Measuring and mitigating voting access disparities: a study of race and polling locations in Florida and North Carolina

Mohsen Abbasi
mohsen.h.abbasi@gmail.com
University of Utah
USA

Calvin Barrett
calvinpennbarrett@gmail.com
Haverford College
USA

Sorelle A. Friedler*
sorelle@cs.haverford.edu
Haverford College
USA

Kristian Lum
kristianl@uchicago.edu
University of Chicago
USA

Suresh Venkatasubramanian
suresh@brown.edu
Brown University
USA

ABSTRACT

Voter suppression and associated racial disparities in access to voting are long-standing civil rights concerns in the United States. A history of violent explicit discouragement has shifted to more subtle access limitations that can include long lines and wait times, long travel times to reach a polling station, and other logistical barriers to voting. Our focus in this work is on quantifying disparities in voting access pertaining to the overall time-to-vote, and how they could be remedied via a better choice of polling location or provisioning more sites where voters can cast ballots. However, appropriately calibrating access disparities is difficult because of the need to account for factors such as population density and different community expectations for reasonable travel times.

In this paper, we perform one of the first large-scale studies of voter access to polling locations, using real-world voter data from Florida and North Carolina in the 2020 general election. We develop a methodology for the calibrated measurement of disparities in polling location "load" and distance to polling locations based on a novel normalized distance metric to model the voter experience of distance. We find that voter turnout is reduced when this normalized distance to polling locations increases, and that non-white voters had to travel further to the polls in Florida (using this normalized distance) than White voters. We also introduce algorithms, with modifications to handle scale, that can reduce these disparities by suggesting new polling locations from a given list of identified public locations (including schools and libraries). The developed voting access measurement methodology and algorithmic remediation technique demonstrates that better polling location placement is possible.

KEYWORDS

Voting Access, Fairness, Measurement, Clustering

*Corresponding author

1 INTRODUCTION

Convenient access to voting is a crucial component of U.S. elections, and such access has been shown to have an impact on voter turnout—voters who have to travel further to the polls or wait longer are less likely to vote [9, 25, 42]. Longstanding historical and continuing discriminatory efforts have targeted and suppressed Black voter access to the polls in the United States [4]. For example, in the 2016 US presidential election, voters in predominantly Black neighborhoods waited 29% longer at polling locations than those in white neighborhoods [11]. In this paper, we study *voting access* and racial disparities through the lens of distance from voters to their polling locations as well as the number of people per assigned polling location in Florida and North Carolina in the 2020 U.S. general election.

1.1 Related Work

There exist a rich literature on polling location placement and the way that the "cost" of voting impacts voter turnout, with "cost" including the distance to polling locations [7, 10, 20, 25], voter wait times [11, 42], and other logistical barriers [16]. There is general wide-spread agreement across this literature that increasing distance to polling locations decreases voter turnout [9, 10, 20, 25], though studies in some locations find no effect [7]. Mechanisms for study include individual voter distances to polling locations [20, 28, 30], precinct-based averages [25], as well as analyses that consider discrepancies on the boundaries of precincts [7, 10] and changes to polling locations [9, 15] as ways to control for multiple other factors. Studies have been conducted at the municipal [28] and county level [9, 20, 25], including counties with both urban and rural sectors [20], and recent state-wide studies using precinct boundary [7] and polling location change [15, 49] designs. To our knowledge, there have not been state-wide studies of polling distance and turnout with a design that allows for the inclusion and analysis of all state voters.

Mohsen Abbasi, Calvin Barrett, Sorelle A. Friedler, Kristian Lum, and Suresh Venkatasubramanian

The role distance to polling stations plays in turnout is complicated by and correlated with income, race, neighborhood density, and other factors that make it hard to directly examine the impact of distance [7, 10, 20, 25, 28, 45]. Some studies have documented nonlinearities in the relationship of voter turnout to polling locations, with rural voters far from polling locations demonstrating a surprising increase in turnout [20, 25]. Hypotheses for this phenomenon include increased use of mail-in voting [7, 15, 20, 24] and more direct and uncongested travel routes in rural areas [25]. There is also evidence that vehicle availability plays a role in voter turnout based on polling location distance [10, 28], and some studies have considered the potential for political manipulation of polling site locations for partisan gain [30, 45, 47]. Attempts to disaggregate race from other correlated factors have found that Black and Hispanic voters on average live closer to the polls, but do not exhibit differences in sensitivity to polling location distance [7] or evidence of race-based partisan manipulation of polling locations [45]. Voter cost as measured by waiting time has been shown to differ markedly by race [11] and longer wait times decrease future voter turnout [42].

A few efforts have considered the placement and impact of better placement of polling locations [7, 10, 30]. These efforts focus on increasing overall turnout by considering a change in placement of polling locations in: one precinct by hand as a demonstration [30], across nine municipalities by considering census block populations with weighted distances to polling locations [10], and in one state by determining a voter turnout model and choosing a turnout-maximizing location. However, there has been little to no research emphasizing *equitable access* to voting. The main algorithmic tools we use are drawn from the literature on *fair clustering* [3, 6, 12–14, 33, 34, 37, 41, 43, 44, 48], a sub-field of algorithmic fairness [2, 19, 22, 27, 35, 50] which has gained a lot of attention in recent years. Fair clustering has often been used as a tool for redistricting [23, 36, 39] and even fairness in redistricting [46]. The closest related work to the fair polling location algorithms we introduce are in the context of facility location (e.g. [17, 18, 38]); however, previously proposed solutions often are either tailored to specific problems or cannot scale to the data sizes we consider here.

1.2 Our Contributions

In this paper we develop a novel methodology to quantify and calibrate disparities in access to voting using measures that act as proxies for the time taken to travel to a polling station and the waiting time to vote. This investigation is conducted within a specific context: that of racial voter access disparities in Florida and North Carolina during the 2020 general election. Previous studies have identified that there are many confounding factors when analyzing the experience of distance to a polling location, such as vehicle access and community norms for reasonable travel times [7, 10, 20]. We introduce a normalized distance based on distance to the closest school or library that allows voter turnout trends to be more clearly understood and racial differences analyzed. The use of this normalized distance allows a study design that includes all voters across two states with individual geocoded distances to polling locations.

We find a clean trend of decreasing voter turnout when the normalized distance increases. We found that when normalized distances are considered, non-white voters in general had to travel further to the polls in Florida while in North Carolina any racial disparity in travel was smaller. There was some racial variation in "load" (our proxy for the waiting time to vote) but this did not suggest any substantive impact. We further develop scalable algorithms that can redistribute polling locations so as to improve access disparities. These polling location placement algorithms are able to use realistic options for polling sites, such that their outcomes are actionable. Our algorithmic interventions (both to minimize travel time as well as waiting time) improved access disparities across the board, while making visible any tradeoff between disparity mitigation and resources required.

2 DATA COLLECTION AND PREPROCESSING

To conduct our study, we need voter registration information (including voter race and address), polling location information, as well as latitude and longitude for each address as well as for alternate or new polling locations introduced. This latter determination is known as *geocoding*, and is relatively difficult and costly. The final data collected presents additional privacy risks to voters beyond what is publicly available and thus we do not make this data public.

