Analysis Reflection Paper

I.   Introduction: **The effect of gerrymandering on the political ideology of the House of Representatives Members**

The political ideology among members of the House of Representatives (HOR) is becoming increasingly polarised. In my research paper, I explore the correlation between gerrymandering and the political polarisation of the House of Representatives (HOR) members between 2010-2020 (using the 2010 redistricting cycle). I contribute to the existing literature on polarisation by contending that gerrymandering - the re-drawing of district lines for the protection and unfair advantage of incumbents – is a catalyst of political polarisation in the HOR. The correlation between gerrymandering and political polarisation of the HOR has been disregarded in the past due to a lack of outcome analysis. However, I hypothesise that in gerrymandered districts, members of the HOR obtain more extreme views to please their most vociferous voters, moving away from the median political ideology of their district to the median of their party. A comprehensive analysis of these theories should shed light on whether there is a correlation between gerrymandering and the political polarisation of HOR members.

I.   Method 1: Statistical Analysis of district-wide Data

One way to approach this is through a quantitative analysis of nominal data for numerous districts. Using two binary variables (gerrymandering in the 2010 cycle on the X-axis and polarisation of the HOR members on the Y-axis), I will carry out this analysis on 10 districts, half of which have been gerrymandered in the 2010 redistricting cycle, and half which have not. To gauge whether a state has been gerrymandered, I deem an efficiency gap of over 7% to indicate a gerrymandered state (Stephanopoulos 2014). I will be choosing these districts by searching for states that have been subject to legal challenges, for example North Carolina and Maryland. To assess whether the HOR member holds polarised political ideology, I use

the DW-Nominate score, in which a score of 0.3 or above indicates polarised ideology (Poole and Rosenthal 1985). With this statistical approach, I hope to observe a positive correlation for both the 'gerrymandered-polarised' and 'not gerrymandered-not polarised' outcomes, so that a correlation can be drawn between gerrymandering and the polarisation of political ideology of the HOR. To then prove my second hypothesis that HOR members' political ideology shifts towards the extreme to please their voters, I continue with a quantitative analysis approach. I will find data on voting patterns of citizens within each of the 10 districts from 2010 -2020. The X-axis represents the ten-year period, and the Y-axis represents the percentage of votes received. I will have one line for the HOR member and one line for the primary opposing candidate. I expect that the gap between the two lines (representing the percentage of voters received by the HOR member and opposing candidate) to increase, indicating increased polarisation of the HOR member to please their voters. The ideal type of data would be having this data available for every district for rounded representation. It would also consider other metrics of gerrymandering such as compactness and packing and cracking. For the second part of my method, it would be ideal to have the political opinions of the voters disaggregated by the extremity of their views, and what percentage the candidate obtained from each of those disaggregate groups. This data might be difficult to find in the real world because extremity of voting pattern is not often collected. More generally, real-world data sources that approximate these ideal features individually should be in abundance but finding data that embodies numerous aspects of gerrymandering might be more difficult. As a statistical analyses, it is clear cut whether a correlation exists.

II.   Method 2: Sentiment Analysis on District-wide Data

A further way to approach this analysis is through a modern natural language processing approach to count the occurrences per year of political terms stated by House of Representative members related to either Republicans or Democrats (e.g., "LGBTQ+ rights"

for democrats and "Border security" for Republicans) between 2010 (after the redistricting cycle) - 2020. The data that will analysed will be found on the internet, for example political news articles and video campaigns. Although this is a quantitative approach, the data I am modelling is qualitative, as the 'buzz words' in political campaigns indicate the characteristics of the jargon occurring at each political event. I will choose two districts, one that has been gerrymandered and one that has not and observe if over ten years the frequency of either side's words increase. I hope to observe an increase in frequency of the words in the gerrymandered district, to show a quantitative trend between gerrymandering and political polarisation of the HOR members to please their more extreme voters. In an ideal data set, there would be a predetermined list of political terms that are related to either Republicans or Democrats which serve as buzz words. This predetermined list will serve as a table which will be used to monitor the counts of each buzzword. Availability of resources is vast, from Government data sources to news archives. Large amounts of text data can be studied. This method would also be easy to interpret as there are simply numerical outcomes to compare, and with the temporal nature of the study shows a change over time which would strongly support my hypotheses.

III.   Limitations

Both methodological approaches to this analysis pose limitations. The first part of my statistical analysis can only produce four outcomes because of its binary structure, and so will not elucidate the extent of the connection between gerrymandering and political polarisation of the HOR member. Although this data is easy to interpret, it oversimplifies both the terms 'gerrymandering' and 'polarisation,' by using only one measure. Simplifying them is assuming that one metric embodies each word. The second data set faces the same problem, oversimplifying polarisation by using one metric. It is also important to distinguish between causation and correlation, as there are many other factors that I cannot consider that influence

the voting patterns of citizens which are unrelated to polarisation. There are also external validity issues. Although there might be a correlation in the tested districts, the limited scope of my method accounted for by the limited data of districts means it is difficult to generalise my analysis. As I am choosing districts that are specifically significantly gerrymandered and districts that are not, there is an issue in representativeness in that it might not be reflective of other districts.

There are also limitations for the sentiment analysis approach. Namely, there are internal validity issues such as the misinterpretation of sources, or inaccurately contextualising key words. If key words are used sarcastically, or in the context of opposition, they do not accurately represent the incumbent's ideology, rendering the results inaccurate. The reliability of the sources is also in question. In addition, there are limitations as to generalisability of the results, as this test only compares two districts. The time-consuming nature of this study makes it unfeasible to do on a larger scale. Finally, it is likely that there will be bias in the keywords chosen, making for inaccurate results. There is often an overlap between words, and so it is difficult to even come up with a key-words table that accurately embodies both sides of the political spectrum.

Overall, I think the quantitative statistical analysis is more favourable than the sentiment analysis approach for multiple reasons. First, the data sets are easy to find, readily available, and straightforward to interpret. For the second data set in my first method, although one must accept that there are numerous reasons for voting patterns, being able to draw a correlation is indicative of an important relationship. In comparison, it is difficult to find sources for the sentiment analysis approach, and even harder to process the data. Although setting the variables of gerrymandering and polarisation as binaries oversimplifies the complex nature of this issue, the results would prove a correlation between the two. Both datasets in my first method face the limitation of specific metrics for complex issues. If I

have time, perhaps I can find separate data on other metrics to measure "polarisation" and

"gerrymandering," and observe whether the results indicate the same correlation.

Works Cited

Poole, Keith T., and Rosenthal, Howard (1985). A spatial model for legislative roll call analysis. American Journal of Political Science, 29 (2), 357-384.

Stephanopoulos, Nicholas. O (2014). The Measure of a Metric: The Debate over Quantifying Partisan Gerrymandering. Columbia Law Review, 114 (7), 1683-1742.