

# Measuring institutional variation across American Indian constitutions using automated content analysis

Rebecca Cordell      Kristian Skrede Gleditsch      Florian G Kern      Laura Saavedra-Lux  
University of Texas at Dallas      University of Essex      University of Essex      University of Essex

*Published in: Journal of Peace Research, <https://doi.org/10.1177/0022343320959122>*

## Abstract

Effectively measuring variation in institutions over time and across jurisdictions is important for examining how institutional characteristics shape political, social and economic issues. We present a new dataset of American Indian and Alaska Native (AIAN) constitutions and a new approach for measuring variation in polities using machine learning techniques. Existing data on AIAN institutions have largely been based on costly and time-consuming expert coding and survey approaches, where the end product will become obsolete once institutions change. Our automated content analysis of AIAN constitutional documents allows for more flexible and customizable measurement of the variation, using a larger corpus of data than existing approaches, limited by data collection and coding costs. We consider variation in judicial institutions, previously shown to play a crucial role in AIAN development, and compare our machine-coded measures to existing hand-coded data for a sample of 97 American Indian constitutions. We show that machine coding replicates expert coded data. Our approach can be easily extended to other topics, including the executive, and shows the potential of automated measures to complement or confirm traditional coding of political institutions.

**Keywords:** constitutions, institutions, content analysis, American Indian and Alaska Native nations

**Corresponding author:** [Rebecca.Cordell@utdallas.edu](mailto:Rebecca.Cordell@utdallas.edu)

## Introduction

Effectively measuring variation in institutions over time and across jurisdictions is important to scholars who study how institutional characteristics shape political, social and economic issues including development, the rule of law, and change in political institutions. We present a new dataset of American Indian nations constitutions and polities to contribute to the discussion on measuring institutions. Political scientists have long neglected the study of American Indian and Alaska Native (AIAN) nations (Ferguson, 2016), even though the pressures these communities face reflect important issues including poverty and inequality. US 2014 census estimates indicate that 28.3% of single-race AIAN individuals lived in poverty; ‘the highest rate of any race group’.<sup>1</sup> Variation in AIAN polities can provide unique information on how the development of political institutions can shape local political, social and economic outcomes.

Previous studies have investigated how AIAN institutions influence various governance outcomes (e.g. Akee et al., 2015; Cornell & Kalt, 1995a, 1995b, 2000; Dippel, 2014; Gover, 2010). However, they have generally relied on measures of institutions developed by costly and time-consuming expert coding and surveys. Crucially, such data are static and will become outdated if institutions change; a serious concern as many AIAN nations have pursued reforms. The California Valley Miwok Tribe, for example, has amended its constitutions at least six times since 1999.

We present an original digitalized database of American Indian constitutions. Our data allows us to code and measure institutional characteristics directly from the content of constitutions. We demonstrate the value of this data, using a text-as-data approach to analyze the content of 97 American Indian constitutions. We use a dictionary approach and supervised

---

<sup>1</sup> [https://www.census.gov/content/dam/Census/newsroom/facts-for-features/2015/cb15-ff22\\_AIAN\\_month.pdf](https://www.census.gov/content/dam/Census/newsroom/facts-for-features/2015/cb15-ff22_AIAN_month.pdf)

machine learning to construct institutional measures of AIAN polities. We focus on judiciary institutions as this is considered essential for AIAN nations' stability and development (Cornell & Kalt, 2000).

Our analysis demonstrates that machine coded approaches applied to constitutional texts recover well the results from expert coding of judiciary functions and guarantees of judicial independence. Our approach allows for more comprehensive measures of AIAN institutional variation, using a larger corpus than existing approaches limited by high data coding costs. The use of automated approaches to analyze the content of AIAN constitutions is fully reproducible, analyses can be updated whenever the underlying documents are amended or new information becomes available, and the analytical approach can easily be expanded to other topics in constitutions including the executive. Our analysis highlights the general value of using automated approaches to measure political institutions. Although expert surveys and hand coding remain valuable for identifying institutional characteristics, automated measures are a viable alternative and supplement to traditional coding practices. This is reinforced by Marquardt (2020) and Krüger and Nordås (2020) in this volume, who discuss and propose solutions to improve on possible measurement error, given the propensity for expert coders to disagree or diverge.

## **Comparative perspectives on constitutions**

### *Worldwide constitutions*

There is a large body of work on measuring political institutions using constitutions and their effects. Tsebelis & Nardi (2014) argue that longer country constitutions are associated with lower GDP per capita because detailed constitutions prevent governments from adopting measures to combat economic shocks. Moreover, detailed constitutions are amended more frequently because

they are more likely to contain restrictive provisions that impede the government's ability to pass necessary or desired legislation.

Focusing on judicial institutions in constitutional texts, Melton & Ginsburg (2014) analyze the relationship between de jure and de facto judicial independence. They find that constitutions that outline the selection and removal of judges enhance judicial independence in practice, using the Comparative Constitutions Project data. Hayo & Voigt (2007) similarly argue that judicial independence vested in formal rule enhances effectiveness, as measured by average term length of judges, number of judges, constant income of judges, changes to legal rules, and statements on degree of judicial independence.

#### *AIAN constitutions*

A number of studies examine how constitutional characteristics and institutions affects governance, policy, and social and economic outcomes in AIAN nations. Cornell & Kalt (2000) argue that self-governing AIAN institutions and 'cultural match', or the closeness of adopted political institutions to indigenous cultural norms, are key to economic development. They classify whether there is an independent judiciary and whether the government is organized under a strong chief executive, a strong legislature, or a general council for 70 AIAN constitutions (Cornell & Kalt, 1995b: 17). They find that AIAN nations with a strong-chief-executive and strong-legislatures display higher economic development than nations with general councils, which they attribute to the greater separation of powers. Additionally, Cornell & Kalt (1991) show that nations with an independent judiciary and directly-elected chief executive have almost 20% higher employment than those without an independent judiciary or those governed by general council.

The cultural match of constitutional design matters for economic development because of perceptions of legitimacy of government, as suggested in Cornell & Kalt's (1995a) qualitative analysis of the White Mountain Apache of the Fort Apache Reservation in Arizona and the Oglala Sioux Tribe of the Pine Ridge Reservation in South Dakota. Even though both nations have adopted constitutions based on the 1934 Indian Reorganization Act (IRA), their pre-constitutional governmental forms differ considerably regarding the structure, scope, location and source of political authority, with the Apache nation having greater cultural match to their constitution and better levels of economic development than the Sioux.

