Proceedings of the Tenth AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2022)

Crowdsourcing Perceptions of Gerrymandering

Benjamin Kelly*, Inwon Kang*, Lirong Xia

Rensselaer Polytechnic Institute
benjaminkelly@me.com, inwon.kang04@gmail.com, xialirong@gmail.com

Abstract

Gerrymandering is the manipulation of redistricting to influence the results of a set of elections for local representatives. Gerrymandering has the potential to drastically swing power in legislative bodies even with no change in a population's political views. Identifying gerrymandering and measuring fairness using metrics of proposed district plans is a topic of current research, but there is less work on how such plans will be perceived by voters. Gathering data on such perceptions presents several challenges such as the ambiguous definitions of 'fair' and the complexity of real world geography and district plans. We present a dataset collected from an online crowdsourcing platform on a survey asking respondents to mark which of two maps of equal population distribution but different districts appear more 'fair' and the reasoning for their decision. We performed preliminary analysis on this data and identified which of several commonly suggested metrics are most predictive of the responses. We found that the maximum perimeter of any district was the most predictive metric, especially with participants who reported that they made their decision based on the shape of the districts.

Introduction

Gerrymandering is the practice of creating districts for legislative maps that manipulate the results of an election to advantage whichever party is drawing the districts. In the United States, this practice is most commonly seen in elections for the United States House of Representatives, where each state is divided into a certain number of districts based on population data collected during the census each decade. These maps are often drawn and chosen by state legislatures, which allows parties in control of those bodies to create maps which give their candidates greater chances to win a large number of Congressional seats during the election. Gerrymandering that discriminates on voters based on race has been deemed unconstitutional (Shaw v. Reno (1993) and Miller v. Johnson (1995)). However, partisan gerrymandering along political lines remains legal in many states and the Supreme Court has ruled that federal courts lack authority to decide cases about such gerrymandering (Rucho v. Common Cause (2019)). Recently, the North Carolina State Supreme Court decided that partisan gerrymandering violated their state constitution (Harper v. Hall (2022)). Other states have sought to set up bipartisan redistricting commissions. Nevertheless, due to its political advantages and legality in many states, gerrymandering remains prevalent across the United States.

It would be ideal if one could construct an algorithm to detect gerrymandering as an unbiased way to view potential redistricting plans, but such efforts often are futile. A gerrymandered map has no precise definition and no specific legal criteria, and so conventional algorithms are difficult to construct. Similarly, using machine learning algorithms presents its own set of challenges, as the collection of data would require samples of both gerrymandered and non-manipulated maps, which again requires a definitive notion of what defines a gerrymandered district map. Even if such a standard were to be created and detection algorithms were able to exist, these efforts would still fail to consider the voting population's perceptions of gerrymandering. Legal guidelines could differ from people's perceptions of potential maps, causing rifts where 'objectively' fair maps appeared unfair. If drawing fair districts is the ultimate goal, it is just as important that any 'fair' proposals are also seen as such by voters to ensure trust in the election's results. Furthermore, voters might be fooled into thinking a district map is fair when it is actually heavily biased or manipulated.

Therefore, we designed a survey and produced a dataset using crowdsourcing that could be used to gain insight into how people view district assignments. We hope that with this data we can inform future decisions in creating fair and trustworthy districts. We present our choices in designing the survey, the dataset itself, and some preliminary analysis of the data on which features were the most predictive of perceived unfairness.

Related Work

Real Life Examples

Previous works on gerrymandering work either concretely with real life examples or more theoretically with abstractions of the problem. (Clelland et al. 2022) proposes redistricting plans for the state of Colorado. Using ensemble analysis, the authors generate multiple versions of the district plans applied on to the voter map and prune out the maps that

*These authors contributed equally.

produce outlier results. (Guest, Kanayet, and Love 2019) apply weighted k-means to pick the best redistricting plan for each state using real life data from the US census. They find that they are able to improve the compactness of districts in every US state. (Herschlag et al. 2020) apply a similar ensemble technique using Markov Chains to generate possible maps in the state of North Carolina to compare the three redistricting proposals.

There are projects open online sharing their efforts in redrawing the district maps in various states. (Bycoffe et al. 2018) analyze the current district maps of the US in 2018 to understand the influence of the districts exert on the election outcome. They show the possible results of elections using maps drawn following certain criteria such as favoring one party or focusing on the shape of each district. Their results show that biased redistricting policies greatly influence the outcome of future elections, especially compared to more 'even' maps that are drawn when keeping the district shapes in mind. (Garg et al. 2021) simulate elections US House of Representatives elections under various voting rules and election systems. They consider multi-member districts in which a district can cast votes for multiple candidates, and find that single transferable vote (STV) with independent commissions can achieve proportional outcomes in every state. (DeFord et al. 2021) apply the partisan symmetry metric defined by (Katz, King, and Rosenblatt 2020) as a new fairness metric in designing voting districts in states of Utah, Texas and North Carolina. They find that using partisan symmetry alone can lead to unforeseen consequences in the election results.

