Visualizing and Analyzing Voting Records from Historical Documents

Cantareira, Gabriel Dias
gabriel.dias_cantareira@kcl.ac.uk
King's College London, United Kingdom

Cole, Nicholas
nicholas.cole@history.ox.ac.uk
Oxford University

Abdul-Rahman, Alfie
alfie.abdulrahman@kcl.ac.uk
King's College London, United Kingdom

Introduction

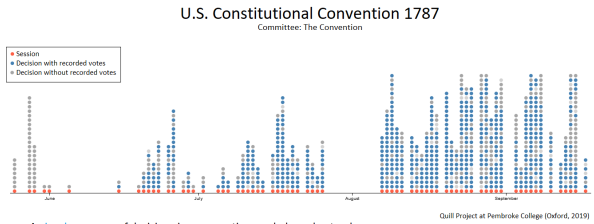
The Quill Project (Cole et al. 2017) is a research initiative that investigates text analysis and data visualization with the aim of assisting users in understanding modern foundational legal texts and constitutional conventions. Quill contains multiple well-organized datasets built on an event-based structure, representing the negotiation of legal texts during legislative drafting for constitutional conventions - processes that can last for several months or years. As each discussed text is the result of hundreds of amendments proposed by several actors over an extended period of time, visual presentation and navigation can become quite difficult.

In this submission, we expand the Quill platform to add features focused on extracting meaningful knowledge from what Quill classifies as decisions: changes or additions that required voting by the attending participants, to which votes and outcome are documented. For each decision, it is possible to track people involved, how they voted, and which other events are related to it. This information can be used to explore the historical context for behavior observed in voting data. We can subdivide our course of action and challenges into three tasks: finding appropriate time windows for observation, dealing with missing or incomplete data, and building an analytical model to observe relevant features and draw conclusions.

Finding Appropriate Time Windows

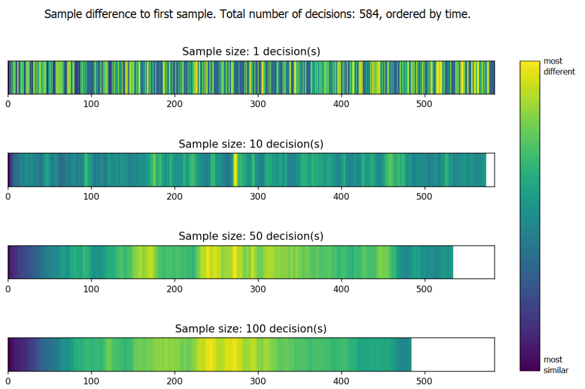
While the use of visualization techniques to observe voting behavior and political positioning is relatively well-established (da Silva et al. 2018), most existing approaches assume there is a semantically appropriate temporal sampling structure available, i.e., time windows to observe and compare voting records are already well defined by historical context. We argue that it is important for a visualization system to be able to work in both directions: users can employ their expertise to select historically relevant dates to observe, but should also be able to use the system to identify noteworthy behavior that could then be traced back to historical events. In the same manner, the size and number of samples to visualize is also important: when visualizing relationships between voters, the entire timespan of the dataset could be viewed as a single sample, but changes that occur inside this timespan will not be easily observable; alternatively, using every single decision as an individual sample would result in hundreds of visualizations to be compared, which is impractical. Therefore, our aim is to aid users in selecting a low number of samples that contain a high amount of variance to one another.

We navigate through this matter using two approaches: the first is to generate a summarized representation of the whole timeline as to when decisions occur and which type they belong to. Users may interact with them and obtain further details, which can also be used to filter the data to be visualized. The second approach is to estimate overall behavior changes over time: groups of people that voted in agreement but started to vote in opposition, and vice-versa. This is done by comparing the differences in pairwise relationships between each voter, using the decisions within the selected time windows as multiple dimensions.



A visual summary of decisions in a convention can help understand the amount of data available and how it is laid out over time.

This representation shows how different voter relationships are in each moment in time when compared to their average. It can be used to identify outliers and points of interest. The perceived outliers can change with different sample sizes.

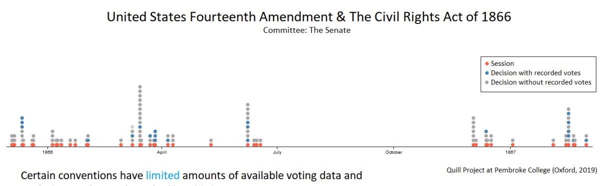


If we choose to start our exploration at the beginning of the timeline, this representation highlights possible spots in time to make a second observation - it shows how different voter relationships become over time

Dealing With Missing or Incomplete Data

A fundamental challenge in working with historical records is data inconsistency. Records can be incomplete or noisy, either due to being lost to time, damaged, or never being kept in the first place. In Quill voting records, not all decisions contain recorded votes. Many of them are identified as anonymous, to which no votes were publicly cast. For others, minutes and drafted documents may reveal causes for such absence, such as a decision being an unanimous agreement without a need for casting individual votes.