Akee et al. (2015) explore the institutional differences between AIAN nations who adopted tribal constitutions during Republican versus Democrat presidential administrations and the long-term economic effects. They code constitutions of 70 American Indian nations based on whether the executive is indirectly elected, the tribal council is directly elected, an independent judiciary exists, staggered elections can be found, and the term length or size of tribal council. Nations that adopted constitutions under a Democrat administration tended to implement an indirectly elected (i.e. parliamentary) chief executive, and nations that adopted constitutions under a Republican administration tended to implement a directly elected (i.e. presidential) chief executive – with higher economic development under the former.

These studies shed valuable light on AIAN constitutional texts, but the measurement coding procedure is costly and time-consuming, and it is difficult to replicate the coding decisions for individual constitutions. In the following, we explore how a text-as-data approach can complement existing approaches.

## **AIAN constitution making**

AIAN constitutions embody the foundational principles of AIAN nations, defining the structure, functions and authority of their institutions and branches of government (Flies-Away et al., 2007). Miller (2011: 17) argues that ‘a close examination of [AIAN] constitutions demonstrates the different powers that tribal governments exercise under their constitutions and the different ways in which tribal communities have chosen to create and control their governments’. There is considerable variation in both the content and origin of contemporary AIAN constitutions.

Some AIAN nations already maintained constitutional documents before the 1934 IRA, but many adopted constitutions based on an outline of topics (Rusco, 2000: 305n7) or model constitution (Wilkins, 2004: xxiii), provided by the Bureau of Indian Affairs (BIA).<sup>2</sup> The drafting of IRA-era constitutions reflected an attempt by the US government to decrease federal control of American Indian affairs and to increase self-governance.

The IRA intended to provide American Indian nations ‘with the modern governments that would allow them to participate more fully in democratic, American-style political processes’ (Lemont, 2006: 19). IRA constitutions typically provide for a representative tribal council with legislative powers, tribal chairperson directly or indirectly elected, little or no provision for judicial institutions/functions, little enumeration of powers of the various parts of government, and a requirement that the BIA approve any changes (Cornell & Kalt, 1995b). American Indian nations were given two years to vote on the act’s provisions, with 181 tribes accepting the act and 77 tribes

---

<sup>2</sup> Cohen (1971) notes that the Department of Interior had over 60 AIAN nations’ constitutions on record before 1934. Since many constitutions reflect IRA outlines, the documents may share similar content, even if many have subsequently been revised.

rejecting it (Wilkins & Kiiwetinepinesiik Stark, 2011).<sup>3</sup> It is estimated that 35-40% of AIAN nations now govern under an IRA constitution, with the remainder having a non-IRA constitution in place or operating without a written constitution (Miller, 2011).

Following the adoption of the Indian Self-Determination and Education Assistance Act in 1975, AIAN nations were provided greater opportunities to manage local government and local issues in law and practice (Akee et al., 2015). Many AIAN nations have subsequently revised initial constitutions to be ‘both practically effective in meeting the challenges and circumstances the nation faces and culturally legitimate in the eyes of the people’ (Kalt, 2007: 109).

### **AIAN constitutional database and sample**

We have collected and digitalized a machine-readable database of AIAN constitutional documents. This corpus allows for systematic analysis of the content and sources of constitutional change, including the extent to which AIAN nations are able to adapt constitutions depending on specific challenges or circumstances, or revise the content of IRA-era constitutions.

The availability of AIAN constitutions, amendments and codes varies considerably. Many documents are available through the National Indian Law Library (NILL), although not all are digitalized. Gover’s (2010) collection also helped us locate past and present constitutional source material. We retrieved newer constitutions from the official AIAN websites where available. Yet, not all tribes share their constitutional texts, and not all tribal jurisdictions actually have constitutions. Additionally, we ordered and scanned many documents from different libraries. Our

---

<sup>3</sup> American Indian nations in Oklahoma and Alaska Native nations were originally excluded, but the 1936 Alaska Act and the Oklahoma Indian Welfare Act extended the IRA to all federally recognized AIAN nations (Wilkins & Kiiwetinepinesiik Stark, 2011).

AIAN constitutional database includes over 2,700 tribal constitutions, amendments, codes, treaties, and further legal documents.

In this study we only consider AIAN polities with constitutions, American Indian constitutions in the lower 48 US states, and do not include Alaska Native and Hawaii indigenous nations. We exclude secondary codes or laws. As of February 2019, there were 573 federally recognized tribal entities in the United states, 342 within the contiguous 48 states (and 231 in Alaska), with some tribes jointly administering jurisdiction (Bureau of Indian Affairs, 2019). We focus only on *active* constitutions in 2010 before adoption of the Tribal Law and Order Act of 2010. This law enabled tribes to adopt greater sentencing powers and may have triggered changes that would make comparisons to previous analyses of AIAN constitutions challenging. This leaves us with a sample of 97 American Indian constitutions.

Figure 1 displays the adoption year of constitutions in our sample. The oldest active constitution, approved in 1935, is the Santa Clara Pueblo, New Mexico, and the most recent constitution, approved in 2010, is the Absentee-Shawnee Tribe of Oklahoma, who we record as amending their original constitution twice. Most constitutions in our sample were adopted during the 1980s through to the 2000s (67 constitutions, 69%) – with the greatest number adopted in the 2000s (31 constitutions, 32%), and the lowest number adopted in the 1950s and 2010 (3 constitutions respectively, 6%).

---

Figure 1 in here

---

## **Methodology and data**

### *Automated content analysis*



Recent innovations in machine learning and increased digitalization of texts have made automated text analysis widely used in political science. These methods help researchers analyze large corpuses of texts and provide opportunities to replicate human coding in a more systematic and timely manner. Computerized approaches can be used to estimate actors' ideal points from political texts or organize and classify documents into pre-defined or new categories (Grimmer & Stewart, 2013). For example, Laver et al. (2003) and Slapin & Proksch (2008) use word frequencies in party manifestos to estimate location on a policy spectrum. Owens & Wedeking (2011) and Benoit, Munger & Spirling (2019) develop approaches to measure the complexity of texts such as Supreme Court opinions using specific dictionaries and human pairwise judgment data that capture cognitive complexity, sophistication and difficulty. Scholars have used a variety of text sources, including online blogs, social media data, speeches and newspaper articles (Hopkins & King, 2010; Schrodtt & Brackle, 2013; King et al., 2013; Jamal et al., 2015).