Theoretical Aspect of Gerrymandering

While works on new methods of drawing district maps are important, understanding what metrics should be used in evaluating district maps remains critical as well.

(Ramachandran and Gold 2018) discuss the limitations of current methods which average district 'shape' metrics used by officials, and discuss using outlier detection. Outlier detection of voting outcomes in a group of possible maps can be used to determine if a map is unnatural. This approach was also used by (Clelland et al. 2022) and (Guest, Kanayet, and Love 2019), where they generated multiple instances of district plans and chose the map producing the median of the election outcome. (Wang 2016) focus on measuring the fairness of each state's district plan, looking at the compactness of each district as well as its political makeup.

Other works have focused on representing gerrymandering as more common mathematical problems in order to reason about its potential effects and hardness. (Cohen-Zemach, Lewenberg, and Rosenschein 2018) describe representing gerrymandering as a graph problem and show that such a representation is NP-Complete. (Borodin et al. 2018) discuss the importance of population density in gerrymandering and that the gerrymandering power of a political party is amplified by its support in rural areas.

While it is important to be able to distinguish whether a map is natural or not, using the ensemble approach of generating every possible instance and picking the best one can be a very computationally intensive approach. If one had a better understanding of what features indicate a given popular perception of the assignment, the possible space of maps can be greatly reduced. Our work presents a dataset of over 2000 comparisons of different maps and human responses on which map appears more fair. This data sheds some light into how people tend to evaluate district maps.

Column	Value
Per each question	
IsGerrymander	[Map A, Map B]
Considered features	{Shape, Winner, Population, None}
Suggestions	Text
Per each user	
User ID	Unique string
Feature importance	{[0, 10], [0, 10], [0, 10]}
Explanation	Text
Survey summary	Text
Response key	UUID
IsSpam	Boolean

Table 1: Columns present in the final dataset.

The Gerrymandering Perception Dataset

As our main contribution, we present the The Gerrymandering Perception Dataset. This dataset is composed of responses from 482 unique participants. Each participant answered 5 different questions comparing two district maps applied on the same population distribution and selected which maps appeared to be more 'fair'. For each question, the participant also selected the features of the map they considered when making their decisions. We provided the participant three potential features: *district shape*, *population distribution*, and *election outcome*, allowing for multiple selections. In order to account for cases where the participant considered features that were not presented in the available options, we also asked them to provide an optional written response on what other features they might have considered. At the end of the survey, we asked the participants to assign a score to each of the three features, assigning highest score to the feature they considered the most important throughout the survey. We also asked the participant to share their overall decision process in the survey, asking them to elaborate why they made their decisions the way they did.

Description of Dataset

The final dataset is composed of 25 different columns for each response. These columns include the anonymized user ID, responses to the three questions for each of the 5 pairs of maps presented, and four questions asked at the end of the survey and five other values used to filter spam responses. A detailed description of each column present in our dataset can be seen in Table 1.

Map Design

We generated simplified maps of a 64 voter city who are split amongst three political parties. These voters were arranged

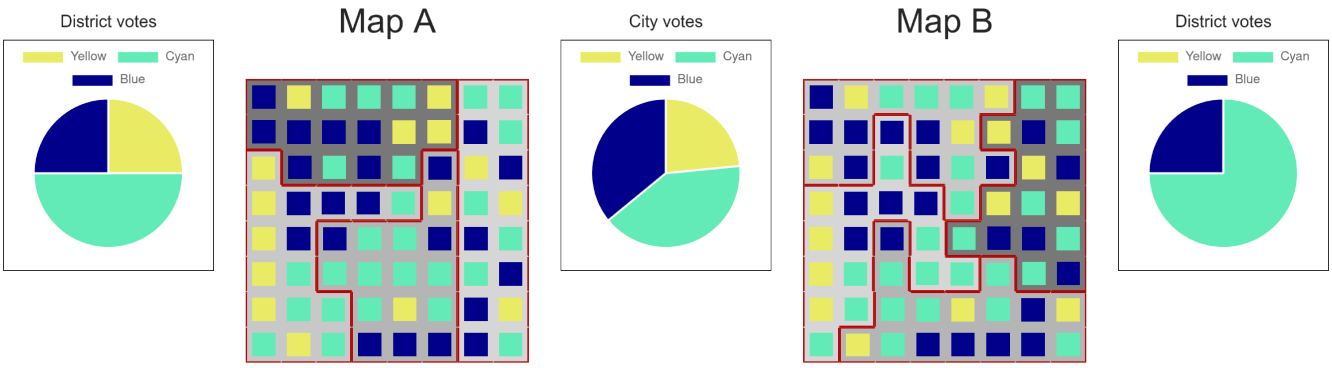

Figure 1: Screenshot of the main survey question.