2.1 Voting rolls and polling location data

Voting rolls, or voter registration data, usually include the name, address, other voting-related information, and race for all registered voters. Florida and North Carolina were chosen as focus states for our case study because the states are large, have substantial non-white populations, and provide freely available voting rolls that include race information.[1] There are approximately 15.1 and 7.3 million voters in the records for Florida and North Carolina, respectively. Data collected for both states for each voter included: voter's residential address; county, precinct, and congressional district information to determine the voter's associated precinct location; and race as defined in the voting rolls. For the analysis of voter turnout in Section 3.3 this data was augmented with information about whether the individual voted on Nov. 3, 2020 (available from the same sources).

Both Florida and North Carolina require their residents to vote at the polling place designated to their precinct,[2] and we use the terms polling location and precinct site interchangeably. We collected all polling location addresses for Florida and North Carolina[3] and determined each voter's polling location based on their listed county and precinct information from the voting rolls. 6068 and 2662 polling sites are listed in the Florida and North Carolina records, respectively. In order to calibrate distances to local understandings of

---

[1] Florida voting rolls originally accessed via http://flvoters.com/downloads.html, now available at https://web.archive.org/ or by mail request from https://dos.myflorida.com/elections/data-statistics/voter-registration-statistics/voter-extract-disk-request/. Our analyses use the 10/13/2020 data. North Carolina voting rolls are available at: https://www.ncsbe.gov/results-data/voter-registration-data. Our analyses use the 11/03/2020 snapshot.

[2] https://dos.myflorida.com/elections/for-voters/voting/election-day-voting/, https://www.ncsbe.gov/voting/vote-person-election-day

[3] Florida polling locations were downloaded from https://dos.myflorida.com/elections/for-voters/voting/. North Carolina polling locations were downloaded from: https://www.ncsbe.gov/results-data/polling-place-data. Our analyses are based on the November 2020 general election polling locations.

travel times (see Sect. 3) and to identify potential alternative polling location sites (see Sect. 4), we collected the addresses of all schools and libraries in Florida and North Carolina.[4] In total, the addresses for 5128 and 3265 schools and libraries were collected in Florida and North Carolina, respectively. Schools and libraries were chosen since these are public sites that are generally well-distributed around a state and are often sites chosen to serve as polling locations.

2.2 Geocoding addresses

To measure voter distance to the nearest polling location, we need to geocode voter residence, polling location, and schools or libraries that could serve as possible polling locations. We used the ArcGIS tool[5] for our geocoding queries. In the ArcGIS geocoder, a *match score* controls how closely addresses have to match their most likely candidate in the reference data. The minimum match score is set to the default value (80), as are all other internal parameters. When geo-locating an address, the address is compared to records in a database which each have an associated latitude and longitude. The record with the highest match score with that address is chosen as the match and that record's associated latitude and longitude are assigned to the address. If there are no records with a match score above the minimum match score, no match is returned. By using the default minimum match score of 80, we allow for some spelling mistakes and other slight address differences between the address and its match, while maintaining a high degree of match accuracy.

2.3 Missing data analysis

In order to ensure that no eligible voters were missing from our voting rolls, we chose the latest possible voting roll snapshot that occurred before the 2020 general election. To validate its completeness, we compared to the public voter turnout rates. In North Carolina, we calculate an overall voting rate of 75.0%, which matches exactly with official reports on voter turnout.[6] After dropping individuals without a geolocated home address or polling location, we still calculate a 75.0% voting rate. In the Florida data, we found that some voters who did vote in the Nov. 3, 2020 general election were not present in the latest pre-election voter roll data (Oct. 13, 2020), perhaps because of a lag in reporting registrations. Additionally, the voter rolls we consider include more people than represented in the official turnout numbers, possibly due to voter removal from the rolls just before the election[7] or due to differences in which voters are countered as part of this rate. The resulting voter turnout rate in our data is 72.2%, which differs from official reports of 77.2%.[8] After dropping individuals with polling locations or addresses that we were not able to geocode, we are left with an overall rate of 71.6%.

We proceed with our analyses using this Oct. 13, 2020 snapshot of Florida voting rolls.

Our lists of valid polling locations, schools, and libraries are fully complete.[9] However, the geocoding process did not successfully return latitude and longitude matches for all given addresses; in some cases, the provided address was misspelled, incomplete, or potentially missing from the ArcGIS database. Within each state, approximately 91 percent of voter addresses were successfully geocoded. In Florida, 6068 out of 5159 polling locations were successfully geocoded, and in North Carolina it was 2250 out of the 2662 polling locations. Of the school and library addresses, 4499 and 2711 were successfully geocoded and used in the analyses for Florida and North Carolina, respectively.

Voters with residential addresses or polling locations that could not be geocoded were excluded from the data used for analysis in the remainder of this paper, as were polling locations that could not be geocoded. Specifically, in all the following analyses 11.8 and 5.3 million voter records were included for Florida and North Carolina, respectively. In order to ensure this missing data did not skew our voting access disparity analysis, we conducted a missing data analysis. We verified that a) the geocoding success rates were not themselves racially skewed, and showed that b) the resulting racial distributions for both states were similar to the original distributions (see Appendix A for details).

3 VOTING ACCESS DISPARITIES IN FLORIDA AND NORTH CAROLINA

In this section, we examine the accessibility of polling sites in the 2020 general election by race as categorized in the Florida and North Carolina voter registration data. We consider two measures of access to polling places:

Distance: We use the great-circle distance[10] as a proxy for the travel distance/time it takes for voters to get to their assigned polling site.

Polling site load: For each voter, we use the number of voters assigned to a polling place as a proxy for their wait time. We assume that all polling sites process votes at the same rate.

These measures are not be perfect proxies for the *time* it costs voters to participate in elections. Geographic distance does not account for variability in travel times in different areas - such information is available via APIs, but was not viable for this analysis given the expense. The polling site load values do not account for differences in number of machines, impact of voter identification requirements, or other voter processing time differences that may be significant and different by demographic group [32]. However, these measures provide us with an attainable yet useful enough tool to analyze and compare voters' access.

