Parties in Non-Partisan Settings

Roll Call Votes in the 1973 Louisiana State Constitutional Convention

A. David Jackson
University of Missouri
Professsor Baekkwan Park
ajm7b@umsystem.edu

December 5, 2024

Introduction: State Constitutional Conventions and Roll Call Votes

In this paper, I seek to analyze how atypical legislative bodies (constitutional conventions in this case) act when they do not have the same incentives for reelection as normal legislatures. To do this, I look at the roll call votes of the 1973 Louisiana state constitutional convention. In this introduction, I provide some background information on why this analysis is important and interesting. After my discussion here, I move on to discussing how I gather the data I am analyzing. I then discuss how I pre-processed the data, turning my unstructured digital archival records into structured data. I then move on to discuss how I analyzed my data. I follow up this discussion with my empirical findings before concluding with the shortcomings of my analyses and the implications of both my analysis and my findings.

In normal legislative bodies, legislators work together as parties in order to draft policies. Parties are often seen as essential to the functioning of political systems for at least one reason. In diverse legislatures with many different constituencies with different interests, there are often collective action problems. Each individual legislator is held to a certain standard by their constituencies and have a pressure to produce certain policies that they promise. In order to follow through on their promises, they need a "team," or a party to overcome collective action problems. Parties thus exert power over individual legislators by pressuring them to support party policies in order for their policies to receive support. Constitutional conventions, at least in the American tradition, are temporary legislative bodies that organize for a single objective (writing or amendming a constitution) and then dissolving (Jameson 1887). Delegates to these conventions do not have the same incentives for reelection as typical legislators (Mayhew 1974). The 1973 Louisiana state constitutional convention is a good example of this because most of the delegates are do not hold elected offices outside of the constitutional convention (Barnidge 1974). Theoretically, party cohesion should be lower in this constitutional convention.

Analyzing party cohesion using roll call votes can be something that is somewhat difficult outside of the context of the United States because, in practice, parties usually vote as a block.

However, in the United States, elected officials have an incentive to both the parties, and to their individual constituencies. As such, there is usually enough variation in the voting patterns of legislators that roll call votes that they are useful to determine how much party cohesion there is. In the American politics literature, particularly in the study of Congress, DW-NOMINATE or one of its variants are typically to analyze roll call votes.

The purpose of this paper is to gather and analyze roll call data from the 1973 Louisiana state constitutional convention, but it is part of a larger objective that would be to create an algorithm that can extract roll call votes from archival sources to turn them into usable data for political scientists. I chose to analyze the Louisiana convention because it was one of the most recent state constitutional conventions so I assumed that the archival record would be in a better shape to analyze that older records as a test run for the code I am developing.[1]

Data Collection

The first step of my research project was gathering the data I wanted to use for my analysis. I decided that I would use an archival record of a state constitutional convention, and as I mention above, I decided to use the 1973 Louisiana convention. Finding the data was not too difficult. Archives.org has a large amount of digitized archival records so I found the convention in two volumes. These volumes contained the verbatim records of the debates of the constitutional convention, as well as the recorded votes−which are what I am interested in in this analysis. The roll call votes are embedded within the recorded debates. In other words, extracting roll call votes requires going through the entire text and pulling out the votes themselves. The website had the convention records in multiple digital formats, including PDF and txt files. Because my analysis in Python requires data in a txt format, I just downloaded that version instead of downloading the PDF and then transforming the data into a txt file. One thing I discuss below in my section on shortcomings is that the txt file that Archives.org had for the convention had bad OCR with led to many issues with my pre-processing and analysis.

Because the data I wanted to analyze is roll call votes and not the debates of the convention in their entirety, the next step in my project was to divide the text into parts that contained votes,

[1]As I discuss below, my assumption was wrong.

and parts that did not. To do this, the first thing I did was divide the text into separate blocks. For volume one, I split the record by the word "AMENDMENT." I will explain why I chose this as my cut point below. Using this as my cut point divided volume one into around 2800 blocks. I then only kept blocks that had the word "YE." I used YE instead of YEA or YEAS because the OCR software that Archives.org used did not always correctly translate the words into one of those two words, which would have been more correct. A search of YE suggests that there is around 650 votes that include the word YE in them. After playing with a few different cut points, I found that "AMENDMENT" and "YE" allowed me to retain the most amount data. I managed to keep 508 blocks out of the roughly 650 votes.[2]

For Volume 2, I used a similar process, except instead of using "AMENDMENT," I used "Proposal." Using Proposal gives me just under 1900 text blocks, and when i keep only the blocks that have YE in them, I only have 377 blocks. Unfortunately, this loses me a good amount of votes because the second volume of the convention has around YE 440 times. Unfortunately, I was not able to identify a reliable cutting point that allowed me to keep more votes. This is made more difficult because the OCR for volume two is in much worse shape than the first volume.