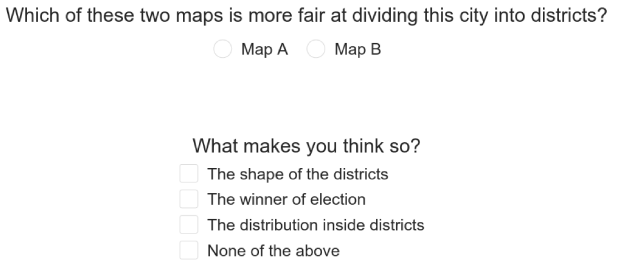

Figure 2: Screenshot of the reason question for main survey question.

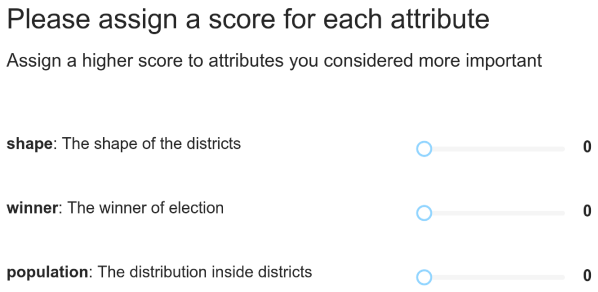

Figure 3: Screenshot of the final survey question.

in an eight by eight grid and then divided into four districts. This construction is a simplified representation of real district maps, but our goal was to create maps that were simple enough that the information could be conveyed to the participants easily. More complicated, real world style maps contain too much information for respondents to easily parse, and so by simplifying our model we allowed them to focus on all relevant data contained within our maps. We discuss some of the limitations this representation causes in a later section. While we initially designed a set of 15 maps by hand that examined different features, we realized that in doing so we might introduce our own biases by presenting maps we thought ourselves to be particularly fair or unfair. In particular, if we strived to create a selection of fair and unfair maps, we would be doing so using our preconceived notions of what features caused maps to be fair or unfair. Thus, we instead randomly generated the assignments to avoid such bias tainting our results. We first constructed five voter layouts that served as the distribution for each of the questions. Then, for each map, we generated five random district assignments splitting the voters into four districts of 16 voters each.

Survey Design

The main part of the survey was composed of five questions, each of which asked the participant to pick the map that they considered more fair out of two possible options. Each main question was composed of a required response about which map is more fair and the reasons behind the choice, followed by an optional written response on any possible suggestions on other features that the participants considered. Screenshots of the main question page can be seen in Figures 1 and 2.

In order to make sure that our participants understood the possible changes made to the election outcome for different district plans, we included a pie chart to indicate the voter preference in the overall population and the election outcome for each district plan. As can be seen in Figure 1, the center pie chart describes the overall population's political preference, while the pie charts on each side describe election results from the respective district plans. When hovering on each block, the user was also able to view the voter distribution inside the district it belonged to.

In the final page of the survey, we asked the participants to rank the importance of each aspect of the district assignments as well to provide a written explanation of how they made their decisions throughout the survey. Participants were asked to assign a score between 1 and 10 for each feature, namely *Shape of districts*, *Winner of election* and *Population distribution of districts*, with the higher score indicating a higher importance. This question provides some ground truth for each participant's response. Although the participants also marked which features they took into consideration when comparing each pairs of maps, they were

not asked to provide a specific ranking for them. Thus, we added the final question as a ground truth measure of the participant's overall utility of these features. The written response served a similar purpose and also allowed us to look into the key words or phrases the participants used to describe their decision making process.

Data Collection

We ran five batches of 100 instances of the survey on Amazon's Mechanical Turk platform. We did not restrict the demographics of our participants, and did not have access to such data. Each participant was compensated with 40 cents for their response, which was calculated with the estimation of five minutes per response and the federal minimum wage of 7.25 dollars per hour. In order to be able to display the maps and the related statistics, we built a web platform using React.JS and collected the responses using Google's Sheets API. Figures 1, 2 and 3 show the two types of questions that were present in our website [1]. Figure 4 shows Kendall Tau's correlation between feature values of the pair of maps that the participants in the dataset faced. Aside from the voter misrepresentation being correlated to other values, which is to be expected to a degree, we see that the other feature values are not very strongly correlated to each other, in order to provide a fair ground for the survey.