3.1 Distances and polling site loads

The results in Figure 1 give the distance and polling site load for voters, broken down by race, in Florida and North Carolina. We observe that in Florida, White voters travel the longest distance

---

Mohsen Abbasi, Calvin Barrett, Sorelle A. Friedler, Kristian Lum, and Suresh Venkatasubramanian

	Florida						North Carolina				
	Distance (miles)			Polling Load (# Voters)			Distance (miles)			Polling Load (# Voters)	
Race	Mean	Med.	Max	Mean	Med.	Race	Mean	Med.	Max	Mean	Med.
Wht. not Hisp.	1.26	0.92	35.32	4805	4272	White	1.69	1.27	40.67	3503	3105
Hispanic	0.89	0.68	28.51	4374	3824	Black/Afr. Amer.	1.32	0.92	21.55	3581	3106
Blk. not Hisp.	0.87	0.66	35.32	4160	3604	Undesignated	1.51	1.10	25.68	3750	3302
Unknown	0.97	0.73	35.32	4460	3905	Other	1.40	1.03	16.87	3922	3418
Asn./Pac. Isl.	1.05	0.85	17.36	4903	4395	Asian	1.23	0.95	18.18	4204	3740
Other	1.04	0.79	20.04	4758	4237	Amer. Ind./AK Nat.	1.97	1.54	15.99	2751	2397
Multi-Racial	1.02	0.78	16.11	4729	4216	Two or more races	1.32	0.94	16.06	3802	3327
Amer. Ind.	1.24	0.86	22.96	4656	4079	Nat. Haw./Pac. Isl.	1.65	1.16	9.37	3922	3326
Overall	1.12	0.82	35.26	4628	4079	Overall	1.58	1.16	40.62	3561	3127

Figure 1: Florida (left) and North Carolina (right) distance and polling site load values, broken down by race. Distance is measured based on the mean, median, and maximum distance (in miles) for voters to their closest polling location. Polling side load gives the mean and median values for number of voters assigned to each racial group's polling locations. Maximum values are the same across all groups; the maximum size polling location for Florida is 18369 and for North Carolina is 13669. The (non-weighted) numbers across all voters is given for all measures in the last row.

to polling locations among the groups with respect to both mean and median values. Asian or Pacific Islander voters have the largest number of voters at their respective polling places. In North Carolina, American Indian voters experience the longest trip to polling locations while Asian voters encounter the most crowded polling locations. Prior studies have shown that variations in distance to polling sites, even as small as those observed in Figure 1, can have a significant effect on voter turnout [28].

3.2 Normalized distances and polling site loads

The same distance to the polls might be perceived differently in rural as opposed to urban areas, due to factors such as differences in road traffic, differences in access to personal vehicles, or the norm for everyday travel distances (see, e.g., [25]). In order to determine whether the disparities we observed based on the measured distances to polling locations (in miles) are burdensome or otherwise unusual in voters' lives, we next explore two different methods for *normalizing* these distances.

Distance Normalization: Schools and libraries. In the first, we consider the distance to a polling location relative to other distances that voters might travel regularly and which might be considered reasonable. For this analysis, we choose to use schools and libraries as reference locations, with the motivation that these sites are distributed so that all residents have access to a local school and library. An important caveat is that this choice may hide voting access disparities that accrue to the same voters as school and library access disparities. However, our additional analyses in Sections 4 and 5 will consider the possibility of opening new polling locations, and since schools and libraries also serve as good polling locations, they provide for a useful analysis reference.

Distance Normalization: Median pairwise distance to other voters. In order to normalize based on residential density (which may impact voter turnout [25]), we estimate how densely populated each voter's neighborhood is based on the median of their pairwise

distances to other voters in their precinct. For each voter, we find this median and divide the voter's absolute distance to their polling site by that median.

Load Normalization. In order to normalize the polling site loads, we divided each group's median site load by that of the majority group, i.e. White voters in both states. Thus, White voters will always have a normalized load of 1.

In Figure 2 the normalized distances to voting locations and loads are presented for Florida and North Carolina voters. We note that although called distances, these values are ratios and cannot be compared to the values in Figure 1 directly. Instead, our goal here is to bring into attention how considering normalized distances changes the ordering among different races in terms of their distance to polling locations. In the state of North Carolina for example, on average American Indian voters experience the largest distance to polling locations while Black voters have the shortest travel distance when absolute distances are considered (Figure 1). However, when normalized distances are considered as presented in Figure 2, the conclusions change. When normalizing to the nearest school or library, Black voters face the longest average travel distance to their polling site and White voters the shortest in both Florida and North Carolina. In Florida, this pattern is the same when considering the median of school/library normalized distance. In North Carolina, this pattern changes when considering the median. Normalizing to voter density tells a slightly different story, particularly in North Carolina, where Asian voters experience the longest mean density-normalized travel distance and American Indian and Alaskan Native the shortest. This highlights the importance of careful consideration of how distance to polling location is measured. Different choices among reasonable alternatives can lead to different conclusions about which group is most privileged with respect to polling access. Ultimately, the numbers themselves cannot tell us which normalization scheme (if any) is most appropriate and we must rely on common sense and contextual understanding to make

	Florida, normalized access						North Carolina, normalized access				
Race	Normalized Distance school/lib.		voter dist.		Norm. Load	Race	Normalized Distance school/lib.		voter dist.		Norm. Load
	Mean	Med.	Mean	Med.			Mean	Med.	Mean	Med.	
Wht. not Hisp.	1.65	1.09	1.19	1.02	1	White	1.69	1.01	1.08	0.98	1
Hispanic	1.86	1.26	1.35	1.10	0.89	Black/AfAm.	1.95	1.05	1.09	0.98	1.00
Blk. not Hisp.	1.98	1.24	1.20	1.02	0.84	Undesignated	1.83	1.04	1.11	0.99	1.06
Unknown	1.84	1.21	1.29	1.06	0.91	Other	1.83	1.06	1.14	1.02	1.10
Asn./Pac. Isl.	1.90	1.22	1.26	1.07	1.02	Asian	1.82	1.07	1.21	1.06	1.20
Other	1.84	1.18	1.26	1.05	0.99	AmerInd./AKNat.	1.91	1.00	0.92	0.90	0.77
Multi-Racial	2.02	1.22	1.26	1.05	0.98	Two+ races	1.90	1.05	1.13	1.01	1.07
Amer. Ind.	1.76	1.15	1.15	1.01	0.95	NatHaw./PacIsl.	1.77	1.05	1.10	0.98	1.07

Figure 2: Mean and median distance to polling location, normalized by distance to closest school/library or median distance to other voters in the precinct. Load is normalized to the median load experienced by the majority group.

a subjective determination of which metric is most appropriate in this case.

The median of the nearest school or library-normalized distance is useful for assessing access disparities based on distance to polling location. As discussed above, it is a reasonable proxy for the distance a person regularly travels for everyday tasks. We prefer it to the median pairwise distance normalization because this method for normalization may behave unintuitively for residents of multi-family housing. For example, residents of a large apartment building would likely have very small median distance to other voters in their precinct, which may not reflect the distance they regularly travel. Going forward, we present our results in terms of the school/library-normalized distance.

3.3 Effects of distance on voter turnout

Finally, we explore the relationship between (normalized) distance to polling location and voter turnout. For each state, we place individuals into one of 100 bins based on the quantile of their calculated distance to the polls. Within each bin, we calculate the rate at which those individuals voted. The average distance to the polls for individuals in that bin is then plotted versus the rate (with a trend line produced via Loess smoothing) in Figure 3 (top row). Figures plotting the voting rate by distance percentile (which tends to exaggerate the effects of outliers along the x-axis less) as well as curves disaggregated by race are given in the Appendix. Surprisingly, voting rate *increases* as voters get farther from the polls, finally decreasing very slightly for voters very far from their polling location. We note that there are very few voters who travel these long distances, so for the vast majority the trend is simply increasing as a function of distance. This result seemingly contradicts the literature (see Sec. 1.1); however it does not account for neighborhood density, vehicle access, or other factors known to be relevant to the causal interaction between poll distance and voter turnout.

