

POLI3148 Data Science in Politics and Public Administration Project Poster

CONSTITUTIONAL PREAMBLES DATA-DRIVEN QUANTITATIVE ANALYSIS

INTRODUCTION

Despite the lack of legal effect *stricto sensu*, Constitutional Preambles are where one finds "explicit expression of ideological values" of countries (Law, 2016). Although Constitutional Preambles are known for "being the most varied and idiosyncratic" (Law, 2016), researchers have at the same time conjectured that "writers of preambles borrow from the language of other preambles" (Ginsburg et al, 2014). This data-driven analysis research project aims to shed light on these claims. Through investigating the distribution of words across Constitutional Preambles worldwide and the sentiments they express, this project ventures into unearthing patterns, peculiarities and commonalities among Constitutional Preambles. Ultimately, it hopes to encourage researchers in comparative politics to "take preambles seriously" (Frosini, 2019).

DATA

This research project draws upon three datasets. First, a *Dataset on All Constitutional Preambles* in force as of 2021 that is self-constructed. This dataset includes the country, full text of the Constitutional Preambles, Constitutional Groupings which the preambles belong to (e.g. socialist constitutions) and the years in which the constitutions were enacted. It is retrieved from the Comparative Constitutions Project. Second, the *Freedom House Index* available on the Freedom House website. Among others, the dataset contains variables including the Freedom House Score of a country and the Freedom House Status the scores correspond to (e.g. "Free", "Partly Free", and "Not Free"). Third, the *Historical Index of Ethnic Fractionalization Dataset (HIEF)* which includes a variable that measures the degree of ethnic division of countries (Drazenova, 2019). It is downloaded from Harvard Dataverse.

RESEARCH METHODOLOGY

This research project proceeds in three phases with the use of RStudio Cloud. *Phase One*, the wrangling phase that consists of four steps. First, wrangling and cleaning up of the three datasets which include, for example, dropping of irrelevant variables and harmonizing of names of variables to enable the merger of three datasets. Second, the merger of three datasets into one meta dataset. Third, tokenization of the text (i.e. the Constitutional Preambles). Fourth, wrangling of the tokenized text, such as removal of stopwords and stemming. *Phase two* consists of exploratory data analysis of the (tokenized) meta dataset. Specifically, the length of Constitutional Preambles across all constitutions and constitutions within each region and constitutional group, the distribution and frequency of words appearing in the Constitutional Preambles, the relationship between the length of the Constitutional Preambles and other variables such as Freedom House Scores and HIEF. *Phase Three* is a lexicon-based sentiment analysis, using AFINN, NRC and BING lexicon sentiment datasets as dictionaries. The sentiment of each Constitutional Preamble and Constitutional Preambles within each region and constitutional group, such as "anger", "joy" and "positiveness", expressed in terms of different scores, are explored, visualised, and analysed.

LIMITATIONS

There are four limitations associated with this research and its methodologies. First, some data points are missing due to incompleteness of certain datasets. For example, the HIEF Dataset does not contain the Ethnic Fractionalization Index of certain countries. Second, the datasets may not necessarily lend themselves to far-reaching conclusions compared to what one may have expected. Although it turned out to be true that there are indeed similarities in the tokens of words across different Constitutional Preambles, it does not mean that these words are necessarily used in the same way within the specific context of drafting, and hence does not totally support a conclusion that drafters have copied from each other as what Ginsburg et al. postulated. Third, the Constitutional Groupings variable may be arbitrary as it is constructed based on the literature and common understandings. They do not exhaust all possible categories. Indeed, some are marked as having "No Grouping". Fourth, a lot of words used in the Constitutional Preambles are quite unique and may not feature in the lexicon sentiment dictionaries. They are, therefore, not amenable to lexicon-based sentiment analysis.