Data Filtering

In order to weed out data from participants who did not pay enough attention to the survey content or are automated bots, we first asked the users to pass a Captcha field in the beginning of the survey. The participants are only allowed to proceed in the survey if they pass the Captcha check. To make sure that our participants also understood the task being presented, we added a required written response at the end of the survey on describing the summary of the survey to serve as an attention check. This response was examined manually, and only those who passed this check were marked as non-spam on the final dataset, which resulted in 416 valid participants out of the total 482. Only the non-spam responses were considered in our data analysis. In this process, we noted that while some participants barely passed the attention check with responses that almost seemed random, some responses were very detailed in both the description of the survey and their decision process, providing a good insight into how the participants were handling the questions. In the end, our dataset was gathered through two different stages of filtering, namely the captcha field in the beginning and the attention check at the end of the survey.

Hypotheses

We had several hypotheses when collecting our dataset about what types of features we expected to see be most predictive. Broadly, we imagined there being two categories of features: election result based and district shape based. The former would depend on the difference between the distribution of preferences of the voters and the distribution of district winners after the election. Such features get to the heart of the effect of partisan gerrymandering, which can lead to large differences in election winners including turning minority parties into significant majorities. We expected to see a significant effect when parties earned far more or far less seats than they 'should' based on the population distribution.

The other type of feature we expected to learn was district shape features. Rather than being related to the population distribution and election results, these features were solely based on the geometric shape of the districts. Real world discussions about gerrymandering often include district shape as evidence of gerrymandering, as highly complex and 'weird' shapes allows for more control and potential for manipulation. Additionally, when looking at redistricting plans (especially highly complex real world examples), the shape of the district is what is visually apparent to a viewer on first inspection. These intuitions led us to believe that 'weird' district shapes would be more likely to be interpreted as unfair. Defining 'weird' is a question with much debate and many metrics have been proposed. Part of our goal was to use several commonly suggested metrics and see which matched how people viewed the district assignments.

Data Analysis

To understand the dataset we collected, we did exploratory data analysis to find out if a correlation exists between the features of each district assignment and the responses. We were interested in finding out whether our participants would focus on the shape of the district or consider other features of the map, such as the winner by population or winner by district. So we focused on features that either described the shape of the districts or the election outcome. In our analysis, we found that our participants primarily considered the shape of the districts, even when they responded that they considered the election outcome in their responses.

Perceived Fairness of District Layouts

In our survey, each map was expressed through a eight by eight grid. Each grid block represented a unit population, colored by its political preference and which district it belonged to. While this grid format cannot be said to be faithful expressions of real-life maps, it allowed us to present a variety of maps that our participants could easily understand. For each pair of maps compared, the political preferences of each block stayed equal, while only the district scheme changed. This pairwise comparison allowed us to calculate some features only related to the district layout in order to find out what features people considered when making their choices. Out of the features tried, Table 2 describes the features we found had the most correlation to the final choices made.

To find the correlation between these feature values and the perceived fairness of the presented maps, we calculate the difference between the feature values of the two maps. For two map features F_{mapA} and F_{mapB}, the resulting feature values are calculated as $F_{mapA} - F_{mapB}$. In case $mapB$ is selected as the more fair map, we negate the difference in order to generate one distribution for the fair map. Using this method, we are able to check the distribution of our features for maps that were selected to be 'more fair', shown

[1] https://inwonakng.github.io/gerrymander-map-builder/

Feature	Description
Diameter	$\frac{\textit{Maximum diameter of district}}{\textit{District area}}$
Convex Hull	$\frac{\textit{Minimum convex polygon}}{\textit{District area}}$
Border length	$\frac{\textit{Border length of district}}{\textit{District area}}$
Votes misrep.	$\frac{\textit{District votes}}{\textit{Population votes}}$
Winner Ratio	$\frac{\textit{votes for winner}}{\textit{total votes}}$

Table 2: Features used in analyzing results.

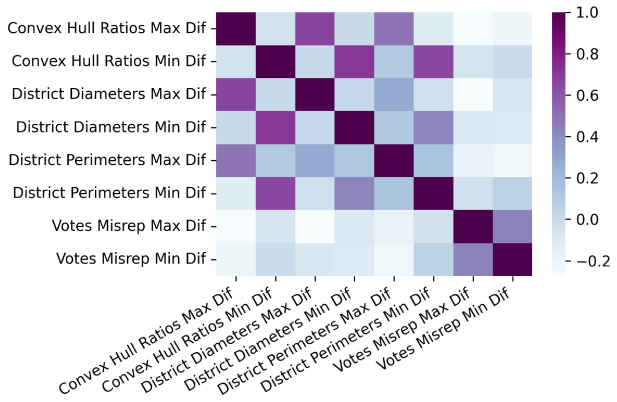

Figure 4: Correlation matrix of feature values used in maps

in Figure 5. The values in this figure show the distribution of the difference in the maximum of each feature value for the agents who selected the first map. The feature values for agents who selected the second map were negated. When we checked the distribution of the calculated feature values, we found that only those related to the shape of district showed any trend.

