATEN, L. Van der; HINTON, G. Visualizing data using t-sne. **Journal of Machine Learning Research**, v. 9, n. 2579-2605, p. 85, 2008. One citation in page 36.

MARINO, R. **A valsa dos partidos, de Collor a Dilma**. 2014. Available from Internet: <http://www.todasasconfiguracoes.com/2014/04/24/a-valsa-dos-partidos/>. 2 citations in pages 23 and 25.

MEREPRESENTA.ORG. **MeRepresenta**. 2015. Available from Internet: <http://www.merepresenta.org.br/>. One citation in page 31.

MOODY, J.; MUCHA, P. J. Portrait of political party polarization. **Network Science**, Cambridge Univ Press, v. 1, n. 01, p. 119–121, 2013. One citation in page 29.

MORGENSTERN, S. **Patterns of legislative politics: Roll-call voting in Latin America and the United States**. [S.l.]: Cambridge University Press, 2004. One citation in page 16.

MUNROE, R. **A History of The United States Congress, Partisan and Ideological Makeup**. 2012. <http://xkcd.com/1127/large/>. Accessed: 2015-03-31. 2 citations in pages 26 and 29.

PAULO, E. de S. **Basômetro**. 2013. Available from Internet: <http://estadaodados.com/basometro/>. One citation in page 23.

PAULOVICH, F. V. et al. Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping. **Visualization and Computer Graphics, IEEE Transactions on**, IEEE, v. 14, n. 3, p. 564–575, 2008. One citation in page 36.

PERER, A.; SHNEIDERMAN, B. Integrating statistics and visualization for exploratory power. 2009. 2 citations in pages 25 and 27.

POOLE, K. T. Nonparametric unfolding of binary choice data. **Political Analysis**, SPM-PMSAPSA, v. 8, n. 3, p. 211–237, 2000. One citation in page 18.

POOLE, K. T. **Spatial models of parliamentary voting**. [S.l.]: Cambridge University Press, 2005. One citation in page 20.

POOLE, K. T.; ROSENTHAL, H. A spatial model for legislative roll call analysis. **American Journal of Political Science**, JSTOR, p. 357–384, 1985. One citation in page 18.

POOLE, K. T.; ROSENTHAL, H. Patterns of congressional voting. **American journal of political science**, JSTOR, p. 228–278, 1991. One citation in page 21.

PRIESTLEY, J.; FRANKLIN, B. **A new chart of history**. [S.l.: s.n.], 1769. One citation in page 26.

RADAR Parlamentar. 2015. <http://radarparlamentar.polignu.org/>. Accessed: 2015-07-22. 2 citations in pages 23 and 26.

REES, K.; CITRARO, D. Voteeasy: Helping voters find political matches. **IEEE VisWeek Discovery Exhibition**, 2011. One citation in page 31.

ROSENTHAL, H.; VOETEN, E. Analyzing roll calls with perfect spatial voting: France 1946–1958. **American Journal of Political Science**, Wiley Online Library, v. 48, n. 3, p. 620–632, 2004. 2 citations in pages 19 and 23.

SHNEIDERMAN, B. The eyes have it: a task by data type taxonomy for information visualizations. In: **Visual Languages, 1996. Proceedings., IEEE Symposium on**. [S.l.: s.n.], 1996. p. 336–343. ISSN 1049-2615. One citation in page 39.

WEBCITIZEN. **VotenaWeb**. 2009. Available from Internet: <http://www.votenaweb.com.br/>. One citation in page 31.

ZUCCO, C. J. Ideology or what? legislative behavior in multiparty presidential settings. **The Journal of Politics**, v. 71, p. 1076–1092, 7 2009. ISSN 1468-2508. One citation in page 16.

APÊNDICE A

A.1 CivisAnalysis: introdução

Cidadãos ativos e informados são essenciais para uma democracia saudável. A democracia essencialmente criada em cidades-estado hoje em dia é aplicada em grandes Estado-nações, na sua principal variante moderna a democracia representativa. Devido ao imenso tamanho das democracias representativas manter-se informado sobre assuntos públicos é uma tarefa desafiadora. Nas democracias representativas a mais importante ferramenta do cidadão é o voto, que é como os cidadãos fornecem *feedback* para ajustar o sistema democrático. Para poder votar efetivamente, um cidadão necessita informação política consistente, especialmente no que diz respeito a seus representantes. Os dados políticos referentes a representantes podem ser encontrados em todas as partes de suas vidas públicas: discursos, entrevistas, reuniões, etc. Mas estes dados podem mascarar o efetivo posicionamento do político. As votações no plenário do Congresso são especialmente importantes pois têm um impacto direto na sociedade. Como representantes políticos se comportam em votações no plenário é, portanto, uma fonte inestimável de dados que pode ser usada para manter os cidadãos informados.