To our knowledge, we are the first to apply machine learning techniques to AIAN constitutional texts. Spirling (2012) uses a text-as-data approach to analyze treaties between Native American nations and the US government from 1784-1911 and the impact of institutional change and relative power. He finds that treaties became harsher overtime, particularly during the mid-1820s when the US gained power and during war when the US government gained military strength that facilitated stronger demands.

### *Text data processing*

To pre-process the documents, we converted all AIAN constitutions to plain text files (UTF-8). We convert all text to lower case and remove unnecessary white space, punctuation, encoded text and non-ascii characters, and stem the text using the Porter (1980) algorithm to reduce words to

their base root form (e.g. ‘judges’ is condensed to ‘judg’). This helps to remove aspects of the text irrelevant to our analysis and avoid over-counting the same terms and concepts (e.g. capitalized and stringed variations of words). We use this text for calculating the word count of terms included in our dictionaries. To improve the efficiency of our supervised machine learning model, we remove all numbers and stop words (e.g. ‘are’, ‘at’, ‘by’, ‘from’, ‘that’, ‘the’).

## **Exploring American Indian constitutions**

### *Constitution length*

We first compare the total length of constitutions (i.e. total word count). For this purpose, we create a document term matrix (DTM) that counts the number of times a term appears in each of the constitutions. The columns in the DTM correspond to the unique words and the rows correspond to the American Indian constitution. We then take the sum of each row to calculate the total word count of the constitution.

The constitutions in our corpus vary considerably in overall length. For example, the St. Croix Chippewa Indians of Wisconsin has the shortest constitution in our sample (approved in 1942) with only 1,423 words, seven articles and three bylaws. By contrast, the 1992 Menominee Indian Tribe of Wisconsin constitution is nearly thirteen times longer, with 18,406 words based on 20 articles and six bylaws. The average length of a constitution in our sample is 5,410 words. Figure 2 displays the distribution of constitution length in our sample.

---

Figure 2 in here

---

Figure 3 displays the average word count of constitutions by decade adopted. On average, constitutions adopted from 1980-2010 were 28% longer than constitutions adopted during the earlier years in our sample. In general, we observe a steady increase in average length over time – with constitutions produced in 2010 on average 20% longer than constitutions produced during the 1930s. However, this growth is not entirely consistent as constitutions adopted during the 1930s in our sample were on average 29-40% longer than those adopted during the 1940s and 1950s. Using the Wilcoxon signed-rank test, we find a statistically significant increase in the average word count from constitutions produced over 1935-1969 to those produced during 1976-2010 at the 99% confidence level. This suggests that the length of constitutions matters and length of newly adopted American Indian constitutions has increased over time. However, the change in average word count is only statistically significant at the 99% confidence level for constitutions produced during the 1930s and 1940s.

---

Figure 3 in here

---

### *Complexity of judicial institutions*

We measure the depth and breadth of judicial laws via two methods. First, we measure the length of a constitution's section on the judiciary by segmenting this section from the main body. We then create a DTM that counts the number of times a term appears in each of the constitutions and take the sum of each row to calculate the word count for each judiciary section. Second, we build a dictionary of terms associated with the judiciary and create a DTM that counts the number of times a term in the dictionary appears in the constitution and take the sum of each row (see Table

I).<sup>4</sup> We select our terms using existing dictionaries that define and provide synonyms for the term ‘judiciary’ and include any relevant terms from our corpus that contain their base root form (e.g. ‘law’, ‘bylaw’, ‘unlaw’). To ensure that these terms are relevant to the discussion of AIAN judicial institutions and account for variation in language and topics across constitutions, we only include those terms that feature in multiple constitutions in the sample.

---

Table I in here

---

There is considerable variation in the judicial section length of constitutions in our corpus. For example, 52% of constitutions do not have a dedicated section on the judiciary. The longest section on the judiciary in our sample is the Turtle Mountain Band of Chippewa Indians of North Dakota’s constitution (approved in 2006) which consists of 1,766 words – containing information on the purpose of the judiciary, the establishment, judicial powers, the selection of judges, the judicial board, implementation and saving clause and reservation powers of the tribal council. The tribe with the shortest judicial section in their constitution (excluding tribes who do not have one) is the Lower Sioux Indian Community in Minnesota (approved in 2007) at 46 words. The information included in this section simply states that the council establishes a court of domestic relations to handle all misdemeanors, disputes and maintenance of law and order on the reservation. Figure 4 displays the distribution of judicial section length (i.e. word count) in our sample across tribes, with a mean of 231 words.

---

<sup>4</sup> For alternative measures of textual complexity see Owens & Wedeking (2011) and Benoit et al. (2019).

---

Figure 4 in here

---

To account for the possibility that an AIAN nation does not have a specified judicial section but mentions the judicial function somewhere else in the constitution, we explore the word count of terms associated with our judiciary dictionary. We observe substantial variation. For example, for the St. Croix Chippewa Indians of Wisconsin, the terms included in our judiciary dictionary appear only 8 times in the constitution (approved in 1942). The Menominee of Wisconsin unit has the greatest amount of terms associated with the judiciary in the constitution (approved in 1992), with terms in dictionary appearing 286 times. Figure 5 displays the distribution of the number of terms included in our judicial dictionary appearing in a constitution with a mean value of 57.

---

Figure 5 in here

---

### *Judicial independence*

We measure levels of judicial independence by building a dictionary of terms associated with an independent judiciary including statements on judicial independence, judicial authority, judicial tenure, the selection and removal of judges, and salary insulation (see Table II). We construct our dictionary by selecting terms that correspond to the measures of judicial independence presented by Melton & Ginsburg (2014) (developed by the Comparative Constitutions Project) and Hayo & Voigt (2007). Namely, judicial independence, judicial tenure, selection procedure, removal procedure, limited removal conditions, and salary insulation. We then create a DTM that counts the number of times a term in the dictionary appears in the judiciary section of an American Indian constitution and take the sum of each row.