After dividing the text of both volumes into blocks and only keeping those that have YE in them, I saved them as their own txt files. I then developed a code that ran through each of those new files and divided the text again into the voting item,[3] yea votes, nay votes, not voting, and absent voters. To do this, I used " roll"[4] "YE" for yea votes, "NA" for nay votes, "NOT "[5], "ABS" for absent voters, and an end cut point of "Total." An issue I ran into was that sometimes there is only either absent or not voting items, while other times there are both items. My code recognizes this and searches for not voting items first, then absent, skipping not voting if there is no item, or skipping absent if there is no item. If there are both, it captures both of them. I then save the text of each of these items for each of the kept blocks and save them into a csv

[2] I should note that there are more votes that do not begin with YEAS because of bad OCR. However, systematically, YE is the best identifier of a vote. Other votes that do not begin with YEAS, YEA, or YE have many different words that precede them so any blanket identifier would not be sufficient. I should also mention that in the actual archival records, all votes begin with YEAS, so this is really an issue with OCR, not with the code.

[3] This is the text that likely contains the substance of the vote, or the text between AMENDMENT or Proposal, and the cut point for the yea votes

[4] Note that there is white space before the word "roll" in this case.

[5] Again, there is white space. Here though it is after "NOT"

file. I then combine both of the csv files.

To summarize my data collection process, I downloaded two raw txt files of the digitized 1973 Louisiana state constitutional convention from Archive.org. I then cut each of the files into blocks, saving them as new, unique txt files. After doing this, I went through at cut the new files into voting items, yea items, nay items, not voting items, and absent items. While the structure of the data was beginning to take shape, it was not quite ready for any type of empirical analysis. It was still an unstructured−albeit organize− text. The next step in this project was to transform the organized text into structured data that I could use by pre-processing the data.

Data Pre-Processing

Data pre-processing is the process of transforming the data from unstructured data which is not suitable for any type of empirical analysis, into structured data that I can perform an analysis on. In this case, I want to transform the data from a list of names in each of the categories I discuss above (yeas, nays, not voting and absent) which act as columns for each voting item, into data where the rows are still each voting item, but the columns are each delegate to the convention. For the purpose of my empirical analysis, which is to use W-NOMINATE for roll call scaling, I need each data point to be a 1, 0, and a -1 for yea votes, not voting, and nay votes respectively.[6] To do this, I went through multiple steps of pre-processing and data cleaning.

With my data in rows for each voting item, where columns are the vote item, and then the text containing the yeas, nays, not voting, and absent items, and with my final goal in mind, the first thing I did was create a list of each person who voted in the constitutional convention. This means I had a list of each of the delegates to the 1973 Louisiana convention, though this list only contained their last names because when roll call votes were taken only the last names were recorded.[7] This list is essential because it works as a type of ground truth since there are many OCR problems specifically with the names of of convention members. I will discuss why I made this list more after talking about my pre-processing further. For now, however, I put this

[6]W-NOMINATE is a version of DW-NOMINATE that does not have a dynamic temporal aspect. Because this is only one legislative meeting, W-NOMINATE is superior to analyzing the roll call votes than DW-NOMINATE or another type of roll call scaling would be.

[7]There are a few exceptions to this. There are two sets of delegates who have the same last names and so I also include their first initial in this list.

list to the side for a later point where it becomes essential.

The second step I identified in pre-processing the data is making the names within each of the columns of my unstructured data into tokens. I do not tokenize the voting item column because it is not necessary for my analysis using W-NOMINATE. Tokenization of each of the columns was quite simple. I decided I would tokenize based on white spaces in the data. This means that every word in each column is transformed from a group of letters into a single item. Simply cutting at white space had a few problems that I had to fix. Most of these problems were with the two voters with the last name Jackson and the two voters with the last name Landry. To deal with these issues, I create a list of exceptions to when the words are tokenized at blank spaces for these last names. While tokenizing, I also remove most symbols−though I do have some exceptions−and make every token lowercase. After tokenizing each word in each of the columns, I create an all words dictionary of every word that contains all of the tokens from all columns except the voting item column.

After I tokenized the words in each column, I had to find a way to get the tokens that were supposed to represent the names of delegates to be able to be identified by the name of the delegate. This was tricky for a number of reasons. The first reason is that, as I mention above, there are major problems with the OCR of the documents. This means that some names are misspelled, or completely messed up. Because of this, its not so simply to just use the list of delegates I created above to extract the names from each list of tokens in each of the columns. The second reason is that there is sometimes lots of extra text after the not voting or absent item−whichever is the last item. This is because, after the final cut word, which as I discuss above is Total, there is more text. If the final cut word does not appear, my code runs until it finds the word, which may be a significant way after where it should be. This is, again, caused largely by bad OCR, as every vote in the physical archival record, does have that as the digitized photo of the record.