FINDINGS (SELECTED) AND ANALYSIS

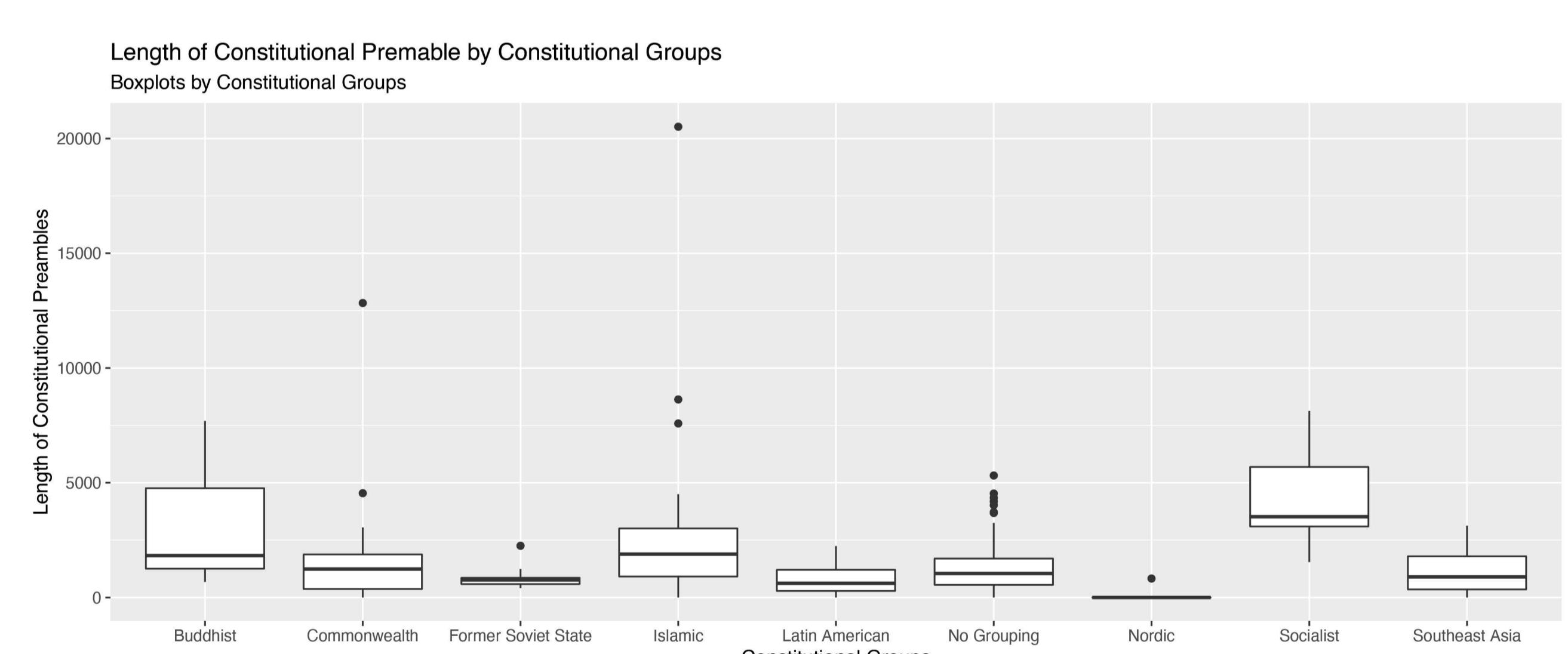
Most Frequent Words in Constitutional Preambles

The following word cloud shows the most frequently-appearing words in the corpus of all Constitutional Preambles currently in force worldwide. Words like "nation", "people" and "freedom" feature prominently. This is not surprising because frequent reference to these words help strengthen the legitimating narrative that the state truly represents the people. More interestingly, these words feature equally prominently in Constitutional Preambles of Islamist countries (which one may expect fewer mention due to lesser need to appeal to the "people" or 'freedom' to justify the state), and Former USSR countries (which one may expect more of such mention due to historical experience). This lends credence to Ginsburg et al's constitutional convergence thesis.



Length of Constitutional Preambles

The following boxplot shows lengths of Constitutional Preambles by Constitutional Groupings (including, e.g., Islamist countries, Buddhist states etc.). Result shows that Constitutional Preambles of Socialist countries are on average the longest, followed by Buddhist and Islamist. The result can be explained by the fact that Socialist countries, which establishment and legitimacy may depend less on the people's direct mandate but more on ideology, crave for the entrenchment of their narratives to cement their legitimacy. Constitutional Preambles lend themselves to such entrenchment.



Lexicon-Based Sentiment Analysis of Constitutional Preambles

The following table depicts polarity scores (and their mean) of a sentiment analysis of Constitutional Preambles using AFINN lexicon sentiment dataset. The higher the mean, the more "positive" Constitutional Preambles of that group are. It is interesting that Buddhist states' Constitutional Preambles turn out to be most negative, especially given Buddhist's teachings which stress peace of mind.

Constitutional_Groupings	afinn_polarity_score	afinn_polarity_score_mean
Buddhist	-42.7	-14.2
Commonwealth	126.2	11.5
Former Soviet State	11.2	0.8
Islamic	-93	-3.7
Latin American	41.2	2.4
No Grouping	135.6	1.9
Nordic	10.4	10.4
Socialist	70.9	14.2
Southeast Asia	11.3	1.9