---

Table II in here

---

We again find significant variation on this measure in our sample. Since 52% of the constitutions in our sample do not have a dedicated section in their constitution on the judiciary, this group of tribes score a 0 on this measure. The Turtle Mountain Band of Chippewas of North Dakota includes the greatest number of references to an independent judiciary (with terms in our judicial independence dictionary appearing in the judicial section 93 times). Their constitution approved in 2006 mentions the term ‘independ’ 3 times, ‘elect’ 24 times and ‘term’ 11 times. In comparison, the tribes with the least amount of terms included in our judicial independence dictionary appearing in a judicial section (excluding tribes which do not have one) is the Lower Sioux Indian Community in Minnesota and the Tule River of the Tule River Reservation in California. The only term included in our dictionary that appears in their constitution approved in 2007 and 1974 is ‘author’ just once. Figure 6 displays the distribution of the number of terms included in our independent judicial dictionary appearing in the judicial section of a constitution, with a mean value of 9.

Unpacking the constitutions qualitatively helps illustrate the validity of our measure. If we compare constitutions which are in the 80th percentile of our judicial independence measure with those at the 60th percentile, a clear difference becomes apparent in how detailed judiciary independence is specified. The constitutions in the 60th percentile contain topics in their judiciary section such as definition of the establishment of judicial institutions, selection procedure and

number of judges.<sup>5</sup> Constitutions in the 80<sup>th</sup> percentile contain all of those topics, but also include topics such as salary insulation, limitations to removal procedures, and the power for judges to review and hold to account the actions of the governing council.<sup>6</sup>

---

Figure 6 in here

---

Table A1 (see Online appendix) illustrates three American Indian constitutions that our judicial independence dictionary model identifies as having some of the least, average, and most references to an independent judiciary in the judicial section of their constitution. The Yankton Sioux Tribe of South Dakota has a constitution with the least information on the judicial function (except for a short passage relating to the selection procedure of judges). The Turtle Mountain Band of Chippewa Indians in North Dakota has a constitution with the most topic coverage on the structure of the judiciary (except on the topic of salary insulation). Table A1 also demonstrates the substantive differences in institutional design. For example, the Mescalero Apache Tribe in New Mexico selects judges through the judicial council (indirectly elected) and the Turtle Mountain Band of Chippewa Indians in North Dakota selects judges through a general population vote (directly elected).

There is striking variation on section length, judiciary dictionary and judicial independence dictionary measures. While unfortunately outside the remit of this article, these measures will provide researchers with new opportunities to explore how various social, economic and political

---

<sup>5</sup> For instance, constitutional documents of the Chippewa Cree Indians of the Rocky Boys Reservation, the Sisseton-Wahpeton Oyate, or the Lower Brule Sioux Tribe of the Lower Brule Reservation.

<sup>6</sup> For instance, constitutional documents of the Ely Shoshone Tribe, or the Sac and Fox Nation.

outcomes are shaped by the structure and content of AIAN constitutions, which previous research suggests matter greatly (e.g. Cornell & Kalt, 2000, and Akee et al., 2015).<sup>7</sup>

### *Supervised machine learning*

We use a supervised machine learning model to produce alternative measures of the complexity of judicial institutions and levels of judicial independence, trained on a labeled set of human coded American Indian constitution data by Cornell & Kalt (2000). Their data includes a set of constitutional variables for 64 American Indian nations, hand coded using an archive of American Indian constitutions acquired from the BIA in 1990. The two constitution measures that are most relevant to our analysis include the JUDICIARY variable (i.e. whether there are references within the constitution on the structure of judicial institutions) and the INDEPEND variable (i.e. whether there are guarantees within the constitution on the independence of the judiciary).

There is an overlap of 38 constitutions in both Cornell & Kalt's (2000) data and our current sample. We use three different supervised machine learning algorithms (Support Vector Machine (SVM), Naïve Bayes and Logistic Regression) to classify which constitutions in our corpus contain information on the structure of judicial institutions and an independent judiciary using the training data DTMs for the 38 overlapping constitutions.<sup>8</sup> We train the models on 70% of the training data using repeated k-fold cross validation and evaluate the accuracy of our model using the remaining

---

<sup>7</sup> Table A2 (see Online appendix) displays the Pearson correlation coefficients for every paired combination of our dictionary measures. The weakest (positive) correlation is between constitution length and independent judicial dictionary measure at 0.48. The constitution length and judicial section length are also weakly (positively) correlated at 0.52. This suggests that there are substantial differences between the aggregate amount of information included in a constitution (total length) and the topics themselves (judicial institutions).

<sup>8</sup> For the Logistic Regression, we use a boosted algorithm that performs variable selection to improve the accuracy of the model.



30% of the data.<sup>9</sup> This procedure splits our training data into 5 k-subsets, with each subset removed from the sample and trained on all other subsets, repeated 5 times. Each model finds patterns in the training data that relate the word count of terms contained in a constitution to Cornell & Kalt's (2000) binary variables JUDICIARY and INDEPEND. Constitutions that the model classifies as containing information on the structure of the judiciary and judicial independence are assigned the value of 1, and 0 otherwise. We select the algorithm that produces the greatest accuracy on the test data to assign binary variables to the 97 constitutions in our corpus.

### **Evaluating machine coding and human coding approaches**

To evaluate our text-as-data approach and the potential for automated coding procedures to be used to analyze the content of American Indian constitutions, we compare our machine coded measures to an existing alternative: Cornell & Kalt's (2000) human coded American Indian constitution data (see Fariss Kenwick and Reunig 2020 for a similar approach to assessing model validation). First, we use our dictionary approach to replicate the hand coding procedure of Cornell & Kalt (2000) for their JUDICIARY and INDEPEND binary variables that measure whether a constitution includes information on the structure of the judiciary and judicial independence. Second, we use a supervised machine learning method to replicate the Cornell & Kalt (2000) variables by training a SVM, Naïve Bayes and Logistic Regression model on 70% of the Cornell & Kalt (2000) data.

For the dictionary JUDICIARY variable, we follow Cornell and Kalt's (2000) coding rules and create a dummy variable that codes constitutions 1 if the following terms from our judiciary

---

<sup>9</sup> We select a 70/30 split to better evaluate the precision and recall of the model on our test data as it leaves more constitutions in the test data.

dictionary are mentioned, and 0 otherwise: ‘judici’, ‘judg’, ‘prejudici’, ‘judiciari’, ‘adjudg’, ‘adjud’, ‘quasijudici’ and ‘court’.<sup>10</sup> This is the case for all 38 constitutions in our sample that overlap with the Cornell & Kalt (2000) data. For the dictionary INDEPEND variable, we again follow Cornell & Kalt’s (2000) coding rules and create a dummy variable that codes constitutions 1 if the term ‘independ’ is mentioned in the judicial section of the constitution, and 0 otherwise.<sup>11</sup> This is the case for only 5% of constitutions in our sample (2), with the remaining 95% not including the term ‘independent’ in the judicial section of the constitution (36).