To combat both of these issues, I used the "fuzzywuzzy" package in Python. This package allows users to create fuzzy tokens, based on set tokens. Using fuzzywuzzy, I had all tokens that were 85% similar to the names of the delegates attached to the name that they were similar to in a dictionary so that the dictionary was something like: {Delegate Name : [similar token 1, similar token 2, etc.]}. This solves the first issue. By attaching 85% similar tokens to delegate names, I am able to include tokens that OCR may have made unrecognizable if I were to use the untouched name of delegates as the only token that was examined. Tokens that were not related to any of the delegates names were placed as values in the key "unspecified." After creating this new dictionary, I removed the unspecified tokens from the dictionary. I also I remove all tokens that are less than 3 letters. This solves the second issue I discuss above. It also removes words like "total," numbers, page breaks, or other tokens that are not of interest for my observation.

After tokenizing and removing all unrelated tokens from data, I then use my new dictionary where the key is the delegate name and the value is the fuzzy token, to replace each instance of the fuzzy token with its key. This makes my data finally usable. Each fuzzy token, with the hundreds or even thousands of variations that exist in the data, are replaced with one of the 132 delegate names from my list of said names so that each of my columns data only have lists that contain delegate names are in each item. I then remove any rows that do not have yea and nay vote items or have empty lists of delegates. While there might be some importance in these data points I lose, because I am focusing on using W-NOMINATE, I need both the yeas and the nays. After making these important changes to the contents of my data, I adjust its structure. In Python, I wrote a code that creates a new dataframe where each row is still the voting item, but each column is a delegate. In each cell, a candidate is coded as either a 1, 0, -1, -99, or NA for if they voted yea, did not vote, voted nay, if they were absent, or we do not have data for them respectively. My data is now essentially ready for me to analyze.

To summarize my pre-processing, I first created a list of all the delegates who would vote in the convention. This acted as my base truth. I then tokenized all of the columns of my data, created a list of all the tokens in all of the columns, and then created a dictionary that attached all of the tokens to one of the delegates name's or to an unspecified key if they were not similar to any of the delegates' names. I then cleaned the dictionary, removing unnecessary or incorrectly identified tokens. I then replaced all of the tokens with the delegate's name that are the keys that the tokens are values for. I then rearranged the data so that it was formatted with each voting item being a row and each delegate being a column with the cell values being how a delegate voted on a voting item. Now my data is ready to create a spatial map using the

W-NOMINATE method of roll call scaling.

Data Analysis

In this section I discuss and describe roll call vote scaling, provide some preliminary analyses of the pre-processed data−including an explanatory factor analysis of the data, and then discuss the W-NOMINATE method of roll call scaling. I discuss the empirical findings and what these findings suggest in the next section. Because my data is relatively small (n=957, k=132), machine learning methods are would be unable to be used properly. Also, since the analysis I am performing with my data deals with parties in the constitutional convention, and not with the actual content of voting items, the insights from utilizing a machine learning mechanism might not be too much more beneficial than using the techniques that have often been used for analyzing roll call votes in legislatures. For a project that looked more at the content of each of the votes in addition to roll calls and partisanship. or a project that looked at more roll call votes over many time periods or different legislatures, machine learning techniques would be much more appropriate and useful.

Roll call votes are often used in American politics to examine how legislators and parties behave. In the United States and other systems that use elections by geographical districts legislators have incentives to satisfy both the parties they are members of, and the constituencies that actually elect them. Because electoral districts can have a wide variety of interests, even when the partisan division of districts are similar, legislators who are part of the same party can have different policy preferences. This means that legislators will not have the same preference on every issue, even though they might generally present similar policy platforms. Other ways to measure parties (both in and outside of the United States) positioning in the spatial theory literature include using manifestos, using voter's perceptions of parties, using coalition data, or analyzing the public speeches of party leaders. However, roll call votes are useful where enough variation exists for them to be used because they are the actual decisions parties and members of those parties regarding policies.

While roll call votes have been used for a while to analyze how legislators act in legislative bodies, the techniques to empirically study them have changed significantly over time. For a while, correlation matrices, factor analyses, and cluster analyses were some of the primary techniques for trying to understand how legislators and parties made decisions in legislative bodies (MacRae 1965 discusses these. See also Clausen and Van Horn 1977; Jackson 1971; Willets 1972). While pairwise correlation does have some benefits, because there are a large amount of legislators in any legislative body, it is not the most efficient method for analyzing roll call votes. Factor analysis was particularly useful because it suggested how many dimensions the political issue space had. In a typical, two party system, such as in the United States, strong parties would be evident in that they had one or maybe two dimensions. Having one dimension suggests, thought does not necessarily mean, that the primary driving force of votes in a legislature are partisanship or ideology. Two dimensions means that there might be some other type of influence, though in the United States it could be ideology divided. In other words, one dimension might be the social ideology, while a second might be economic ideology. Cluster analysis seeks to find either clusters of roll calls or legislators who vote similarly. These types of analyses are very effective, and I performed them[8] to gain some preliminary insights into the data I gathered.