In addition to the importance question at the end, we also had the participants select which features they considered for every comparison they made. These features were the same as the features asked in the final question, namely *Shape*, *Winner of election*, and *Population distribution of districts*. To find out if the participants actually considered these reported features when making their decisions, we examined the same distribution for different groups separated by their response to which features mattered the most. However, we found that the trend was almost the same for all groups. This apparent contradiction led us to conclude that even though some participants reported having considered other features, the most dominant feature in terms of correlation to the final decisions are all shape-based.

Participant Reported Feature Importance

In the final question of the survey, we asked the participants to share the ranking of features in order of their perceived importance, i.e. higher score for more important features. This question can be seen in Figure 3. The responses to these questions indicate that most people place the 'Shape' feature in either the top or second importance. The second most popular feature was the distribution of the political preferences inside the districts, followed by the final winner of the election.

It is interesting to note that even though some participants indicated that they considered features other than shape, our analysis show that most groups seemed to only consider the shape when making their choices in the comparison. This odd behavior may be due to the fact that the district shape the first feature apparent upon inspection or that population distribution or election simulations are more complicated for people to consider fully. The distribution of the ranking of these features can be seen in Figure 6.

Written Responses

We also looked at the responses to the written question included at the end of the survey. While not all participants were interested in sharing an in-depth description of their own reasoning, we found that some participants provided insightful comments for their decisions. Some of these more useful comments provided by the participants can be seen in Table 3. The distribution of the frequency of the top 20 mentioned words are shown in Figure 7. Interestingly, we found that even those who reported that the winner of the election mattered the most still mentioned the words 'shape' and 'look' most frequently, suggesting that even though the respondents claimed to have considered the election results first, they still considered the shape to be more important. These words are highlighted in red in the figure.

Experimental Results

To further analyze the dataset we collected, we made use of well established machine learning models to fit our data. In our experiments, we used logistic regression, linear regression, linear SVM, decision trees and XGBoost. These models were chosen for their interpretable nature, so that we can examine the importance of each feature learned by them. We found that training the models on the entire dataset, or even training on dataset filtered by the reported feature importance yields poor results, with accuracy scores around 60%. However, when training the classifiers for each individual, we found more success. Each participant has a unique decision making process, and each model should learn the individual preferences rather than learning on the entire population. Each participant responded to five questions on map pairings, and by flipping their responses and inverting the feature values, we used ten data points per individual. This was possible because our questions were pairwise comparisons which resulted in binary labels of 1 and -1. By flipping the labels and negating the feature differences, we were able to augment our dataset for each agent. These models were used for a classification task, where the input vector was the feature difference of the two maps, and the target label was 1 or -1, depending on if the first map was chosen to be more fair or not.

Learning from Individual Participants

For individual participants, we split the ten data points into a training set of eight samples and tested on two samples.

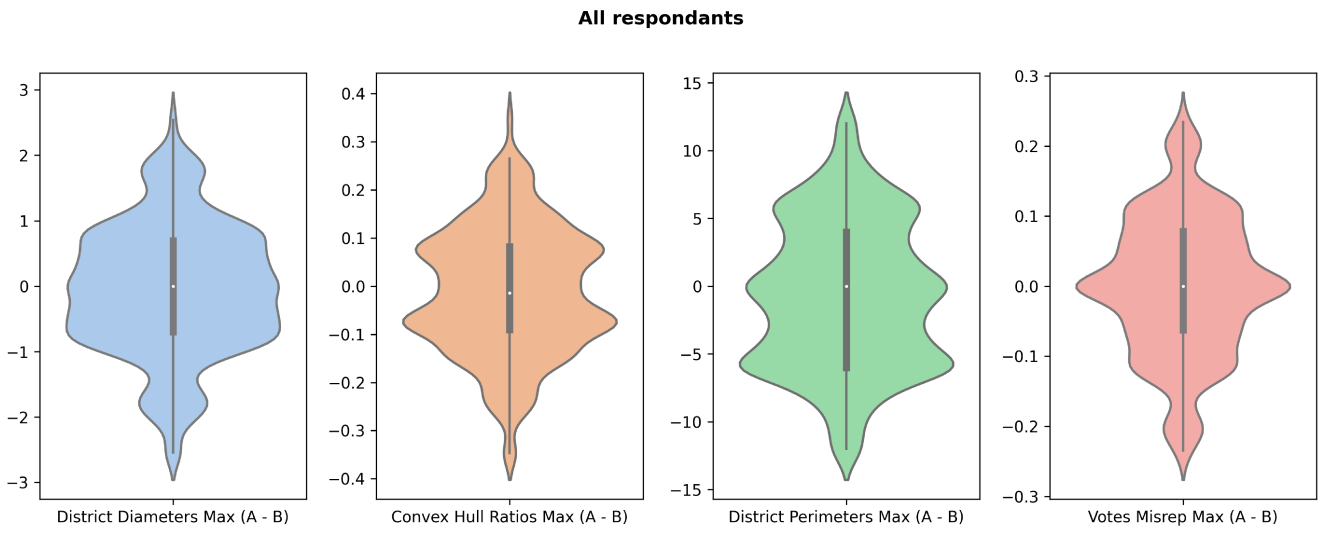

Figure 5: Distribution of difference in feature values.