To assess the accuracy of our dictionary approach and supervised machine learning models, we report the percentage of correctly classified constitutions, the precision (the ratio of correct positive predictions to the total predicted positives) and recall (the ratio of correct positive predictions to the total number of true positives and false negatives) estimates. The accuracy of our supervised machine learning models is determined by taking the mean accuracy over each of the 5 k-sub-samples (for 30% the training data held out for training).

These approaches illustrate how automated coding methods can be used to replicate the human coding of American Indian constitutions in the future. While there is a 20-year difference

---

<sup>10</sup> For the purpose of replicating Cornell & Kalt’s (2000) JUDICIARY variable, we exclude the following terms from our judiciary dictionary that do not directly relate to the structure of judicial institutions: ‘justic’, ‘bylaw’, ‘law’, ‘unlaw’, ‘lawandord’ and ‘jurisdict’. However, we obtain the same level of agreement when we account for all terms in our judiciary dictionary as constitutions that mention the structure of the judiciary tend to include more general terms related to the judiciary.

<sup>11</sup> For the purpose of replicating Cornell & Kalt’s (2000) INDEPEND variable, we do not include any additional terms from our judicial independence dictionary as the focus of their coding is restricted to whether there are guarantees within the constitution on the independence of the judiciary. Accordingly, when we account for all terms in our judicial independence dictionary, we obtain lower levels of agreement as there are many more constitutions in our sample that include terms associated with an independent judiciary as opposed to explicitly describing the judiciary as ‘independent’.

between our corpus and the constitutions that Cornell and Kalt (2000) analyzed, our sample includes many constitutions that have not been amended during this time period – and the amendments that have been made do not always pertain to the judiciary.

The results indicate that our dictionary approach achieves high levels of accuracy, assigning the same classification as Cornell & Kalt (2000) to between 82-84% of constitutions in the overlapping sample (see Table III). Our supervised machine learning method also obtains high levels of accuracy, assigning the same classification as Cornell & Kalt (2000) to on average 90% of constitutions in 30% of the test data (see Table IV and VI). Both approaches have high levels of precision and recall for the JUDICIARY variable (1 and 0.842 for the dictionary approach and 1 and 0.9 for the supervised machine learning models).

However, the precision and recall of the INDEPEND variable produced by both methods is low as neither approach is able to correctly classify the true positives identified by Cornell & Kalt (2000) in the overlapping sample (i.e. constitutions that they code as containing references to an independent judiciary). For the dictionary approach, this is because our search for the term ‘independ’ is too narrow and only features in a small number of constitutions in the overlapping sample. For example, when we search for all terms included in our judicial independence dictionary, the precision and recall improve (correctly identifying 80% of the true positives identified by Cornell & Kalt 2000) but the accuracy decreases as our models assign the value of 1 to constitutions that Cornell & Kalt (2000) coded 0. This is likely because the content of these constitutions regarding an independent judiciary have been updated and changed over the past 20 years (since the Cornell & Kalt dataset is based on American Indian constitutions acquired in 1990 and our sample includes constitutions up to 2010). For the supervised machine learning method, these results are likely due to the small number of constitutions assigned a 1 by Cornell & Kalt

(2000) that the model is trained on (only 5 constitutions) and similarities between those constitutions in the training data assigned a 1 and 0. Both of these results call for more recent labeled data to compare our judicial independence measures that could be built using our larger AIAN constitutional database and hand-coding approaches.

---

Table III in here

---

---

Table IV in here

---

---

Table V in here

---

These results demonstrate the potential for automated approaches to recover human-coded measures and also complement and improve on human coding. Our measures assign a different classification to Cornell & Kalt (2000) to some constitutions where content has been added that mentions terms related to the judiciary in the main body of the constitution and ‘independence’ in the judicial section. In these cases, we find references to the judiciary and judicial independence in a more recent constitution for the same tribe that did not exist in an older constitution that Cornell & Kalt (2000) coded. Another reason for disagreements in coding of the same tribe’s constitution is due to the restrictive nature of our JUDICIARY and INDEPEND replication variables. For example, if we only search for the term ‘independ’ we will miss other terms that refer to the judicial function and judicial independence (that human coders can detect). Fortunately, our dictionary and supervised machine learning method are able to overcome this limitation and assign correct coding as they will consider multiple terms associated with these concepts. These

results show the potential for automated approaches to code American Indian constitutions and validate Cornell & Kalt's (2000) coding scheme. These findings are impressive given that our dummy variables used for this model evaluation exercise truncate the variation contained in our original measures, yet still capture something akin to existing alternatives.

## **Conclusion**

There is demand for better data and measurement of institutions, and this article shows how a text-as-data approach can be employed to capture variation across American Indian institutions. Existing approaches of measuring political institutions relying on hand-coding or expert surveys are time consuming and specific to the time when coded, which is problematic when the institutions are evolving. We show that content analysis provides a feasible alternative that can replicate existing expert-based coding and be easily extended as institutions are revised. Existing approaches and text-as-data design need not be seen as alternatives but can complement each other and help make data coding more reliable and replicable. Moreover, the methods presented in this article can easily be extended to other topics in constitutions; highlighting the value of these methods for measuring political institutions more generally.

Our database of the constitutions and politics across AIAN nations and the automated coding and estimation of institutional variables will allow us to further explore the role that constitutions play in shaping various governance outcomes. First, scholars could evaluate what determines the content of institutional provisions in constitutions. For our sample, American Indian constitutions reflect the public and political discourse and context of the time, and one might explore whether more recently adopted AIAN constitutions make greater reference to topics such as gender equality, rights for the elderly, and special needs groups. Second, scholars may also ask

about the effects of constitutional texts on political, social, and economic outcomes. In the spirit of the Cornell & Kalt (2000) and Akee et al. (2015) studies, scholars can investigate how AIAN constitutions' properties affect outcomes such as violent crime, for which judiciary institutions are likely to be particularly relevant.