In order to account for the voter experience of distance—whether their polling location is closer or farther than locations they regularly travel—we repeat this analysis by dividing by the distance to the nearest school or library. Once again, we threshold this normalized distance, creating 100 equally sized bins, and consider the average normalized distance within that bin versus its voting rate

(see Figure 3 bottom row). From these figures, a tidier story emerges in which the farther from the polls voters live (normalized for typical travel times), the less likely they are to vote. These effects can be significant: those with polls that are about the same distance from their residence as other amenities (i.e. a normalized distance of one) in both states had voter turnout 1% and 1.4% higher than voters whose distance to the polls was twice that of the distance to other amenities in North Carolina and Florida, respectively. Similarly, voters whose normalized distance was five (i.e. had to travel 5 times farther to their polling location than they did to other public amenities) saw voter turnouts that were 2.8% and 2.6% lower. In tight elections, even small differences like these can make a difference.

This analysis thus offers a simple methodology to study the relationship between voter turnout and distance from the polls that contextualizes distance relative to the typical distances voters travel. Other large-scale state-wide research in this area has used precinct-boundary designs (including only voters on opposite sides of a precinct boundary) or polling location change designs (including only voters whose polling place has changed) [7, 10, 47], and thus necessarily excludes some voters and may miss effects if this creates selection bias. The cleanness of our results are suggestive that the normalization we introduce here identifies realistic trends and accounts for many of the factors that other research in this area has accounted for while still allowing the inclusion of all voters in the analysis and providing a true state-wide study of voter turnout.

4 ALTERNATIVE SELECTION FOR POLLING PLACES: ALGORITHMS

Having established a way to measure disparities in access to voting, can we determine where to place polling locations so as to reduce these disparities? In this section, we propose algorithms to reduce access disparities in terms of both distance and load. These methods are scalable and work on the large input sizes necessary to handle states such as Florida and North Carolina. We formulate this problem as different versions of the well-known k-median and k-center discrete clustering problems. In both these problems, we are given a set of points X (the voters) and facilities F (polling locations) in a metric space and the goal is to choose k facilities from F and

Mohsen Abbasi, Calvin Barrett, Sorelle A. Friedler, Kristian Lum, and Suresh Venkatasubramanian
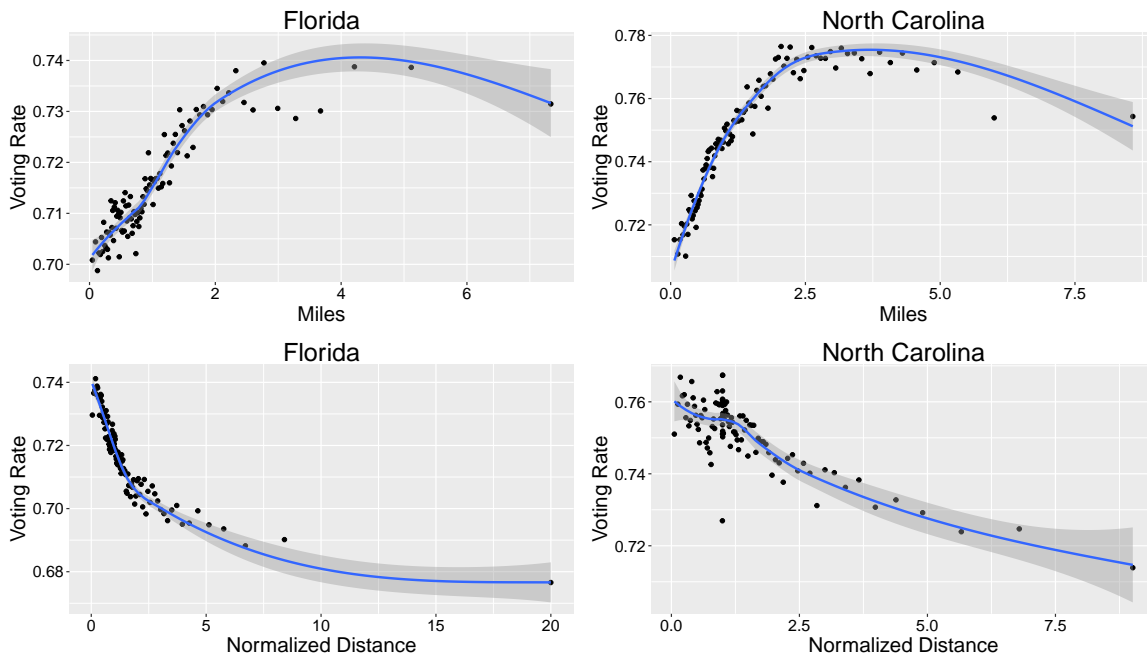

Figure 3: The average distance per bin (top row) and the average normalized distance per bin (bottom row) versus the voting rate for that bin. Bins were determined by thresholding distance such that we create 100 equally sized bins, where the ith bin represents all the people in the ith percentile of the distance from the poll distribution.

assign each point in X to its nearest facility so that the overall cost (measured either as the sum of distances or the maximum distance) is minimized. In general, one assumes that $F = X$, but we will distinguish the two sets in our setting.

We solve variants of the standard k-median and k-center algorithms with additional constraints that enforce the fair access criteria. Our approach here is an extension to the fair clustering literature which has gained momentum in the past few years [3, 6, 12–14, 33, 34, 37, 41, 43, 44, 48]. To ensure the developed algorithms can be applied to large datasets, we employ the concept of *coresets*. Given a problem (e.g. k-median), coresets are small, weighted subsets of large datasets such that the solutions to the problem found on subset are provably close to the solutions found on the original dataset [5]. Our methods first summarize the massive voter dataset X using coreset construction algorithms, and then feed the summarized versions into fair clustering algorithms to produce the desired result.

We introduce three solutions that provide more equitable alternatives to the original polling location assignment: 1) minimize the maximum distance to polling places across different race categories while ignoring load balance; 2) build upon the first method to identify location that also provide a more balanced distribution of voters, at the cost of opening additional polling locations; and 3) use an alternative approach to address both distance and load simultaneously without requiring additional polling locations. All these solutions come with their own advantages and limits.

4.1 Group fair distances

The problem of selecting a predetermined number of polling places and assigning exactly one to each voter, so that the overall distance between voters and polling places is minimized, can be formulated via the well-known problem of k-median clustering. The objective in the k-median algorithm is to select k centers (i.e polling locations) so as to minimize the sum of the distances between points and their associated center: $\arg\min_C \sum_{p \in X} \|p - C(p)\|$ where $C(p)$ denotes the center point p is assigned to. This formulation of the k-median problem can result in arbitrarily large distances for certain sub-populations within voters, as long as the overall average distance is minimized. This is a problem in the context of polling as it may hinder certain individuals' ability to cast their vote, due to longer than necessary travel distances. We formulate this via a variant of the *fair k-median* clustering problem introduced in [1] and introduce an algorithm to address this issue.