Threshold	Number of Pairs
10%	4688
20%	2824
30%	1586
40%	893
50%	454
60%	178
70%	53
80%	20
90%	13

Table 1: This table contains how many pairs of delegates reach the threshold for how correlated their voting behavior is. This excludes any pairs that are 100% correlated, including the diagonal. The total number of possible pairs in the 132×132 matrix, not including the diagonal, is 8861.

In Table 1, I include how many pairs of delegates reach certain thresholds of correlation. Note that the table contains 8661 possible pairs above the diagonal, not including the diagonal.

[8]I should note that I use Principal Component Analysis instead of typical factor analysis because the latter was not able to work with my data

This seems to suggest that there is less correlation than one might imagine in a legislative body, with only 454 pairs out of 8661 possible pairs of delegates voted at least 50% similar to each other. This is only 5% of possible pairs. However, as I mention earlier, a pairwise correlation matrix is not super useful for providing easily digestible results of findings. It also only looks at the covariance between two delegates at any given time. The findings of this table do not actually provide any particularly interesting information, but does provide some preliminary insight into the data. Factor analysis and cluster analysis greatly improve on these very basic preliminary findings

Principal Component	Eigenvalue	Explained Variance (%)	Cumulative Explained Variance (%)
1	25.718123	20.385663	20.385663
2	9.641927	7.642746	28.028409
3	4.961949	3.933126	31.961535
4	3.833868	3.038944	35.000479
5	3.491101	2.767247	37.767726
6	3.363625	2.666202	40.433928
7	2.967439	2.352163	42.786091
8	2.871874	2.276413	45.062504
9	2.598028	2.059346	47.121850
10	2.534947	2.009345	49.131195
11	2.412538	1.912317	51.043512
12	2.343259	1.857402	52.900913
13	2.213723	1.754724	54.655637
14	2.081128	1.649622	56.305259
15	1.924309	1.525318	57.830577

Table 2: This table contains the first 15 principal components out of 132 and their Eigenvalues, the explained variance of each of the principal components, and the cumulative explained variance of each principal component in addition to the ones that took place before them.

For my factor analysis, I use principal component analysis (PCA) instead of typical factor analysis.because normal factor analysis does not work on my data in Python. The most important difference between normal factor analysis and PCA is that the purpose of the former is simply to uncover the number factors, while the purpose of the latter is dimension reduction. Reducing dimensionality is useful in explaining how many components impact the decisions of delegates. Table 2 only has the first 15 of 132 components, but there are a total of 25 factors with an eigenvalue of more than 1, which would suggest that they are the "principal components" of the data. However, over 50% of the variation in the voting of delegates can be explained in the first 12 components, and 32% in the first 3. This suggests that there are roughly 3 strongly influential components effecting how delegates voted on certain issues in the 1973 Louisiana convention. This is an interesting preliminary result because it does suggest that there are more components influencing the votes of delegates than just party, ideology, or some combination of the two, which would probably have resulted in just one or two principal components. Figure 1 is a screeplot of the eigenvalues to also demonstrate which components are important. Using the "elbow rule" while analyzing the scree plot, it looks like the first three principal components are most important. This makes sense since there are probably lots of important dimensions, but not a huge number of dimensions that are extremely important, even if they do explain a notable amount of variation in the voting behavior of delegates.