### **Replication data**

The dataset and replication files for the empirical analysis in this article can be found at

<http://www.prio.org/jpr/datasets>.

### **Acknowledgements**

We thank the National Indian Law Library for providing documents to add to our database and Nihan Ari for assistance with transcribing the constitutions. We thank Joseph Kalt for kindly sharing data. We also are indebted to feedback from conference participants at the American Political Science Association Annual Meeting 2019, Southern Political Science Association Annual Meeting 2018, and at the 2018 workshop on “Beyond Political Violence: Causes and Consequences of Crime and Social Violence” at the University of Essex. We are grateful to the special issue editors and reviewers for helpful comments.

### **Funding**

This project was funded by Gerda Henkel Stiftung, AZ 08/KF/15. Gleditsch would also like to acknowledge other support for his research on crime and text as data from the Economic and Social Research Council, ES/L011859/1, and a British Academy Newton Fellowship to Mauricio Rivera.



## References

- Akee, Randall; Miriam Jorgensen & Uwe Sunde (2015) Critical junctures and economic development: Evidence from the adoption of constitutions among American Indian Nations. *Journal of Comparative Economics* 43(4): 844-861.
- Benoit, Kenneth; Kevin Munger & Arthur Spirling (2019) Measuring and explaining political sophistication through textual complexity. *American Journal of Political Science* 63(2): 491-508.
- Bureau of Indian Affairs (2019) Indian entities recognized by and eligible to receive services from the United States Bureau of Indian Affairs. *Federal Register* 84(22): 1200-1205.
- Cohen, Felix S (1971) *Handbook of Federal Indian Law*. Albuquerque, NM: University of New Mexico Press.
- Cornell, Stephen & Joseph P Kalt (1991) Where's the glue? Institutional bases of American Indian economic development. Working paper, Kennedy School of Government, Harvard University.
- Cornell, Stephen & Joseph P Kalt (1995a) Where does economic development really come from? Constitutional rule among the contemporary Sioux and Apache. *Economic Inquiry* 33(3): 402-426.
- Cornell, Stephen & Joseph P Kalt (1995b) Cultural evolution and constitutional public choice: Institutional diversity and economic performance on American Indian reservations. Working paper, Kennedy School of Government, Harvard University.
- Cornell, Stephen & Joseph P Kalt (2000) Where's the glue? Institutional and cultural foundations of American Indian economic development. *Journal of Socio-Economics* 29(5): 443-470.
- Dippel, Christian (2014) Forced coexistence and economic development: Evidence from native American reservations. *Econometrica* 82(6): 2131-2165.
- Fariss, Christopher J; Michael R Kenwick & Kevin Reunig (2020) Estimating one-sided-killings from robust measurement model of human rights. *Journal of Peace Research* 57(6): XXX-XXX.
- Ferguson, Kennan (2016) Why does political science hate American Indians? *Perspectives on Politics* 14(4): 1029-1038.
- Flies-Away, Joseph Thomas; Carrie Garrow & Miriam Jorgensen (2007) Native Nation Courts: Key Players in Nation Building In *Rebuilding Native Nations: Strategies for Governance and Development*, ed. Miriam Jorgensen. Tucson, AZ: University of Arizona Press: 115-145.
- Gover, Kirsty (2010) *Tribal Constitutionalism: States, Tribes, and the Governance of Membership*. Oxford: Oxford University Press.



- Grimmer, Justin & Brandon M Stewart (2013) Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis* 21(3): 1-31.
- Hayo, Bernd & Stefan Voigt (2007) Explaining de facto judicial independence. *International Review of Law and Economics* 27(3): 269-290.
- Hopkins, Daniel & Gary King (2010) Extracting systematic social science meaning from text. *American Journal of Political Science* 54(1): 229-247.
- Jamal, Amaney A; Robert O Keohane, David Romney & Dustin Tingley (2015) Anti-Americanism and anti-interventionism in Arabic Twitter discourses. *Perspectives on Politics* 13(1): 55-73.
- Krüger, Jule & Ragnhild Nordås(2020) A latent variable approach to measuring wartime sexual violence. *Journal of Peace Research* 57(6): XXX-XXX.
- Laver, Michael; Kenneth Benoit & John Garry (2003) Extracting policy positions from political texts using words as data. *American Political Science Review* 97(2): 311-331.
- Lemont, Eric D (2006) *American Indian Constitutional Reform and the Rebuilding of Native Nations*. Austin, TX: University of Texas Press.
- Kalt, Joseph P (2007) The role of constitutions in native nation building: Laying a firm foundation. In: Miriam Jorgensen (ed.) *Rebuilding Native Nations: Strategies for Governance and Development*. Tucson, AZ: University of Arizona Press, 78-114.
- King, Gary; Jennifer Pan & Margaret E Roberts (2013) How censorship in China allows government criticism but silences collective expression. *American Political Science Review* 107(2): 1-18.
- Marquardt, Kyle M. (2020) How and how much does expert error matter? Implications for quantitative peace research. *Journal of Peace Research* 57(6): XXX-XXX.
- Melton, James & Thomas Ginsburg (2014) Does de jure judicial independence really matter? A reevaluation of explanations for judicial independence. *Coase-Sandor Institute for Law and Economics Working Paper* 612.
- Miller, Robert J (2011) Tribal constitutions and native sovereignty. In: José Antonio Lucero, Dale Turner & Donna Lee Van Cott (eds) *Oxford Handbook of Indigenous People's Politics*. Oxford: Oxford University Press.
- Porter, Martin F (1980) An algorithm for suffix stripping. *Program* 14(3):130–137. <https://doi.org/10.1108/eb046814>.
- Rusco, Elmer R (2000) *A Fateful Time. The Background and Legislative History of the Indian Reorganization Act*. Reno and Las Vegas, NV: University of Nevada Press.

Schrodt, Philip A & David Van Brackle (2013) Automated coding of political event data. In: VS Subrahmanian (ed.). *Handbook of Computational Approaches to Counterterrorism*. New York: Springer.

Slapin, Johnathan B & Sven-Oliver Proksch (2008) A scaling model for estimating time series party positions from texts. *American Journal of Political Science* 52(3): 705-722.

Spirling, Arthur (2012) US treaty making with American Indians: Institutional change and relative power, 1784-1911. *American Journal of Political Science* 56(1): 84-97.

