Reviewer #1: The authors introduce a novel dataset on the Czech Constitutional Court (CCC), encompassing information on decisions, justices, and clerks. In doing so, they make a valuable contribution to the literature and the scientific community by presenting data that extends beyond the well-studied US courts and the increasingly examined Western European courts. I express my gratitude to the editors for inviting me to review this work. Unfortunately, the data is currently inaccessible. I understand that a double-blind review process requires this, but evaluating an article introducing a dataset without access poses limitations on the reviewer's end (e.g. in subsection 4.2 Accessibility). Perhaps the editor can explore the possibility of making the data accessible in the case of resubmission.

I hope that in the next round of reviews, I will be able to share the data by sending the files to the editor.

Having said that, I recognize significant potential in the presented data but recommend that the authors rewrite and resubmit their paper before potential publication. In the following paragraphs, I outline a few points that, in my opinion, are worth reconsidering. Since I believe all the mentioned points can be addressed, I suggest granting a (major) revise and resubmit.

1. The authors introduce the CCC in section two. While they include important information, they also miss information that links to their data structure. For example, they write that "[…] justices are appointed by the president of the Czech Republic upon approval of the Senate" (p. 2). Considering that data on justices is collected as part of the CCC dataset, I would like to know more about the judicial selection procedure. How are nominees chosen? What are the preconditions to become a nominee? How politicized is the process? Are hearings held? How does the Senate make a decision? What majorities are necessary? In my view, this information is necessary to better understand the data generating process. The same holds true for the legal procedures mentioned in paragraph two of section two. In my opinion, a few more details to better understand the societal and political implications of different procedures (including some summary statistics) are necessary. Finally, the authors rightfully claim that including information on clerks is an asset of their dataset. However, in section two, I do not read any technical details and background on how clerks are selected and what is known about clerks in the Czech Republic in general. It seems that the authors have enough room to elaborate on these points.

All these points have been expanded upon in the manuscript.

2. The authors present the structure of the datasets in section three. After reading the section, I am still puzzled by how many tables there are exactly, whether data is repeatedly collected, or linked through identifiers, some "variables [contain] multiple values nested in one cell" (p.7) etc. I believe an easy fix is to include a visualization of the data structure described throughout the section. The authors could also add an annotated decision to the appendix to outline which information is extract from where. Additionally, the authors refer back to Epstein et al. (2014) to characterize their data as a "multi-user dataset" (p. 7). In this regard, I was wondering whether their dataset is more of a database, as multiple datasets can be connected in flexible ways through unique keys (compare p. 7). Considering that databases model real-world objects—both entities [...] and relationships [...]— and capture the structure between them (see Foster et al. 2021, 71).

The schema of the database has been added via its underlying SQL structure. All mentions of dataset have been replaced with the word "database" as I fully agree that what I am presenting is a database. It's just that almost all the previous articles on similar topic have utilized the term dataset.

3. I do have a number of concerns with the figures shown by the authors. Mostly, as the figures are neither well integrated in the text nor explicitly referenced. Below Figure 1 the authors write that "while the caseload of the CCC has grown, the proportions between the admissibility and merits decisions have not changed" (p. 5). I would suggest to than also show these proportions explicitly. Thus, instead of absolute numbers a stacked plot based on percentages might be more compelling. Figure 2 plots the age of the justices at entry. However, no story is told concerning that age. Again, the comment blow the figure appears as a note to the figure. What makes age so interesting to study? Why should law scholars, journalists or social scientists care about this age? The authors

might wish to better integrate the figure or they could drop it. Figure 3 is better integrated as the authors tell a short story about the justices' alma maters. The integration argument made is
interesting, but what are its implications? Also, is it true that a justice who might have grown up in Prague but afterwards studied elsewhere is less integrated than a justice who moved to Prague and studied there?

The figures have been remove as they do not fit the purpose of the said sections and to make space as the rest of the article has been expanded. Remaining tables have been better integrated with the whole text, references to the presented tables have been added in the text.