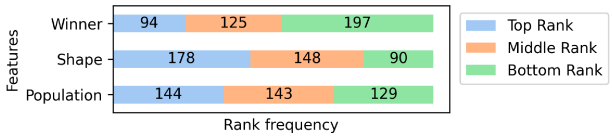

Figure 6: Distribution of the importance ranking of features.

After training each model, we then applied 5-fold cross-validation to ensure that they are not overfitting. This division is especially important because of the small size of our individual datasets. Because each individual dataset contained ten data points split evenly between the two labels, 5-fold was the maximum number of folds that could be applied by taking 1 sample of each label for each test set.

We found that among the tested models, logistic regression and linear SVM performed the best. The averaged performance of these two models can be seen in Table 4. These linear models are also useful for interpretation because they contain a scalar weight for each feature. By looking at the learned weights for each feature, we can deduce which features the model considers to be more important. We find that perimeter, especially the maximum perimeter of a district in each map, has the highest weight for both classifiers. Figure 8 shows the mean weights assigned to each feature by the two classifiers. This result matches our earlier analysis that shape of the district seemed to play the largest role in determining the perceived fairness of a redistricted map. The importance of features related to the election outcomes did not appear to play a large role in either classifiers, further supporting the notion that the shape of a district is most predictive of the perception of fairness.

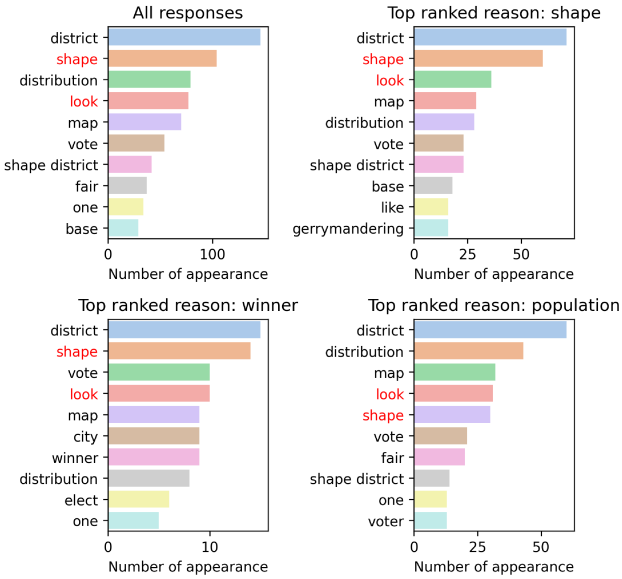

Figure 7: Appearance frequency of top 10 words grouped by the most important feature selected by participant.

Comparison of Results and Hypotheses

We had originally hypothesized that features related to both distribution of seats and the shape of the districts would be predictive of how a district assignment would be perceived. The results of our analysis indicate that we were partially correct in our hypothesis. Our analysis showed that features related to shape were predictive, particularly district perimeter. We were surprised by this result only in that district perimeter is a less suggested metric for evaluating district shape. Other metrics such as the ratio of the area of the convex hull to the area of the district performed worse. This

Top feature	Response
Shape	*"I strongly prefer a more even geographical distribution for districts."* *"To me, roughly equal populations in simple shape voting districts is the best situation. Any attempt to draw odd shape boundaries to skew elections is not a good situation."*
Winner	*"I tried to be as fair as possible, so the majority gets reflected, and second try to give every group a representation if possible."* *"I ranked the fairness by winner of the election, then by regularity of the shapes."*
Population	*"There should be an approximate equal number of different voters in each district."* *"How fairly voting blocs were distributed was my primary concern. I noticed that in a lot of them, one of the three were broken up so they couldn't really have a cohesive voting bloc. Sometimes it was done so only one really had an advantage."*

Table 3: Sample responses from groups who ranked each feature most important.

Model	Train	CV	Test
Logistic Regression	0.9926	0.8728	0.9036
Linear SVM	0.9033	0.7542	0.7294

Table 4: Average performance of the 2 best models.

poor performance may be attributable to having consistent area across districts or the fact that we used a grid-based approach to represent the map, as there was less variation in the feature. We discuss some other limitations of our map designs in regards to shape metrics below. Nevertheless, we were pleased to see that we were able to find a shape feature that performed well on our dataset, particularly when applied to those who selected that they used the shape in their decision process. Even though the correlation matrix in Figure 4 shows that voter misrepresentation is not highly correlated with the shape features, most of our participants seemed to have considered the shape features first.