Tsebelis, George & Dominic J Nardi (2014) A long constitution is a (positively) bad constitution: Evidence from OECD countries. *British Journal of Political Science* 46(2): 457-478.

Wedeking, Justin P & Ryan J Owens (2011) Justices and legal clarity: Analyzing the complexity of US supreme court opinions. *Law & Society Review* 45(4): 1027-1061.

Wilkins, David E (2006) Introduction. In: Cohen, Felix S (ed.) *On the drafting of tribal constitutions*. Norman, OK: University of Oklahoma.

Wilkins, David E & Heidi Kiiwetinepinesiik Stark (2011) *American Indian Politics and the American Political System*. Lanham, MD: Rowman and Littlefield.

REBECCA CORDELL, b. 1989, PhD in Government (University of Essex, 2017), Assistant Professor, Department of Political Science, University of Texas Dallas (2019– ); research interests include state repression, political violence, human rights and measurement.

KRISTIAN SKREDE GLEDITSCH, b. 1971, PhD in Political Science (University of Colorado, Boulder, 1999); Professor, Department of Government, University of Essex (2005–); Research Associate, Peace Research Institute Oslo (2003– ); research interests include violent and nonviolent conflict and democratization.

FLORIAN G KERN, b. 1983, PhD in Politics and Public Administration (University of Konstanz, 2016); Lecturer, Department of Government, University of Essex (2017– ); research interests include political economy of local governance, conflict and development with a focus on indigenous governance.

LAURA SAAVEDRA-LUX, b. 1992, PhD Candidate in Government, University of Essex (expected 2020); research interests include inequalities, post-conflict state building, and peace.

Table I. Stemmed terms included in the judiciary dictionary

adjud, adjudg, bylaw, court, judg, judici, judiciari, jurisdict, justic, law, lawandord, prejudici, quasijudici, unlaw
--

Table II. Stemmed terms included in the judicial independence dictionary

appoint, appointe, author, compens, decid, decisi, disqualif, disqualifi, elect, elector, elig, independ, inelig, power, qualif, qualifi, reappoint, remov, salari, select, serv, term, vote, voter
---

Table III. Percentage agreement, precision and recall for the JUDICIARY and INDEPEND replication variable using a dictionary approach

	Percentage Agreement	Precision	Recall
JUDICIARY	84.211	1	0.842
INDEPEND	81.579	0	0.000

Table IV. Model accuracy, precision and recall for the JUDICIARY replication variable using supervised machine learning

	Accuracy	Precision	Recall
Logistic Regression	0.9	1	0.9
Naïve Bayes	0.9	1	0.9
Support Vector Machine	0.9	1	0.9

Table V. Model accuracy, precision and recall for the INDEPEND replication variable using supervised machine learning

	Accuracy	Precision	Recall
Logistic Regression	0.8	0	0
Naïve Bayes	0.9	0	Undefined
Support Vector Machine	0.9	0	Undefined

Figure 1. Constitutions by year of adoption

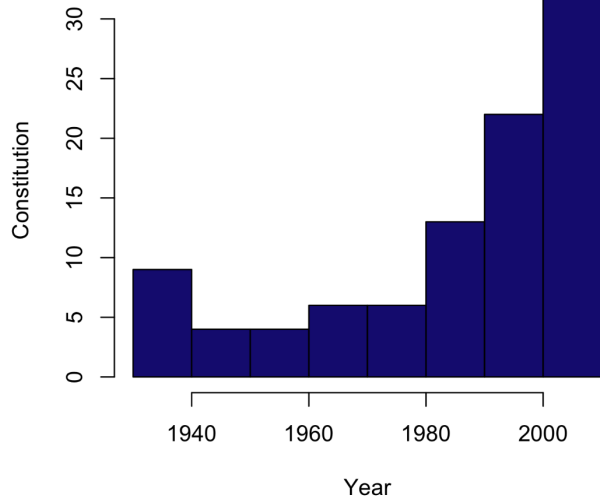


Figure 2. Constitution length, measured as total word count of a constitution

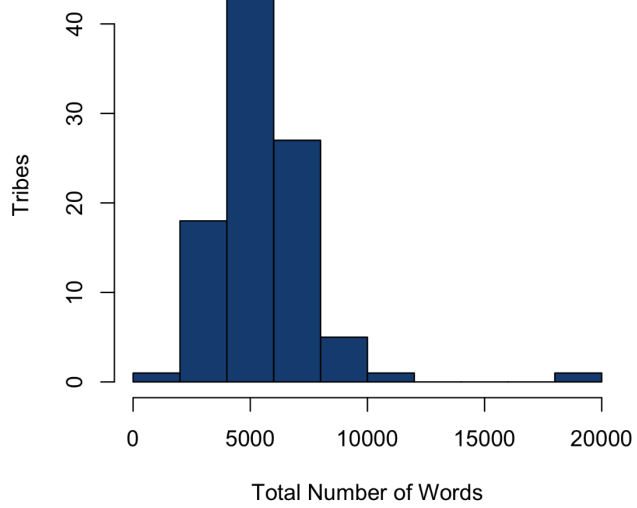


Figure 3. Constitution length (measured as average word count), by decade of adoption