Definition 1 (Fair k-median clustering). Given m groups $X = X_1 \cup \cdots \cup X_m$, a fair k-median clustering algorithm returns k centers so as to minimize the maximum average distance to centers across all groups.

$$\arg\min_{C \in \mathcal{C}} \max \left(\frac{1}{|X_1|} \text{cost}_C(X_1), \ldots, \frac{1}{|X_m|} \text{cost}_C(X_m) \right)$$

where $\mathcal{C}$ is the set of all choices of cluster centers, and $\text{cost}_C(X_i)$ is the sum of distances for group X_i in clustering C.

In the context of polling and the concern around racial disparities, each group X_i in the above definition represents voters of a particular race (based on voter roll categories). To solve the fair

k-median clustering problem, [1] first solve an associated linear program and use a faithful rounding procedure to choose exactly k centers. When voters' addresses can be selected as centers, this method results in expected approximation factor of 4 [1]. Unfortunately, directly using this linear programming-based method is not realistic because it would be prohibitively slow or even impossible when determining polling locations for millions of voters. Therefore, instead of processing the entire input data we propose to use its *coreset* [5]. Denote the average cost of standard and fair k-median clustering on point set X as $\text{cost}_C(X)$ and $\text{cost}_C^f(X)$ respectively, where C is a set of k centers.

Definition 2 ((k, ϵ)-coreset for k-median). Given a point set X in a metric space, the (k, ϵ)-coreset for k-median is a weighted subset S of X, where for each C with size k: $(1-\epsilon)\text{cost}_C(X) \le \text{cost}_C(S) \le (1+\epsilon)\text{cost}_C(X)$

Theorem 4.1. *Given point set $X = X_1 \cup ... \cup X_m$ as input, let S_i be a (k, ϵ)-coreset for group X_i, $i \in 1, ..., m$, separately. Then for all C: $(1-\epsilon)\text{cost}_C^f(X) \le \text{cost}_C^f(S) \le (1+\epsilon)\text{cost}_C^f(X)$ where $S = S_1 \cup ... \cup S_m$.*

Theorem 4.1 (proof is in Appendix D.1) suggests that given a large point set as input one can run any fair k-median algorithm (as defined in Definition 1) on the union of its groups' coresets, and find a clustering with an objective value arbitrarily close (within ϵ factor) to that of running the same algorithm on the original set.

For our coreset construction we use the algorithm proposed by Feldman and Langberg in [21]. In the first step, it uses a bi-criteria k-median clustering algorithm as a subroutine to find initial centers, and assigns weights to each point in the original set based on its distance to the closest center.[11] In our implementation for (α, β) bi-criteria k-median algorithm we use the method due to Indyk [29]. In the second step the points are sampled according to the distribution implied by the assigned weights, and their union with bi-criteria centers are returned as a weighted coreset. If the original point set is defined in a metric space (which is the case in our dataset), then with probability $1 - \delta$ this algorithm returns a weighted ϵ-coreset of size $\frac{c}{\epsilon^2}(k \log(n) + \log(1/\delta))$ where n is the input size and c is a large enough constant. In our implementation this algorithm is used to construct a coreset for every group and the union of these coresets is fed into the fair clustering algorithm to find a distance fair assignment.

In order to empirically evaluate the effectiveness of using coresets from Theorem 4.1, we compare the results of running regular and fair k-median clustering algorithms on a sample dataset, to the corresponding values achieved on its coreset. For this purpose we randomly selected 4000 and 1000 White and Black voters from North Carolina voter records, respectively. The results of regular and fair k-median algorithms on this sample and its coreset are summarized in Appendix Figure 10; the objective values achieved by running both algorithms on the coreset are very close to the values from the entire sample data, which corroborates the results of Theorem 4.1. Both theoretical and empirical results demonstrate that we can use coresets in our analysis and achieve near optimal results. We return to assessing this algorithm on our Florida and North Carolina data in Section 5.

4.2 Group fair distances and balanced assignments

A fair k-median framing of the problem of access can address concerns around traveling to a polling station. But as we have pointed out earlier, another component of the overall time it takes to vote is the load at the polling station itself. The problem of opening and assigning polling locations to voters so that the overall distance to polling places is minimized while maintaining a balanced distribution of voters per location, can be formulated via the *balanced k-median clustering problem*[12] where each facility comes with a capacity constraint on the number of points that can be assigned to it. Previous algorithms for the balanced k-median either violate the capacity constraint or the cardinality (number of centers) constraint. In this section we opt for a method which may open more centers but maintains the capacity constraint. We start by assuming that all facilities have a uniform capacity L. We call the fair variant of this problem *L-balanced fair k-median clustering*. Bateni *et al.* [8] demonstrated how an approximate solution to the k-median clustering problem can be transformed into an approximate bicriteria solution for the L-balanced k-median clustering with slightly worse approximation factors. We closely follow their approach and show this also holds true for the fair k-median variant of the problem in the following theorem.

Theorem 4.2. *Suppose there is an α approximation algorithm for the unconstrained (no balance constraint) fair k-median problem. Then there exists a $(2\alpha, 2)$ bicriteria approximation for L-balanced fair k-median problem.*

We refer the reader to Appendix D.2 for the proof. Theorem 4.2 gives us an $8 + \epsilon$ expected upper-bound to an instance of L-balanced fair k-median problem where at most $2k$ centers are opened. We assess these results in practice on the Florida and North Carolina data in Appendix E.2.1.

4.3 Individually fair polling assignment

The algorithms discussed in sections 4.1 and 4.2 both consider a group notion of fairness to deliver a polling place assignment with more equitable assignments by group. But we can also investigate the issue of inequitable distances to polling locations at an *individual* level with the objective that no voter is too far away from their nearest polling place. Here, we discuss such a method to minimize the maximum distance a voter has to travel to cast their vote using another well-known variant of the k-clustering problem called k-center. Given a set of data points and parameter k, the objective in the k-center problem is to select k centers from the input and assign each data point to one of the selected centers, so the maximum distance between the points and their assigned centers is minimized, i.e. $\arg\min_{C \in \mathcal{C}} \arg\max_{p \in X} \|p - C(p)\|$ where X is the set of data points and $\mathcal{C}$ is the set of all possible k-clusterings.

In our analysis we consider two variants of the k-center problem, both with and without center load constraints. As before we assume a uniform load requirement for each center. Our solution is similar: we build a core-set and then use an unconstrained algorithm to cluster the coreset. To build the coreset we use the distributed weighted balanced k-center algorithm proposed in [40] that splits

---

[11]An (α, β) bi-criteria k-median clustering algorithm opens up to $\beta \times k$ centers, which results in an objective cost smaller than α times the optimal solution.

[12]This problem is also known as capacitated k-median clustering problem

the data into chunks and runs the well-known 2-approximation due to Gonzalez [26] to generate a coreset. Once we have collected and weighted the coresets appropriately, we run a capacitated k-center algorithm that is a weighted variant of the algorithm proposed in [31]. This process has been shown to yield a 49-approximation overall [40].