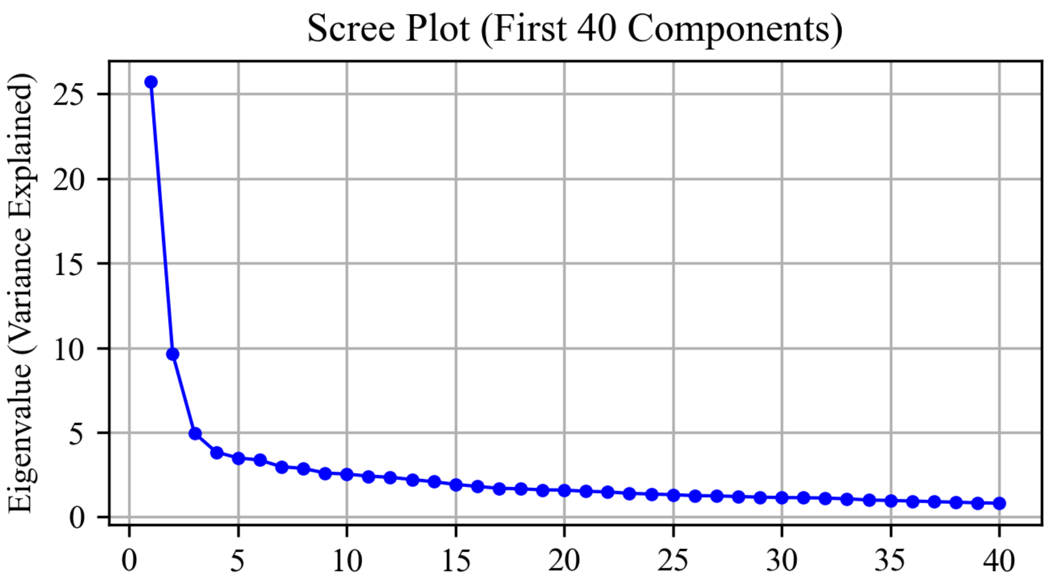

Figure 1: Scree plot showing the eigenvalues for the first 40 principal components.

Using the elbow method, I am able to determine that three clusters is the ideal number of clusters to use. Table 3 has a list of each of the delegates in the clusters they were assigned to. Cluster analysis seems to generally confirm what we saw with PCA. While it does not say anything about dimensionality, it does demonstrate that there are three groups of delegates that act in a similar way.

Table 3: Delegates Grouped by Clusters

Cluster 1	Cluster 2	Cluster 3
Abraham	Aertker	Alario
Alexander	Arnette	Bergeron
Anzalone	Asseff	Blair
Armentor	Avant	Bollinger
Bel	Badeaux	Champagne
Brown	Brien	Dunlap
Burson	Burns	Duval
Cannon	Carmouche	Fontenot
Chehardy	Casey	Gauthier
Colten	Chatelain	Gravel
Corne	Conino	Guarisco
De Blieux	Conroy	Hardee
Derbes	Cowen	Hayes
Deshotels	D'Gerolamo	Haynes
Fowler	Dennery	Jack
Ginn	Dennis	Jenkins
Guidry	Drew	Kelly
Henry	Edwards	Kilpatrick
Hernandez	Elkins	Landrum
Jackson, A.	Fayard	Newton
Jackson, J.	Flory	Rachal
Kilbourne	Fulco	Roy
Landry, A.	Giarrusso	Reeves
Landry, E. J.	Grier	Roemer
LeBleu	Juneau	Schmitt
LeBreton	Kean	Slay
Lennox	Lambert	Soniat
Mauberret	Lanier	Stagg
Munson	Leigh	Thompson
O'Neil	Leithman	Velazquez
Rayburn	Lowe	Vick
Robinson	McDaniel	Wall
Shannon	Martin	Warren
Silverberg	Miller	Wisham
Taylor	Mire	Cluster
Thistlewaite	Nunez	
Tobias	Ourso	
Triche	Perez	
	Perkins	
	Planchard	
	Riecke	
	Sandoz	
	Segura	
	Singletary	
	Smith	
	Stephenson	
	Stinson	
	Stovall	
	Sutherland	
	Tapper	
	Tate	
	Toca	
	Toomy	

Table continued on next page

Ullo
Vesich
Weiss
Willis
Winchester
Womack
Zervigon

Duncan MacRae utilized pairwise correlation and factor analysis in some of his earlier studies of roll call votes (see MacRae 1950, 1952, and 1960), but he eventually developed a type of scaling, building on Guttman scaling−another common type of method used for analyzing votes in congress (MacRae 1965). This cumulative scale of roll call voting is superior to factor or cluster analyses−according to MacRae−because it does not make as many assumptions "about linearity, cardinality, and ... the homogeneity of attitudes" (MacRae 1965, p. 910). This means that his scaling does not force a scholar analyzing data to make assumptions of what policy or ideological dimensions a factor analysis might demonstrate or how many clusters there are and what clusters are important in cluster analysis. Instead, MacRae's cumulative scaling technique simply allows scholars to determine how similarly legislators vote to one another in legislative bodies. I do not utilize MacRae's scaling technique here because it is much more complicated than W-NOMINATE (unlike basic factor and cluster analyses), is more difficult to understand in terms of its output, and has simply been replaced by W-NOMINATE in terms of what it does. I do provide a much more basic Guttman scaling of the data here though.