4. From my perspective, the previous two points I raised can be addressed by improving the structure of section three. As a reader, I wish to have a clear understanding of the datasets and database architecture after going through section three. Unfortunately, this is currently not achieved. Therefore, I would like to underscore the need for visualization of the datasets and data structure accompanied by a narrative developed around that avisualization. This may lead to the conclusion that certain figures are unnecessary, while more fundamental descriptive statistics could be more beneficial. It appears that the authors' objective is to outline the empirical capabilities of the CCC dataset in section four, making it imperative for section three to focus solely on delineating the architecture of the datasets.

The whole section introducing the database has been reworked, it follows along the structure of the database rather than the substance of the variables. Moreover, a clear schema of the database has been added.

5. The authors utilize section four to outline that their data adheres to the tidy principle while presenting empirical examples. In my opinion, not all examples are presented in a compelling manner. For instance, in the section on clerks, the authors state, "such information can serve as a basis for any research on the role of clerks" (p. 8) and continue on page 9 by noting, "surprisingly, given the theoretical expectations, judges seem to prefer clerks equally, whereas female judges seem to hire more male clerks." However, the theoretical expectations are not introduced anywhere in the text. I would personally recommend doing so. Choose an assumption or theoretical expectation from existing research on clerks; the authors already cite relevant studies (p. 9). Afterward, use the CCC dataset to demonstrate whether the assumption holds or if the expectations are confirmed.

The vignettes have been both reworked. The vignette on clerks has been completely rewritten so that it includes more thorough literature review that is clearly linked to theoretical expectations, which I then "test" with concise descriptive statistics. The vignette on separate opinions now includes descriptive statistics instead of a regression model as I believe that fully specifying a regression model is out of the scope of the article.

6. The example in section 4.1.2 is more appropriately placed within the existing literature, and the authors effectively address the point I just raised. They outline the ongoing debate in the literature in the first paragraph of the respective section (p. 9). However, the example has its limitations, as the authors mention, "I manually annotated which justices of the 3rd term were from which coalition." This implies that such information is not included in the CCC dataset. Consequently, I am uncertain about the extent to which the authors present a cohesive application using their data

I added mentions of the foundationality. To this end, I do not believe the database to include all-encompassing data. Quite the opposite: it serves as a foundation and can be easily melded and adopted to any research purpose. Here, I used a lot of data from the database, to which I just added the information on which judge comes from which coalition in line with the Czech legal scholarship. At no point do I attempt or claim to have attempted to include every single piece of information or claim that occurred in the scholarship.

7. The authors seek to introduce a new dataset. In this regard, I would recommend better outlining how data across different datasets in the CCC dataset are connected to extract variables addressing research questions, such as the one on dissenting behavior. In other words, how do colleagues "work" hands-on with the data. Section four is, in parts, too short to outline this, and it seems necessary to present greater detail. You may also

include a short example of commented code in the appendix.
I am not sure I fully follow this remark?

8. In section 4.3, I would encourage the authors to provide additional details regarding the source of their decisions. The authors assert that they retrieve all decisions from the website of the CCC. Is this the official platform where all decisions are publicly disclosed? Does the court consistently publish all its decisions online? Are there any limitations associated with this information source? Although a minor point, in the context of a cross-European comparison, it is not evident that a constitutional court necessarily publishes all its decisions on its website.

This has been cleared up in the text.

Some minor points the author might wish to address: (A) The title page suggests there are two authors, but at some point in the manuscript, the authors start to talk about a single author, using the pronoun "I." (B) The use of the terms "justices" and "judges" does not appear to be consistent throughout the manuscript   . (C) Figures and the table require better labeling and inclusion in the text; for example, use a meaningful label in Table 1 instead of a coded variable name like "coalitionfull_coal." (D) If possible, mention sources for additional information collected in the attached codebook. For instance, you mentioned using Wikipedia and official profiles for judges; therefore, it would be helpful to provide URLs or general sources. (E) Not being a native speaker myself, I understand the challenge of writing in a foreign language. This is why I would suggest a thorough copy-editing and proofread of the manuscript; beginning an article with "although" sound unusual, "to the my best knowledge" (p. 2), "the type of decision being review" (p. 4), the conclusions starts with the incomplete sentence "The presented article introduces a dataset on the Czech Constitutional Court, while aiming at bridging the gap." (p. 12)

All these mistakes have been fixed and the manuscript underwent further proofing.