A.1.1 Motivação

Durante as votações em plenário cada representante escolhe explicitamente uma opção ("Sim", "Não", "Abstenção" e "Obstrução"). Além do voto individual a votação representa um conjunto de votos que pode revelar o comportamento de grupos no Congresso e na sociedade em geral. Cada voto em plenário está diretamente associado a um distrito eleitoral e um partido político; indiretamente, está associado a tratos sociais como religião, gênero, riqueza, etc. Isso faz com que a votação nominal seja um conjunto confiável de dados para entendimento de políticos e posicionamentos dos representantes, bem como das políticas dos partidos em vários aspectos. No entanto, o grande número de votações por ano torna impraticável que os cidadãos

tenham capacidade de acompanhar os votos de seus representantes.

A alternativa aqui proposta para entender o comportamento dos representantes é observar a distribuição dos representantes em um espaço n-dimensional onde comportamentos semelhantes correspondem a posições próximas naquele espaço. Comportamento semelhante traduzimos como votos similares nas mesmas votações em plenário. Uma representação geométrica desse espaço é chamada de "espectro político", o qual fornece informações visuais sobre o posicionamento político dos representantes.

As iniciativas governamentais de dados abertos dão acesso a diferentes conjuntos de dados políticos, facilitando a difusão da informação. Por outro lado, o enorme volume de informações desorganizadas e altamente tendenciosas espalhadas pela Internet pode causar um impacto negativo na assimilação da informação devido ao efeito comumente conhecida como *sobrecarga de informação*. Portanto, é um desafio extrair informações coesas do *big data* político e comunicá-las ao cidadão.

Entender analiticamente uma única votação em plenário sem o auxílio de uma ferramenta é viável, mas conclusões assertivas sobre esta votação vão ser rasas se o autor não tem um bom conhecimento sobre o contexto político. Pode-se também precisar obter informações complexas sobre períodos maiores de legislatura, por exemplo, inspecionar votos de diferentes conjuntos de deputados (partidos, alianças, bancadas, distritos, ...) ligados a conjuntos de votações (possivelmente sobre um assunto específico) dentro de um determinado intervalo de tempo. Uma ferramenta analítica para testar rapidamente hipóteses é, portanto, necessária para podermos tirar conclusões sobre contextos políticos maiores e/ou diferentes.

A.1.2 Objetivos e Contribuição

Este trabalho centra-se nas votações realizadas na Câmara Brasileira dos deputados federais, a câmara baixa. A Câmara é formada por 513 deputados eleitos por representação proporcional em distritos eleitorais para servir termos de quatro anos. O Brasil tem 27 distritos eleitorais, cada um com um mínimo de 8 e um máximo de 70 assentos na câmara; sendo o número proporcional à população de cada distrito. Representação proporcional em grandes distritos eleitorais e leis flexíveis de criação de partidos e alianças eleitorais tornou possível um grande número de partidos eleitos. A legislatura de 2011-2014 teve 21 partidos, e 352 votações em plenário (um total de 114.994 votos).

Utilizando os dados abertos da Câmara dos Deputados federais, desenvolvemos CivisAnalysis, uma ferramenta de visualização que visa facilitar e acelerar a descoberta e compreensão dos padrões de votação do Congresso. Esta ferramenta integra o estado da arte e novas técnicas de visualização para análise de votações

em plenário, usando um conjunto de dados que abrange os registos de voto de seis legislaturas (24 anos), incluindo 914 propostas legislativas votadas através de 2.458 votações nominais (853.952 votos) - bem como as informações de seis eleições presidenciais, incluindo resultados eleitorais e alianças/coligações feitas durante as eleições.

CivisAnalysis é uma ferramenta de análise interativa. Ela fornece diversas perspectivas sobre as votações e análises contextuais de acordo com o desejado período de tempo de legislatura, conjunto de votações e/ou conjunto de deputados escolhido.

As principais contribuições do presente trabalho são:

1. Visualização cartesiana das votações em plenário como uma projeção de um espaço n-dimensional. A visão coordenada dos espaços cartesianos de votação e deputados - a interação sobre um conjunto de votações afeta o espectro político de deputados de acordo com os votos e vice-versa - interação no conjunto de deputados afeta o espectro político de votações de acordo com seus votos;

2. Novo método de criação de uma linha do tempo do posicionamento dos partidos de acordo com análises de votação e a integração de dados eleitorais de longo prazo na análise de votações.