Using a Guttman scale requires binary data. I transform the data to binary data by transforming the yea votes to 1, nay votes, to 0, and treating all other values as missing values. Using Guttman scaling, my data has a Loevinger's H coefficient of 0.773. This suggest that the data fits a Guttman scale well. Substantively, in my analysis, this suggests that delegates clearly support one type of voting item and clearly oppose other types of voting items. While I cannot definitively say anything about how this score reflects partisan division, it does seem to suggest that delegates do follow some kind of general predictive pattern. A few issues with this method is that I am unable to determine dimensionality, or even be sure that partisanship or ideology are the driving forces of the decision making. An H coefficient of 0.773 suggests that, while a Guttman's scale fits the data, it does not fit it perfectly. In other words, some variation is caused by something other than the primary (undefined) dimension. MacRae's scaling would allow me to dive deeper into the relationship between delegates, but using scaling techniques has generally been replaced by DW-NOMINATE or related scaling techniques.

Together, pairwise correlation, factor analysis (PCA here), cluster analysis, and Guttman scaling all provide some useful insights. However, their interpretation is difficult to completely understand or use for other types of research empirically. This is where W-NOMINATE becomes useful. DW-NOMINATE is a roll call scaling technique developed by Poole and Rosenthal (1997) that allows scholars to visually analyze data in a readable graph. W-NOMINATE is a simpler version that is useful when looking at a single, unchanging, legislative session.[9] Essentially, it does combines the benefits of the different techniques I discuss above. It creates coordinates of where voters in a legislative body are in comparison to each other in multiple dimensions, thought the visual map is usually only two dimensional. In a typical NOMINATE map of congress, there are clear divisions between each party, with a significant amount of space in between. In that space, some delegates who tend to be more unorthodox in their voting habits may exist, but generally the division is quite clear. When partisanship breaks down, the division becomes less clear and W-NOMINATE becomes less useful. Because my preliminary work demonstrates that there is low dimensionality and that one or two dimensions explain a significant amount of the variance, W-NOMINATE is a very useful tool in analyzing roll call votes here.

Findings and Discussion

The W-NOMINATE roll call scaling developed by Poole and Rosenthal does not have a package that I could use in Python. However, there is a package in R that was developed, so I utilize this package for my analysis here. The first step in using the W-NOMINATE program is to transform the data from a csv file (what I exported from Python), into a matrix. This matrix drops the delegates names so it is just the roll call vote data. I also replace all -99 values (not voting items) with missing data. I then transpose the data so that delegates are the rows and

[9]Poole and Rosenthal (1997)use both DW-NOMINATE and W-NOMINATE, as W-NOMINATE is a bit simpler when there is no temporal aspect, i.e. when only observing one legislative session.

voting items are the columns. I do this because W-NOMINATE seeks to see scale the legislators, not the votes. After I finish cleaning and rearranging the data, I then simply plug the matrix of roll call votes into the W-NOMINATE package in R. W-NOMINATE generates coordinates for each delegate in two dimensions. Like the methods above, W-NOMINATE does not definitively say what the dimensions are. Theoretically however, it should be clear based on the map where party lines are, where there is a clear gap between two groups or parties who vote differently.

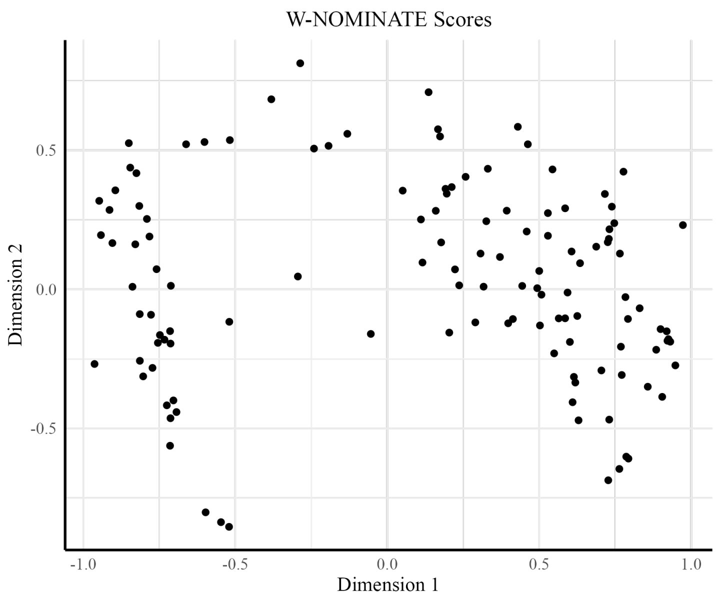

Figure 2: W-NOMINATE coordinates in two dimensions for delegates of the 1973 Louisiana State Constitutional Convention

Using the W-NOMINATE program on my data creates a series of coordinates in two different dimensions which I then plot in Figure 2 below. Because of the nature of the map and the coordinates, I am unable to perform any analyses that are more than observational here. However, as Figure 2 demonstrates observationally, there is a rather clear division between two groups on the first dimension. Figure 3 demonstrates this further, as the first dimension clearly is bimodal while the second dimension is mostly unimodal, suggesting that most division is only on the first dimension and not on the second dimension. These histograms suggest that the first dimension has a major division, while the second dimension does not have too much of a division, though there is a slight division. This means that the delegates are clearly divided into two groups in the first dimension−probably two parties or two ideologies, while the second dimension does not effect cut across parties.