5 ALTERNATIVE SELECTION FOR POLLING PLACES: EXPERIMENTS

We next assess the fair polling site selection methods introduced in Section 4 to analyze the impact they could have on reducing voting disparities in Florida and North Carolina. We evaluated these algorithms based on two options for new polling locations. In the first setting, polling places can be opened at any voter residence. This setting is used to evaluate a best case scenario and is less realistic.[13] In the second setting, the algorithm can only select polling places from a set containing the state's schools, libraries, and 2020 election polling sites. This gives a realistic bound on improvement from better polling location placement, since schools and libraries are often used as polling locations. The full set of numerical results for all algorithms, including both polling selection methods, with normalized and real valued results for each racial group, are given in the Appendix. Here, we focus on the normalized median distance and polling selection from school, library, and existing polling site locations, and on the normalized site loads. Recall from Section 3 that we see this type of normalized distance as the most useful measure of voter distance access assessed. Interestingly, even though the algorithmic guarantees and optimizations introduced in Section 4 are based on the un-normalized distances and polling site loads, we find that the fair algorithms perform well on the normalized measures.

The results, shown in Figure 4, show that both of the fair k-median variants are able to achieve a median normalized distance of 1 for all racial groups across both states. This means that the median voter travels only as far as their closest school or library under these fair algorithm variants. In North Carolina, the real polling location assignments were very close to also matching this result. However, in Florida, the new polling locations from the fair k-median variants would decrease the distance-based access disparities between groups while also decreasing these distances for all groups. Hispanic and Black voters, who experienced the largest median normalized distance, had to travel about 25% farther than the closest school or library under the real polling location assignments and can have that reduced under the fair assignments. In Florida, a 25% reduction in normalized distance translates to about half a percent more in voter turnout, assuming that the trend observed in Figure 3 holds. These reductions can translate into substantial distances for voters. For example, the difference between the minimum and maximum group median distances within Glades County in Florida is reduced from 6.26 miles in the original assignment to 1.21 miles in the assignment produced by the fair k-median algorithm.

The normalized load results demonstrate that the balanced fair k-median algorithm is the most effective at balancing the load

---

[13]Although unusual, there are some jurisdictions that do allow residences (garages) to be designated polling locations, precisely due to restrictions on the number of people per precinct (which we term "load"): https://www.good.is/articles/polling-place-garages-san-francisco

across groups, although the other methods do somewhat help to achieve balance. We considered three different capacities for both the balanced fair k-median and the balanced k-center algorithms. These capacities were $1 + \epsilon$ times the average number of voters per location, where $\epsilon \in \{0.1, 0.5, 0.9\}$ is the control variable, and average is determined by dividing the the total number of voters by the total number of designated polling places. The results in Figure 4 are for $\epsilon = 0.5$. Recall that this capacity choice sets the allowed deviation from a fixed (balanced) load across polling sites. Thus, the capacity choice impacts the number of extra polling places opened up by the balanced fair k-median algorithm, and voter distribution balance in the assignments produced by both balanced fair k-median and balanced k-center algorithms. The results are summarized in Tables 1 and 2. As expected, we see by increasing the capacity fewer additional polling sites need to be opened for the balanced fair k-median algorithm. For both algorithms we see that increasing the capacity results in a larger standard deviation of the number of voters per location, which means voters are less evenly distributed across polling sites.

These methods have distinct strengths and weaknesses and are suitable for different use cases. While the fair k-median algorithm produces an assignment with more equitable distances at a group level as well as offering shorter distances overall, it does not take into consideration the potentially unbalanced loads at polling sites. Still, in practice it performs well when assessing normalized load. The L-balanced fair k-median algorithm solves this issue by introducing a limit on the number of voters that can be assigned to a single facility. Although this method may need to open up additional polling places, this requirement could instead be used to allocate additional resources, e.g. additional poll workers or polling booths. The unconstrained (Appendix E.3) and balanced k-center algorithms introduced in Section 4.3 address equitable access at an individual level. Similar to the k-median methods, the main difference between the two is that one distributes voters evenly across polling paces while the other does not provide such a guarantee. The nice property of the balanced k-center algorithm is that it can be tuned to specify exactly how many voters above the polling place limit could be assigned to it. However, this comes at the expense of less competitive distances to polling sites.

6 DISCUSSION AND CONCLUSION

In this paper, we analyzed voting access disparities with respect to polling locations. We quantified potential racial disparities in terms of distance to nearest polling location and number of people assigned to a given polling location (its "load"). To account for natural variations in population density that might give, e.g., individuals in rural locations the expectation and means of traveling further to the polls, we introduced a methodology using a normalized distance based on the distance to the nearest school or library. Analyzing the turnout of all voters across Florida and North Carolina in the 2020 U.S. general election, we found that turnout decreased as voters had to travel further (using the normalized distance), with voters travelling twice the distance to their nearest school or library experiencing a 1-1.4% decrease in turnout relative to those who traveled only as far as their school or library, with a five times increase in distance leading to a turnout decrease of 2.6-2.8%. Black and
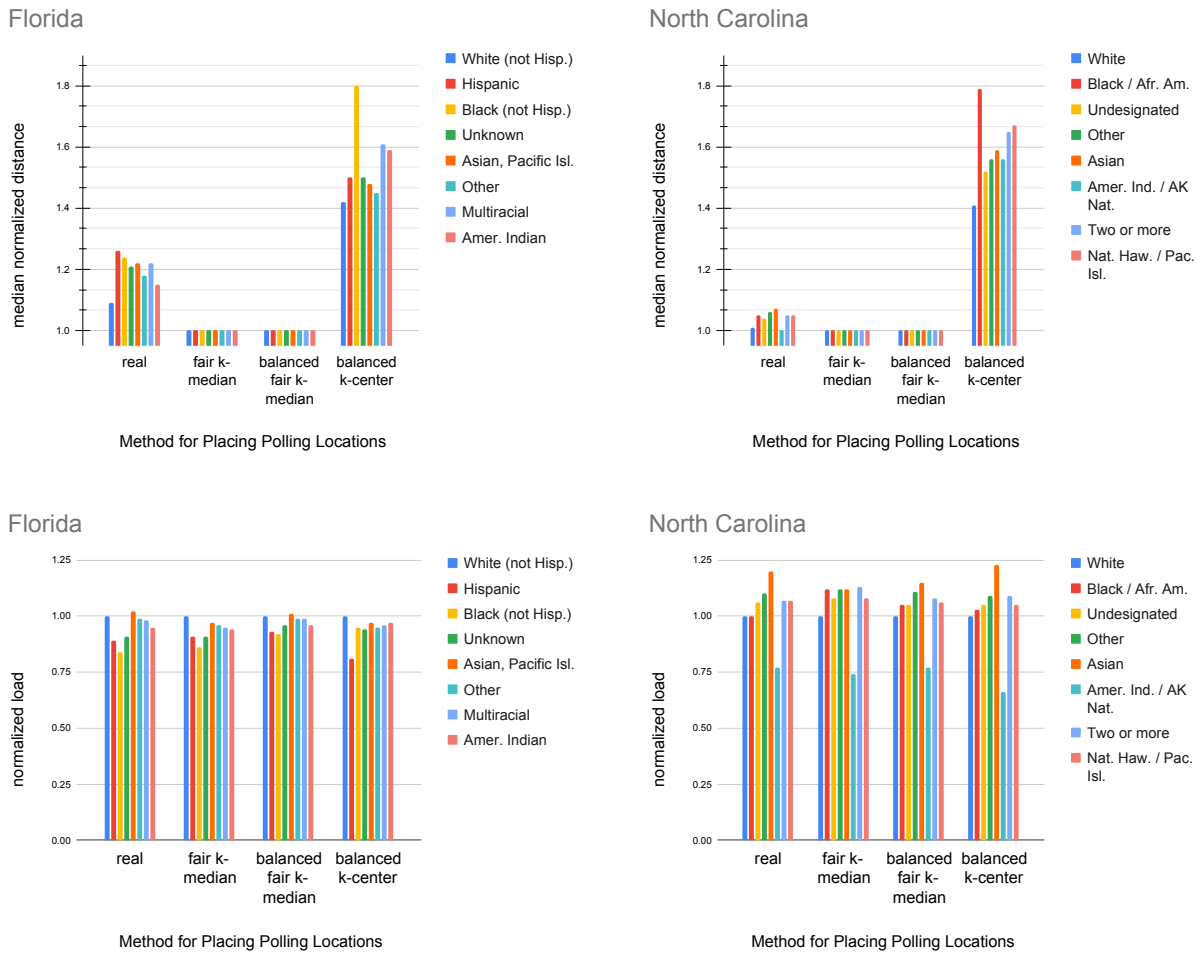
Figure 4: Comparative results for the real polling locations (based on analysis in Section 3) and fair polling location selection algorithms introduced in Section 4. Results are given for Florida (left) and North Carolina (right) based on normalized median values for distances (top) and normalized polling load values (bottom). The polling selection algorithm results shown here allow polling sites to be opened at schools or libraries. Balanced fair k-median and balanced k-center results are for capacity parameter $\epsilon = 0.5$.