Reviewer #2: This article introduces the Czech Constitutional Court (CCC) database — one of the first court databases that focuses on a central or eastern European country. I think this database is an important contribution and opens the door for greater representation in the empirical cases that scholars have the opportunity to use when testing theory.

Framing

In the introduction, the author does a nice job of situating the project in the context of other recent efforts to expand the availability of high-quality data on courts in Europe.

I think this section could be strengthened by highlighting what research questions in the empirical literature — which has tended to focus on the United States and, to a lesser extent, the EU — the database would allow researchers to address in the Czech context. Given the institutional structure of the CCC, some questions in the American and comparative judicial politics literatures will be relevant, while others will not. Can the author provide readers with some specific open research questions?

The author does a nice job of concisely layout out the institutional structure and functioning of the CCC. This provides a good basis for scholars who are not familiar with the CCC to understand the structure and coverage of the database.

At the end of this section, I think it would be helpful for the author to more explicitly highlight reasons why the CCC might serve as a good empirical case for scholars who are looking to test a theory but are not necessarily focused on Czechia and who might not have otherwise considered the CCC as an example. What are the institutional features that are similar to or different from other European courts that would make the CCC a particularly useful case for testing general theories?

Added a couple of paragraphs into the introduction, in which I raise a couple of research questions that could be potentially resolved with the data at hand, discuss its limitation and also discuss what makes the CCC interesting

court to research.

Presentation of the database

I think it'd be helpful if the paper included a diagram or table that summarizes the tables (unit of observation, number of variables, number of observations, etc.) and visually shows how the various tables relate to each other.

The schema has been added.

Vignettes

I like that the author includes two vignettes to highlight how the data could be used, but I think developing these a bit more would help to highlight the usefulness of the database. The vignette on clerks at the CCCis particularly interesting. I'd like to see the author highlight some specific research questions directions related to clerks that scholars can use the database to address — the data on clerks seems like a particularly unique and noteworthy part of the database and I think the author could sell that part more.

I tried my best to "sell that part more". ☺

The second vignette, which applies a hypothesis that has previously been examined in the context of the plenum to small panels, is a good idea, but I think the regression model needs to be more fully specified to be a helpful addition to the paper. Of course, it's not within the scope of the paper to develop a fully-specified empirical test, but I think at least a short consideration of alternate explanations and control variables is necessary. A data visualization that shows that the descriptive statistics are consistent with the theory might be a good alternative to more fully developing a regression model, if there are space constraints.

Went with the alternative.

Technical questions

What pre-processing has been done to the text corpus? Looking at the codebook, it doesn't look like the text has been parsed into paragraphs. For many common NLP tasks (e.g., multi-class labeling, NER tagging, etc.), I think users would want to use a finer unit of analysis than a decision. Is the raw text formatted in a way that it's easy to split up decisions and remove non-substantive text (front matter, footnotes, etc.)?

The information on (the lack of) pre-processing has been added.

I appreciate the author's discussion of reliability and reproducibility. The author notes that the amount of human input has been minimized to increase reproducibility. In my experience, using regular expressions and similar strategies to code variables from metadata or from the text of legal decisions can be surprisingly fraught when the raw metadata or text contains inconsistencies, omissions, and/or errors. How reliable is the raw information from the Nalus website? While reproducibility via an automated process is important, I think it's also important to carefully validate whether automated processes are incorrectly handling edge cases and are appropriately calibrated to check for and account for problems with the raw metadata or text. I'd like to see the author discuss in more depth any validation checks they perform.

I'd also like to see the author discuss the issue of missingness, and whether, based on how the data is collected, there is potential for missingness due to inconsistencies in the formatting or structure of the raw data.

For example, in the codebook for the compositions table, the author mentions that there's a "rather high degree of irregularity" in the formatting of early decisions. Can the author provide a reliability report as supplemental information that applies automated validity checks? For example, with the compositions table, you could check that number of judges is correct or plausible.

A whole section on reliability check has been added, in which I discussed both the NALUS database as well as the reliability of my own data extraction process.

For categorical variables with a limited number of categories, like party_kind in the parties table, it'd be nice if the codebook included a list of possible values that users can have in front of them while working with the data.

I added it to the codebook wherever it wouldn't make it too cluttered.