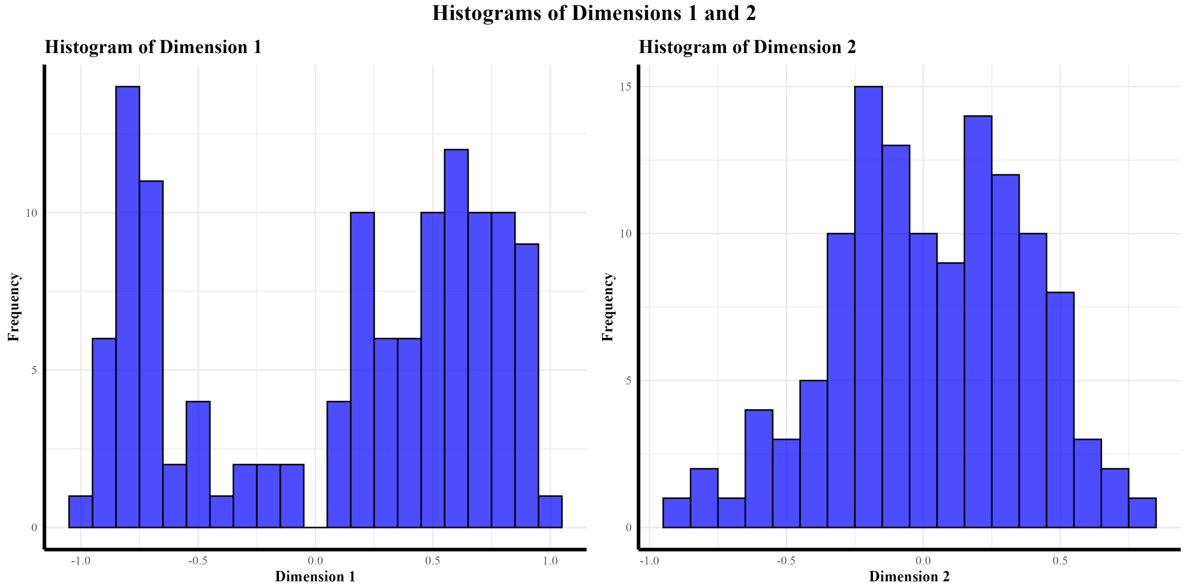

Figure 3: Histograms for each dimension of the W-NOMINATE coordinates.

While my analysis is purely observational, it is clear that partisanship or ideology (again, it is unclear) clearly heavily influences the voting behavior of delegates of the 1973 Louisiana constitutional convention. This is interesting theoretically for two reasons. The first reason is that delegates to the convention largely do not have the same types of incentives for reelection that members of normal legislative bodies have that would effect their behavior. The second is very related to the first and is that most delegates were not any type of elected official and do not have a clear party affiliation. That does not mean, of course, that they are not partisan, just not beholden to the political party the way that normal party members are. This suggests that there would be less division on the first dimension as well. However, this is clearly not the case and there is a strong division that reflects partisanship or ideological division between delegates.

Shortcomings

There are a lot of shortcomings in my analyses here. The most significant shortcoming is not related to the coding or my methods of analyses, but the OCR. Because the OCR does not translate the words of the archival records into usable text clearly, my data collection and pre-processing suffer greatly, which then leads to issues with my empirical analyses and findings. The first major issue with the OCR that I had to deal with is that there a significant number of votes that do not begin with YEAS in the OCR even thought all votes in the original source material clearly has them. This meant that I was not able to collect all the blocks of text that had votes in them early on. There are also issues with the splits at the NAYS, NOT VOTING, and ABSENT cut points. When these cut points did not exist, the voting item before it had the text lumped on to it. The NAYS and either or both NOT VOTING and ABSENT also appear in all votes in the original source material, but the OCR often caused issues with cutting at those words. I could deal with this issue buy performing my own OCR on archival materials or digitized pictures.