A.2 CivisAnalysis: breve discussão sobre as avaliações e trabalhos futuros

Em tempos de crises políticas e de escândalos de corrupção a população tem mais interesse pela vida política, querendo saber como votam os deputados. Informações sobre votações em plenário estão disponíveis na Internet; no entanto, estes registros são apresentados como dados brutos, não bem organizados e difíceis de interpretar. CivisAnalysis visa permitir que qualquer pessoa monitore deputados.

Em geral, os resultados da avaliação do gráfico da linha de tempo dos partidos e do espectro com as diferentes perspectivas da Câmara e distritos foram interessantes e pode-se pensar em melhorias.

A linha do tempo dos partidos e sua inspeção a nível detalhado não recebeu críticas, suspeita ou argumentos duvidosos sobre a expressão da realidade. Na realidade, as pessoas sentiram-se "espantadas" com o infográfico, depois de compreender o espectro político, pois descobriram visualmente o que elas já imaginavam mas não podiam criar em suas mentes ou expressar em língua natural a complexidade política do posicionamento de diversas legislaturas.

É perfeitamente compreensível que as primeiras e principais críticas ao CivisAnalysis tenham sido sobre a falta de títulos nos eixos, legendas e melhores explicações. CivisAnalysis foi criada como laboratório de testes para prototipar novas

ideias e ferramentas de análise de votos. Apesar disso com pouco esforço ela pode ser estendida para um projeto destinado a usuários finais.

Nas avaliações, a representação visual menos compreendida foi o espectro político de votações e seus eixos gerados através do SVD. Este espectro foi criado usando a mesma técnica que o espectro dos deputados, que aparentemente é mais inteligível por se tratar de amostras mais tangíveis, semelhanças de deputados e partidos. Ambos os espectros têm semelhanças e uma relação direta em relação aos dados, votos, e forma. Explicações e informações sobre as técnicas ajudariam os usuários a entendê-las, embora dúvidas sobre o significado dos eixos sejam um problema comum de técnicas de *feature extraction* e discussão de avaliação em estudos de *roll call analysis* que usam redução de dimensionalidade.

O espectro de votações em plenário (ou a idéia do espaço dimensional das votações) não foi totalmente explorado neste trabalho. Talvez a chave para essa idéia seja usá-la em um número reduzido de votações para contar uma história sobre um debate político ou contexto político. Pode ser uma ferramenta poderosa para treinar algoritmos de aprendizagem de máquina ou sistemas de automação de *storytelling*. Usando o nosso método, os votos criam uma relação direta entre as posições das votações e as posições dos deputados em um plano cartesiano. Tirando proveito da facilidade de compreensão da linha do tempo dos partidos pode-se usar esta interessante propriedade cartesiana para exibir as votações de forma alinhada com o posicionamento político dos partidos na linha de tempo.

A.3 CivisAnalysis: conclusões

CivisAnalysis oferece uma visão única da história política da Câmara dos Deputados. Pode-se explorar e comparar padrões de votações de diferentes legislaturas, observar o comportamento de partidos através do tempo, investigar os posicionamentos de partidos, distritos eleitorais, coalizões de governo, coligações, etc.

A visão geral fornecida pela linha do tempo de partido dá ao cidadão uma poderosa ferramenta para compreender e expressar seus sentimentos sobre política. O módulo de inspeção, com suas visões coordenadas e vários recursos de seleção e filtragem em dados brutos, permite aos usuários investigar deputados, partidos e votações em plenário aprimorando sua capacidade de compreender, interpretar, criar e personalizar informações políticas relevantes com evidência quantitativa.

Recursos podem ser adicionados à aplicação para permitir uma investigação mais aprofundada: o rastreamento do comportamento individual de deputados ao longo do tempo; notas de eventos na linha do tempo; navegação sobre dados tabulares, etc. Estima-se que a idéia do espectro de votações é uma ferramenta poderosa que não teve seu potencial totalmente explorado neste trabalho. A simetria entre o espectro

de deputados e o espectro de votações pode ser facilmente usada como suporte para narrar disputas e reposicionamentos no legislativo.

Os dados das votações em plenário têm múltiplas perspectivas possíveis, e durante o projeto do CivisAnalysis cada decisão foi tomada para preservar a integridade destes dados.