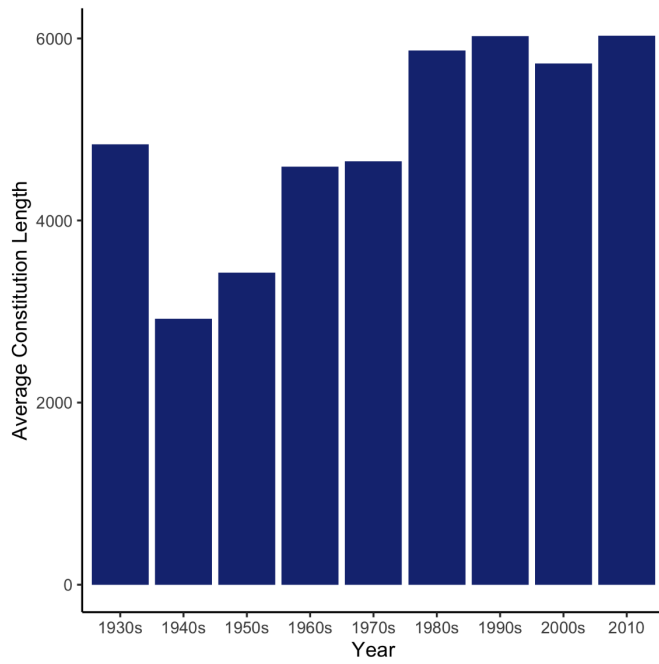


Figure 4. Length of the section on the judiciary, measured as total word count

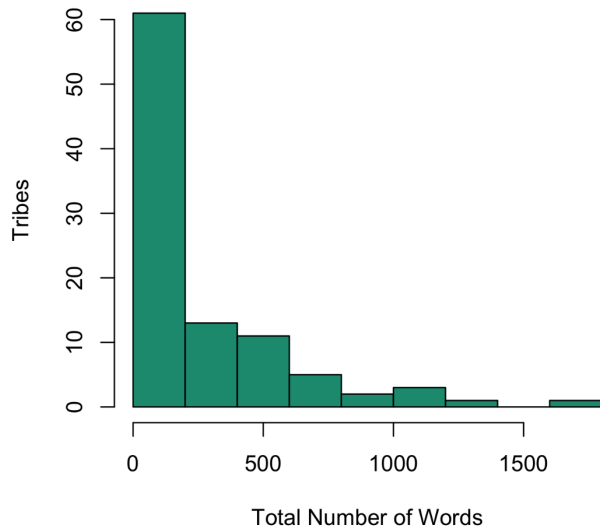


Figure 5. Terms associated with the judiciary included in a constitution, measured as total word count

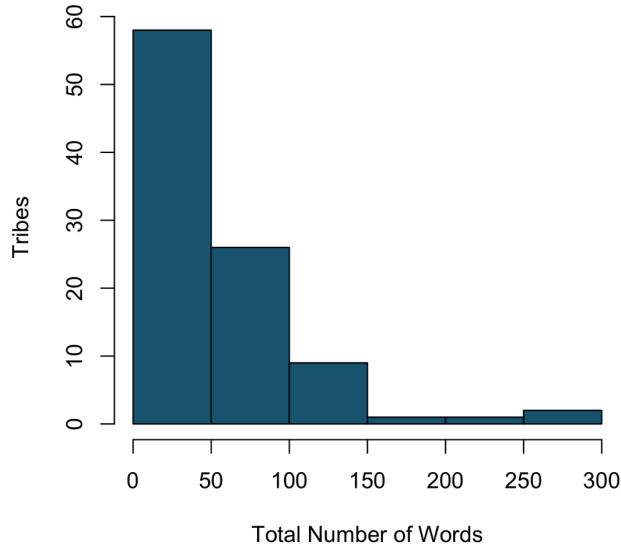


Figure 6. Terms associated with an independent judiciary included in the judicial section of a constitution, measured as total word count

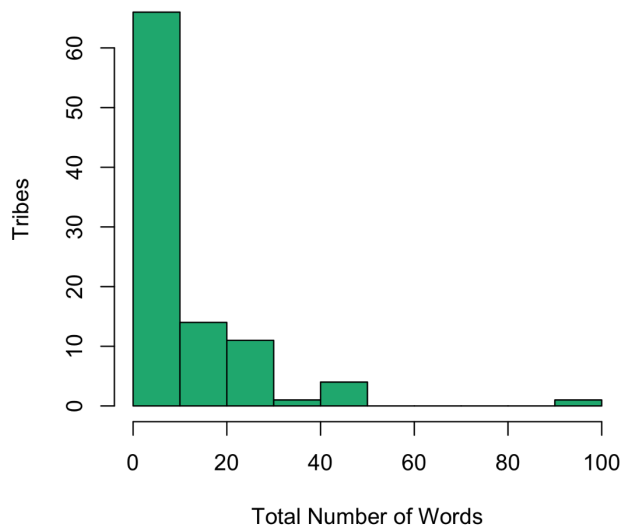


Table A1: Excerpts from three American Indian constitutions regarding the structure of judicial institutions

<b>Tribe</b>	The Yankton Sioux Tribe of South Dakota	The Mescalero Apache Tribe in New Mexico	Turtle Mountain Band of Chippewa Indians in North Dakota
<b>Year of Approval</b>	1998	1988	2006
<b>Frequency of Terms Included in Judicial Independence Dictionary</b>	Least	Average	Most
<b>Statement of Judicial Independence</b>	None	None	‘The Judicial Branch shall have authority to independently develop its operating budget and independently secure funding for its operations directly from funding sources’
<b>Judicial Tenure</b>	None	‘The tenure and salary of tribal judges shall be established by ordinance of the tribal council’	‘The term of office of the Chief Judge of the Turtle Mountain Tribe and all other judges and Chief Clerk of Court shall be four years’
<b>Selection Procedure</b>	‘The judicial power shall remain in the Tribal membership and upon request of five members for any just cause, such cause shall be submitted at any regular Tribal meeting for action.’	‘The tribal court shall consist of a chief judge and two associate judges, appointed by the President of the Mescalero Apache Tribe, with the concurrence of not less than a three-fourths majority vote of the whole membership of the tribal council.’	‘Properly registered candidates for the Office of Chief Judge shall then be subjected to a reservation wide or at large election’
<b>Removal Procedure</b>	None	‘Any member of the tribal council or elected officer of the Mescalero Apache Tribe who, during the term for which he is elected, is convicted of any felony, shall automatically forfeit his office. Any member of the tribal council or elected officer found guilty of a misdemeanor involving moral turpitude, gross neglect of duty, malfeasance in office or misconduct reflecting on the dignity and integrity of the tribal government, may be removed from office by majority vote of the tribal council.’	‘All judges of the judicial branch of government shall be subject to impeachment based only upon cause, as developed by the Judicial Board, only after due process of law is provided. The applicable standard shall be clear and convincing evidence.’



---

**Salary Insulation** None

'The tenure and salary of tribal judges shall be established by ordinance of the tribal council'

None

---

Table A2: Correlation matrix for the dictionary measures, using a Pearson correlation coefficient

	Constitution Length	Judicial Section Length	Judiciary Dictionary	Judicial Independence Dictionary
Constitution Length	1	0.523	0.747	0.484
Judicial Section Length	0.523	1	0.837	0.951
Judiciary Dictionary	0.747	0.837	1	0.781
Judicial Independence Dictionary	0.484	0.951	0.781	1