Our hypothesis that people would use the party distribution and the seat distribution was challenged, even when looking at just those who indicated that they primarily considered those metrics when making their decisions. Even though our maps were generated from a fixed population distribution, we designed the maps so that sufficient amount of them have voting misrepresentation. We expected those features to be the most important to our respondents,

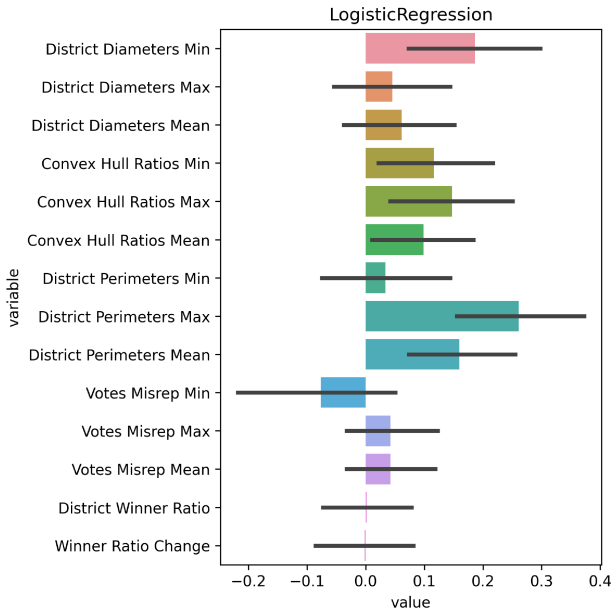

Figure 8: Feature importance learned by logistic regression.

but during our analysis, none of these features ended up showing a strong correlation. Since a change in distribution of seats is often the primary and more obvious effect of partisan gerrymandering, those changes not being an important factor in how people viewed the district maps was surprising and suggested that our participants cared more about the aesthetic shape of the districts than any political difference. This discovery might indicate that people are more easily fooled by 'normal' looking manipulated districts and less accepting of 'fair' maps if they happen to look strange. We might then conclude that when attempting to draw bipartisan and fair districts the seat distribution is the primary concern, but a secondary concern should be ensuring the map has 'normal' shaped districts to allow the map will also be perceived as such.

Limitations and Future Work

As previously mentioned, the design of our maps is intentionally simplified in an attempt to make the questions more friendly for a crowdsourced environment. However, this representation does bring with it some limitations that should be discussed. In order to not overwhelm participants with information, we limited the questions to maps with uniform population density. While easier for respondents to understand and inspect visually, real world populations are far more complicated, which might affect how people perceive maps drawn on those voter maps. Our number of districts was also limited to four, which was intended to make our survey quicker and easier to hold respondents attention. California (at time of writing) has 53 districts, which raises the complexity of the problem significantly. Additionally, all districts in our maps have the same area, while real world

maps might have massive districts for rural areas while other districts representing cities are far smaller. This difference is particularly important since political parties with large support in rural areas were found to have stronger ability to gerrymander districts (Guest, Kanayet, and Love 2019). A difference in district size could be a potential feature used by respondents, and so further work should be done to investigate its effect. Additionally, some metrics like the ratio of district perimeter to district area change based on the size of the district, and may add nuance to its correlation with how participants perceive the districts. We also assume the region being divided has rather plain geography, lacking obstacles like bodies of water or mountains. These features often cause districts to appear strange as they are shaped out of geographic necessity, even including some non-contiguous districts.

One additional limitation is that a dataset such as this would presume to know the outcomes of elections before they occur on the level of individual voters. Obviously, real world election data lacks such precision and so if the public were to be presented a real world redistricting plan they could not be certain of exact election outcomes. Especially for highly competitive or so-called 'toss-up' districts, presenting voters with a guarantee of a certain elections outcome is overly optimistic. Similarly, we assume all voters fall into only three constant political parties, and while many citizens will vote consistently along party lines in most elections, this assumption does discount possibilities for 'swing voters' and differences in popularity between candidates of the same party.

By asking participants to only compare between two possible assignments and choose which they thought was most fair, we also are unable to answer the binary question of whether people believe an assignment is gerrymandered or not. This question is almost an entirely separate problem, as it requires learning how people would define 'gerrymandered' and likely would require a broader context for questions. This question is critical as well and worth exploring in the future.