Table 1: The effect of capacity in the L-balanced fair k-median algorithm on the number of newly opened polling sites and voter distribution

	Florida			North Carolina		
	$\epsilon = 0.1$	$\epsilon = 0.5$	$\epsilon = 0.9$	$\epsilon = 0.1$	$\epsilon = 0.5$	$\epsilon = 0.9$
Number of extra polling sites	1472	690	276	915	432	207
Mean #voters per location	2404	2854	3173	1871	2252	2487
Std. dev. of number of voters per location	1118	1378	1596	860	1106	1317

Hispanic voters had to travel farther to the polls in Florida, leading to a decrease in turnout of 0.5% for the median distance voter.

These voting access disparity results are subject to a number of limitations and should be seen as only the beginning of an investigation into voting access disparities. The measurement methodology we introduce here is able to quantify (normalized) distance to the

Mohsen Abbasi, Calvin Barrett, Sorelle A. Friedler, Kristian Lum, and Suresh Venkatasubramanian

Table 2: The effect of capacity in the balanced k-center algorithm on voter distribution

	Florida			North Carolina		
	$\epsilon = 0.1$	$\epsilon = 0.5$	$\epsilon = 0.9$	$\epsilon = 0.1$	$\epsilon = 0.5$	$\epsilon = 0.9$
Mean #voters per location	4628	4628	4628	3561	3561	3561
Std. dev. of number of voters per location	1503	2331	2806	1255	1821	2150

nearest polling location and the polling location load, but a voter's experience of access to their polling location depends on many additional factors. These include factors directly related to the polling location (e.g., number of voting stations or machines, time to wait in line, accessibility of the location by public transit and for voters with disabilities) as well other societal limitations that may keep people from the polls (e.g., availability of childcare, time off of work to vote, voter intimidation). While distance and load may be reasonable proxies for some of these measures, they do not capture the full set of barriers that may prevent someone from voting. Thus, these measures are most useful as a beginning point by identifying racial disparities that should be addressed, and lack of identification of such disparities in these two measures should not be considered a sign that voting access has been equalized.

Additionally, we introduced multiple algorithmic interventions to assign polling locations to reduce racial disparities in distance and load. These new algorithms could allow elections officials to place polling locations more effectively based on a given list of public location (we use schools and libraries). However, these algorithmic interventions focus only on polling locations and therefore ignore other interventions that may be helpful in reducing voter access disparities, such as arranging rides to the polls or giving workers time off to vote. Universal vote-by-mail could be even more effective at alleviating voting access disparities, as it entirely avoids the distance and load concerns we raise here. Within existing voting rules, the introduced algorithms demonstrate the possibility of mitigating racial disparities in voting access.

REFERENCES

[1] Mohsen Abbasi, Aditya Bhaskara, and Suresh Venkatasubramanian. 2021. Fair clustering via equitable group representations. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 504–514.

[2] Mohsen Abbasi, Sorelle A Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. 2019. Fairness in representation: quantifying stereotyping as a representational harm. In *Proceedings of the 2019 SIAM International Conference on Data Mining*. SIAM, 801–809.

[3] Savitha Sam Abraham, Sowmya S Sundaram, et al. 2019. Fairness in Clustering with Multiple Sensitive Attributes. *arXiv preprint arXiv:1910.05113* (2019).

[4] Stacey Abrams, Carol Anderson, Kevin M Kruse, Heather Cox Richardson, and Heather Ann Thompson. 2020. *Voter Suppression in US Elections*. University of Georgia Press.

[5] Pankaj K Agarwal, Sariel Har-Peled, Kasturi R Varadarajan, et al. 2005. Geometric approximation via coresets. *Combinatorial and computational geometry* 52, 1-30 (2005), 3.

[6] Sara Ahmadian, Alessandro Epasto, Ravi Kumar, and Mohammad Mahdian. 2019. Clustering without over-representation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 267–275.

[7] Gaurav Bagwe, Juan Margitic, and Allison Stashko. 2022. *Polling Place Location and the Costs of Voting*. Technical Report. https://jmargitic.github.io/JM/Margitic_JMP.pdf

[8] MohammadHossein Bateni, Aditya Bhaskara, Silvio Lattanzi, and Vahab S Mirrokni. 2014. Distributed Balanced Clustering via Mapping Coresets.. In *NIPS*. 2591–2599.

[9] Henry E Brady and John E McNulty. 2011. Turning out to vote: The costs of finding and getting to the polling place. *American Political Science Review* 105, 1 (2011), 115–134.

[10] Enrico Cantoni. 2020. A precinct too far: Turnout and voting costs. *American Economic Journal: Applied Economics* 12, 1 (2020), 61–85.

[11] M Keith Chen, Kareem Haggag, Devin G Pope, and Ryne Rohla. 2022. Racial disparities in voting wait times: evidence from smartphone data. *Review of Economics and Statistics* 104, 6 (2022), 1341–1350.

[12] Xingyu Chen, Brandon Fain, Charles Lyu, and Kamesh Munagala. 2019. Proportionally fair clustering. *arXiv preprint arXiv:1905.03674* (2019).