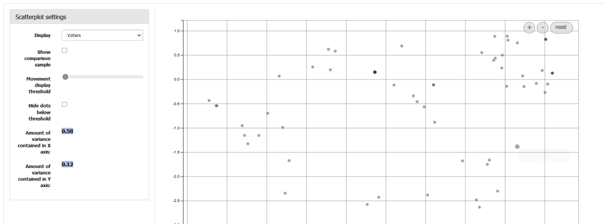
In the sense of observing relationships between voters, decisions with no recorded votes offer no data for visualization. However, we can still use their metadata and related information to provide context and insights to users. Showing the distribution between anonymous and non-anonymous voting can provide insights into periods of contention; cross-referencing decision outcomes to the person that made the proposal can hint at the overall popularity of different people. There is also relevant information in outliers: a vote may look atypical for a person or group for a variety of reasons, and data errors are a possibility.



Certain conventions have limited amounts of available voting data and analysis must be conducted taking this into account

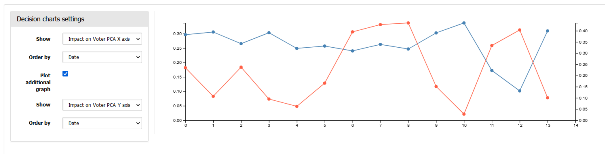
Building an Analytical Model

Once time windows of interest are found and samples are defined, we can then observe them further using computational methods. For this application, we use Principal Component Analysis (PCA) to display all voters as dots in a scatterplot. A PCA projection aims to encode the maximum possible amount of covariance within each dimension without distortion, showing trends and correlation between data points as explicitly as possible. This means voters with similar behavior during the selected time window will be shown close together, while dissimilar voters will be shown far apart.



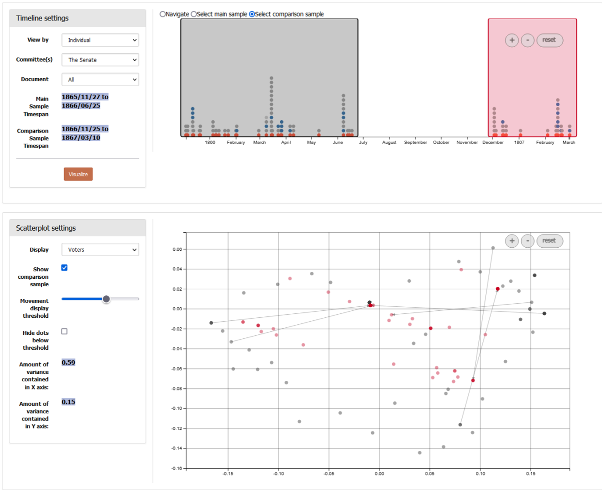
A projection displaying relationships betweeen voters.

There are other projection techniques that can be used in this application – particularly, non-linear methods such as t-SNE and UMAP (Espadoto et al. 2019) can generate accurate neighborhood representations from complex data. However, as PCA is a linear transformation, we can use its features to gather information on the actual meaning of geometric positioning for data: coefficients from PCA analysis inform the influence of each decision on the visualization's axes, and an explanation can be given for the position of voters in the scatterplot.



The graph shows the influence of each decision on the PCA projection's axes, adding more context to geometric positioning.

To enhance this visualization, we can superpose another projection from a different time window and observe changes between the two. Voters with the most expressive changes can be quickly identified, and details on their behavior can be further explored. Normally, two distinct PCA projections do not have comparable spaces; we employ Procrustes analysis to adapt the second projection to the first as much as possible, aiming to highlight differences between the two.



Two different timespan samples are compared as black and red dots.
Arrows show difference in position for the same voters.

Conclusion

Voting records may offer important knowledge on relationships and trends among people or organizations in historical data, and visualization tools provide powerful means to navigate through their features. In this submission, we present tools to be used in this context and discuss their application in the Quill platform, discussing sampling, data availability and presentation.

Bibliography

Cole, Nicholas / Abdul-Rahman, Alfie / Mallon, Grace (2017): "Quill: A framework for constructing negotiated texts-with a case study on the us constitutional convention of 1787". *2017 ACM/IEEE Joint Conference on Digital Libraries* (JCDL): 1-10. IEEE.

da Silva, Rodrigo N. M. / Spritzer, André S. / Freitas, Carla M. D. S. (2018): "Visualization of roll call data for supporting analyses of political profiles". *31st SIBGRAPI Conference on Graphics, Patterns and Images* (SIBGRAPI): 150-157. IEEE.

Espadoto, Matheus / Martins, Rafael M. / Kerren, Andreas / Hirata, Nina S. T. / Telea, Alexandru C. (2019): "Toward a quantitative survey of dimension reduction techniques". *IEEE Transactions on Visualization and Computer Graphics* 27, 3: 2153-2173. IEEE.