Our dataset only considers certain kinds of gerrymandering, namely partisan gerrymandering where the goal is advantage or disadvantage members of a certain political party or group. While much of the gerrymandering in the United States and other regions is done along these political lines, other kinds of manipulation exist that are not covered by this model. For example, there have been attempts to move district boundaries to separate popular representatives from their current constituents to lower their chance of reelection. These and other of these forms of gerrymandering are important to consider, but add too much complexity that we were hoping to avoid. Thus, future work could focus on how to communicate these other kinds of issues to participants.

Finally, our participants were likely from many different demographic groups including multiple nationalities. It is possible that a person's political affiliation or nationality might influence how they would perceive such maps. For example, the prevalence of gerrymandering in the United States might bias its citizens on the issue compared to a citizen of a country which has outlawed gerrymandering. In the future, if data is collected to use on applications specific to the United States or another country, it would be wise to ensure respondents were residents of that country.

Conclusion

In this work, we present a dataset containing people's perception of which redistricting map is more fair. Each participant compared five pairs of our synthetic maps, and marked which one appeared more fair. The participants also reported the features they considered, as well as their reasoning behind their choices for each pair of maps presented. We found that shape-related features were the best indicators for people's perception of fairness in given maps. Although we did our best to include simplified information about the political layout of the region and describe the election outcomes with and without the districts, it seemed that our participants' final choices were not heavily influenced by these factors. In fact, even the participants who reported they considered either population distribution or election winner more than the shape features seemed to have made their decisions mostly based on the shape of the districts.

We initially assumed that both the shape and the seat distribution of the district assignment would be predictive of perceived fairness, but upon receiving our data some of our assumptions were challenged. From our initial analysis of the dataset, we discovered that the shape of the districts appeared to be far more important than proportional seat distribution to people viewing the assignments. From this insight, we conclude that to ensure public trust when drawing fair districts, it is important that the districts are shaped in a reasonable way, perhaps even more so than equal and fair distribution of seats. This finding was surprising and warrants further research to investigate the extent to which real world maps can be manipulated while still appearing 'fair' to many individuals.

The topic of gerrymandering has been gaining more public attention. While it is important that districts are designed with more care to their influence on the election results, it is also important that policy makers should be able to convince the public of the design's fairness. With this work, we hope to shed some light on what non-experts may consider fair. We also hope that our dataset can spark the discussion about the human perception of district design, and add to the ongoing discussion of fair district design by bringing up the perspective of regular voting citizens.

Acknowledgements

We thank all anonymous reviewers from HCOMP '22 for their time and feedback on this work. LX acknowledges NSF #1453542 and #2007476, and #2106983 for support.

References

Borodin, A.; Lev, O.; Shah, N.; and Strangway, T. 2018. Big City vs. the Great Outdoors: Voter Distribution and How It Affects Gerrymandering. In *IJCAI*, 98–104.

Bycoffe, A.; Koeze, E.; Wasserman, D.; and Wolfe, J. 2018. The atlas of redistricting. *FiveThirtyEight*.

Clelland, J.; Colgate, H.; DeFord, D.; Malmskog, B.; and Sancier-Barbosa, F. 2022. Colorado in context: Congressional redistricting and competing fairness criteria in Colorado. *Journal of Computational Social Science*, 5(1): 189–226.

Cohen-Zemach, A.; Lewenberg, Y.; and Rosenschein, J. S. 2018. Gerrymandering over graphs. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, 274–282.

DeFord, D.; Dhamankar, N.; Duchin, M.; Gupta, V.; McPike, M.; Schoenbach, G.; and Sim, K. W. 2021. Implementing partisan symmetry: Problems and paradoxes. *Political Analysis*, 1–20.

Garg, N.; Gurnee, W.; Rothschild, D.; and Shmoys, D. 2021. Combatting Gerrymandering with Social Choice: the Design of Multi-member Districts. *arXiv preprint arXiv:2107.07083*.

Guest, O.; Kanayet, F. J.; and Love, B. C. 2019. Gerrymandering and computational redistricting. *Journal of Computational Social Science*, 2(2): 119–131.

Herschlag, G.; Kang, H. S.; Luo, J.; Graves, C. V.; Bangia, S.; Ravier, R.; and Mattingly, J. C. 2020. Quantifying gerrymandering in north carolina. *Statistics and Public Policy*, 7(1): 30–38.

Katz, J. N.; King, G.; and Rosenblatt, E. 2020. Theoretical foundations and empirical evaluations of partisan fairness in district-based democracies. *American Political Science Review*, 114(1): 164–178.

Ramachandran, G.; and Gold, D. 2018. Using outlier analysis to detect partisan gerrymanders: A survey of current approaches and future directions. *Election Law Journal: Rules, Politics, and Policy*, 17(4): 286–301.

Wang, S. S.-H. 2016. Three tests for practical evaluation of partisan gerrymandering. *Stan. L. Rev.*, 68: 1263.