[13] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. 2017. Fair clustering through fairlets. In *Proc. NIPS*. 5029–5037.

[14] Eden Chlamtáč, Yury Makarychev, and Ali Vakilian. [n. d.]. *Approximating Fair Clustering with Cascaded Norm Objectives*. 2664–2683. https://doi.org/10.1137/1.9781611977073.104 arXiv:https://epubs.siam.org/doi/pdf/10.1137/1.9781611977073.104

[15] Joshua D Clinton, Nick Eubank, Adriane Fresh, and Michael E Shepherd. 2021. Polling place changes and political participation: evidence from North Carolina presidential elections, 2008–2016. *Political Science Research and Methods* 9, 4 (2021), 800–817.

[16] Angela Cools et al. 2022. Parents, Infants, and Voter Turnout: Evidence from the United States. *Quarterly Journal of Political Science* 17, 1 (2022), 91–119.

[17] Tammy Drezner and Zvi Drezner. 2007. Equity models in planar location. *Computational Management Science* 4, 1 (2007), 1–16.

[18] Tammy Drezner and Zvi Drezner. 2011. A note on equity across groups in facility location. *Naval Research Logistics (NRL)* 58, 7 (2011), 705–711.

[19] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*. 214–226.

[20] Joshua J Dyck and James G Gimpel. 2005. Distance, turnout, and the convenience of voting. *Social Science Quarterly* 86, 3 (2005), 531–548.

[21] Dan Feldman and Michael Langberg. 2011. A unified framework for approximating and clustering data. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*. 569–578.

[22] Michael Feldman, Sorelle A Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. 2015. Certifying and removing disparate impact. In *proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 259–268.

[23] Roland G Fryer Jr and Richard Holden. 2011. Measuring the compactness of political districting plans. *The Journal of Law and Economics* 54, 3 (2011), 493–535.

[24] Alan S Gerber, Gregory A Huber, and Seth J Hill. 2013. Identifying the effect of all-mail elections on turnout: Staggered reform in the evergreen state. *Political Science research and methods* 1, 1 (2013), 91–116.

[25] James G Gimpel and Jason E Schuknecht. 2003. Political participation and the accessibility of the ballot box. *Political Geography* 22, 5 (2003), 471–488.

[26] Teofilo F Gonzalez. 1985. Clustering to minimize the maximum intercluster distance. *Theoretical computer science* 38 (1985), 293–306.

[27] Moritz Hardt, Eric Price, Nati Srebro, et al. 2016. Equality of opportunity in supervised learning. In *Proc. NIPS*. 3315–3323.

[28] Moshe Haspel and H Gibbs Knotts. 2005. Location, location, location: Precinct placement and the costs of voting. *The Journal of Politics* 67, 2 (2005), 560–573.

[29] Piotr Indyk. 1999. Sublinear time algorithms for metric space problems. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*. 428–434.

[30] Nick Joslyn, Andrew Bilbo, Jack Arndt, Heidi Berger, and Mark Joslyn. 2020. Distance traveled to polling locations: Are travel costs imposed equally on party members? *The Social Science Journal* 57, 1 (2020), 14–25.

[31] Samir Khuller and Yoram J Sussmann. 2000. The capacitated k-center problem. *SIAM Journal on Discrete Mathematics* 13, 3 (2000), 403–418.

[32] Bridgett A King. 2020. Waiting to vote: The effect of administrative irregularities at polling locations and voter confidence. *Policy Studies* 41, 2-3 (2020), 230–248.

[33] Matthäus Kleindessner, Pranjal Awasthi, and Jamie Morgenstern. 2019. Fair k-center clustering for data summarization. *arXiv preprint arXiv:1901.08628* (2019).

[34] Matthäus Kleindessner, Samira Samadi, Pranjal Awasthi, and Jamie Morgenstern. 2019. Guarantees for spectral clustering with fairness constraints. *arXiv preprint arXiv:1901.08668* (2019).

[35] Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. 2017. Counterfactual fairness. In *Advances in neural information processing systems*. 4066–4076.

[36] Harry A Levin and Sorelle A Friedler. 2019. Automated congressional redistricting. *Journal of Experimental Algorithmics (JEA)* 24 (2019), 1–24.

[37] Yury Makarychev and Ali Vakilian. 2021. Approximation Algorithms for Socially Fair Clustering. *arXiv preprint arXiv:2103.02512* (2021).

[38] Michael T Marsh and David A Schilling. 1994. Equity measurement in facility location analysis: A review and framework. *European journal of operational research* 74, 1 (1994), 1–17.

[39] Anuj Mehrotra, Ellis L Johnson, and George L Nemhauser. 1998. An optimization based heuristic for political districting. *Management Science* 44, 8 (1998), 1100–1114.

[40] Kian Mirjalali and Hamid Zarrabi-Zadeh. 2020. Improved Algorithms for Distributed Balanced Clustering. In *International Conference on Topics in Theoretical Computer Science*. Springer, 72–84.

[41] Maryam Negahbani and Deeparnab Chakrabarty. 2021. Better Algorithms for Individually Fair k-Clustering. *Advances in Neural Information Processing Systems* 34 (2021).

[42] Stephen Pettigrew. 2021. The downstream consequences of long waits: How lines at the precinct depress future turnout. *Electoral studies* 71 (2021), 102188.

[43] Clemens Rösner and Melanie Schmidt. 2018. Privacy preserving clustering with constraints. *arXiv preprint arXiv:1802.02497* (2018).

[44] Melanie Schmidt, Chris Schwiegelshohn, and Christian Sohler. 2018. Fair coresets and streaming algorithms for fair k-means clustering. *arXiv preprint arXiv:1812.10854* (2018).

[45] Michael E Shepherd, Adriane Fresh, Nick Eubank, and Joshua D Clinton. 2021. The Politics of Locating Polling Places: Race and Partisanship in North Carolina Election Administration, 2008–2016. *Election Law Journal: Rules, Politics, and Policy* 20, 2 (2021), 155–177.

[46] Ana-Andreea Stoica, Abhijnan Chakraborty, Palash Dey, and Krishna P Gummadi. 2019. Minimizing margin of victory for fair political and educational districting. *arXiv preprint arXiv:1909.05583* (2019).

[47] Sabina Tomkins, Keniel Yao, Johann Gaebler, Tobias Konitzer, David Rothschild, Marc Meredith, and Sharad Goel. 2022. Blocks as Geographic Discontinuities: The Effect of Polling-Place Assignment on Voting. *Political Analysis* (2022), 1–16.

[48] Ali Vakilian and Mustafa Yalçıner. 2021. Improved approximation algorithms for individually fair clustering. *arXiv preprint arXiv:2106.14043* (2021).

[49] Jesse Yoder. 2018. How polling place changes reduce turnout: Evidence from administrative data in North Carolina. *Available at SSRN 3178184* (2018).

[50] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rogriguez, and Krishna P Gummadi. 2017. Fairness constraints: Mechanisms for fair classification. In *Artificial Intelligence and Statistics*. PMLR, 962–970.