The second issue related to OCR that I dealt with was with the names of delegates in the voting items. Because the OCR was bad, I had to use fuzzywuzzy to tokenize the names of delegates from the list instead of just the names of delegates. This lead to my code not being able to recognize some of the names. Particularly difficult was those who shared the same last names. There were two pairs of delegates with the same last name, where their vote was noted with their last name and first initial. Because I used fuzzywuzzy, these delegates were completely dropped from my analysis. My code was unable to pick them up. There are definitely some delegates who the code did not pick up due to some kind of OCR errors, meaning that there is more missing data in my cleaned dataset than is actually in the data—even just looking at the votes that I was able to keep. A possible solution to this would be to create a machine learning algorithm that is able to recognize names that are not clean from OCR, but with 132 delegates, the number of times I would have to go over each name would be significant, and it is not quite worth it for the data I gather from these analyses

Finally, perhaps the most notable shortcoming of my analyses here is that I utilize more traditional statistical analyses, not machine learning methods. This means that I am able to recognize current patterns from existing data to make interesting observations about certain events, however, the generalizability and predictive power of what I do here is not very strong. I only have a small number of observations and features. Perhaps creating a larger dataset with more roll call votes from different conventions would be useful. Perhaps an area where machine learning could be useful in this is in predicting what the voting items are, based on the text that I cut. Because I do not care about the subject of votes in this project and only focus on partisanship and ideology, I do not do much with those datapoints, but they could be useful and help with future analyses of different legislative session or conventions that use archival data.

Conclusion

In this paper, I explain how I develop a series of code that extracts roll call votes from a couple of volumes of archival records from the 1973 Louisiana state constitutional convention and then use that data to create a spatial model of the voting of delegates to the convention. In order to do this, I extract unstructured text data of the votes, tokenize the unstructured data, and then transform it into a dataset where the rows are voting items, the columns are delegates, and the cells are how delegates voted on a voting issue. I performed this data collection and pre-processing in Python. After creating the dataset, I performed some preliminary analyses using pairwise correlation, factor analysis (PCA), cluster analysis, and an analysis using a Guttman scale. I then move the dataset to R, where I use the W-NOMINATE package in R to create a spatial model. Using all of these analyses, I find (observationally, not statistically) that, despite not having the same incentives or party pressures, partisanship is clearly a major factor in the voting behavior of the delegates at conventions. The analyses I do here are far from perfect. There are significant OCR problems that lead to difficulty extracting and transforming the data. These issues lead to major issues in my empirical analyses. In fact, I lose 9 entire delegates when I use W-NOMINATE on my data because there is so much missing data due to bad OCR. However, because there are 132 delegates and nearly 900 roll call votes, W-NOMINATE is still able to function very well with the data that does exist. The coordinates of delegates at the convention would change slightly if I had less missing data, but probably not so much that the basic observational findings would change. In other words, the specific coordinates of delegates might change, but the overall structure of the spatial model I create would not change. I could build on this paper in the future by performing own OCR on archival materials that would be more accurate, or by analyzing more state constitutional conventions. I could also use the general data extraction model on historical legislative bodies, or adjust it so that it could be used in non-American legislative bodies.

Works Cited

Barnidge, James L. “The Louisiana Constitutional Convention of 1973: The Road to Revision.” *Louisiana History: The Journal of the Louisiana Historical Association* 15, no. 1 (1974): 35–48.

Clausen, Aage R., and Carl E. Van Horn. “How to Analyze Too Many Roll Calls and Related Issues In Dimensional Analysis.” *Political Methodology* 4, no. 3 (1977): 313–31.

Cox, Gary W., and Matthew D. McCubbins. Legislative Leviathan: Party Government in the House. Cambridge: Cambridge University Press, 2007.

Jackson, John E. “Statistical Models of Senate Roll Call Voting.” *The American Political Science Review* 65, no. 2 (1971): 451–70. https://doi.org/10.2307/1954461.

Jameson, John Alexander. *A Treatise on Constitutional Conventions: Their History, Powers, and Modes of Proceeding.* Callaghan, 1887.

MacRae, Duncan. “A Method for Identifying Issues and Factions From Legislative Votes.” *The American Political Science Review* 59, no. 4 (1965): 909–26. https://doi.org/10.2307/1953213.

———. “Direct Factor Analysis of Sociometric Data.” *Sociometry* 23, no. 4 (1960): 360–71. https://doi.org/10.2307/2785690.

———. “Roll Call Votes and Leadership.” *The Public Opinion Quarterly* 20, no. 3 (1956): 543–58.

———. “The Relation Between Roll Call Votes and Constituencies in the Massachusetts House of Representatives.” *The American Political Science Review* 46, no. 4 (1952): 1046–55. https://doi.org/10.2307/1952111.

Mayhew, David R. “Congressional Elections: The Case of the Vanishing Marginals.” *Polity* 6, no. 3 (1974): 295–317. https://doi.org/10.2307/3233931.

Poole, Keith T., and Howard Rosenthal. “A Spatial Model for Legislative Roll Call Analysis.” *American Journal of Political Science* 29, no. 2 (1985): 357–84. https://doi.org/10.2307/2111172.

———. *Congress: A Political-Economic History of Roll Call Voting*. Oxford: Oxford University Press, 1997.

Willetts, Peter. “Cluster-Bloc Analysis and Statistical Inference.” *The American Political Science Review* 66, no. 2 (1972): 569–82. https://doi.org/10.2307/1957800.