

MTV/Paramount Capstone Final Report
Master of Interdisciplinary Data Science
Duke University

Dapo Adegbile, Jasmine Young, Pranav Manjunath

April 3rd, 2022

Table of Contents

1. Introduction

- a. Key Questions

2. Data & Methodologies

- a. Overview of Data Needs
 - i. College Campus Boundary Data
 - ii. College Demographic Data
 - iii. Polling Place Data
- b. Data Methodologies
 - i. Merging Campus Boundary & College Demographic Data
 - ii. Polling Place & College Campus Distance Calculation

3. Results & Analysis

- a. How many and which schools across the country have voting sites on their college campuses?
- b. How has this changed, if at all, from 2012 - 2020?
- c. How does the presence of a polling place on or near campus differ depending on the type of school and student demographics?
- d. For those schools that do not have on-campus voting options, what is the average travel distance to their nearest voting option by foot, public transportation, or car?

4. Conclusions & Limitations

- a. Conclusions
- b. Limitations & Future Directions

Introduction

Youth voted at a record rate in 2018, but of those who didn't turnout in the last election, nearly half cited access as a reason for not voting. According to a 2019 report by The Leadership Conference, more than 1,600 polling places closed since the 2013 *Shelby v. Holder* decision which removed all jurisdictions from Voting Rights Act "pre-clearance" coverage. A large number of those 1,600 closures took place in communities of color, there has been no comprehensive study and no targeted reporting on how the ruling affected students on college campuses as a whole. Through this report, we hope to understand the level of access students at different types of institutions have when it comes to voting.

Key Questions

The questions we focused on this year are as follows:

1. How many and which schools across the country have election-day voting sites on their college campuses? What types of voting sites are they i.e. early voting, same-day voting, dropboxes etc.?
2. How has this changed, if at all, over the last three presidential election cycles (2012, 2016, 2020)? While it is unlikely we can draw any causality, how has this changed since *Shelby v. Holder*?
3. How does the presence of a polling place on campus / the distance to the nearest voting location differ depending on the type of school and student demographics?
4. For those schools that do not have on-campus election-day voting options, what is the average travel distance to their nearest early voting booth by foot, public transportation, or car?

Data & Methodologies

Overview of Data Needs

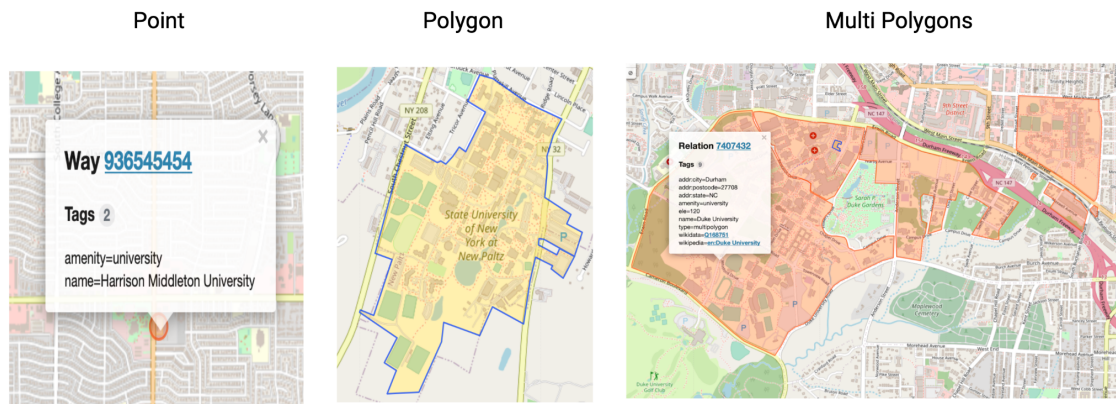
To help answer our project objectives, we require polling place (Election Day and Early Voting) location, college demographics and college campus boundary data. It is necessary to have the latitude and longitude values for the polling locations and the boundaries of college campuses represented by polygons/multipolygon data types. In terms of college demographics, we needed to understand the MSI Type, racial student demographics, and if it is a 2 Yr. or 4 Yr. Public/Private University.

Geocoding is the process of transforming a description of a location, such as an address or a name of a place, to a location on the earth's surface. For our project, we geocoded the polling places addresses to obtain their respective latitude and longitude values.

College Campus Boundary Data

For this project, we needed to determine the location data for college campuses in the US. College campuses could be represented by a point, polygon or a multipolygon. A point

representation would be a single longitude/latitude value present anywhere on the campus and a polygon representation depicts the boundaries of the college campus. When a college is spread out in many places, multi-polygons are used to connect multiple polygons as one university. The illustration below depicts the differentiation between a point, polygon, and a multi-polygon.



The main sources of College Campuses were from:

- OverPass Turbo. A web-based data filtering tool for OpenStreetMap.
- Homeland Infrastructure Foundation Level Data (HIFLD ARCGIS). National foundation-level geospatial data within the open public domain that can be useful to support community preparedness, resiliency, research, and more. ARCGIS is a geographical information system (GIS) software that allows handling and analyzing geographic information by visualizing geographical statistics through layer building maps like climate data or trade flows.

For our analysis, college polygons/multi-polygons are a more accurate representation of the college. We initially used the Overpass Turbo Data to extract college campus information but noticed that many colleges were in points format and the polygons were not extremely accurate. Hence we completely migrated to using the college campus data obtained from HIFLD.

The HIFLD data is stored in a shape file and consists of the college name and its respective college polygon coordinates. There was no cleaning required in terms of modifying column data types or with the format of the data. This data consisted of 5396 colleges. With the HIFLD ArcGis data, there were few cases where specific departments, labs, apartments were specified as individual college names. For example, along with Duke University, Fuqua School of Business, Sanford School of Public Policy, University Apartments were also included as separate colleges. In order to maintain uniformity, we removed entries from the dataset through string pattern matching techniques.

College Demographics Data

The college demographics dataset was given to us by the +1 Team at MTV. This data was collected as a part of the “Students Learn Students Vote Coalition”, and used as a tracker to coordinate their efforts in enrolling campuses in NSLVE. The data consists of columns including College Name, City and State, Institution Type, MSI Type, Student Racial Breakdowns, and other demographic information.

The +1 Team at MTV began with institutions in the National Student Clearinghouse, and added campuses that are designated as minority-serving institutions in the Rutgers Center for MSI Directory. The list was cross-referenced with all the campuses in the National Center for Education Statistics and IPEDS. They cut out for-profit schools, specialized medical schools, and beauty schools. We are confident that this dataset reasonably represents the college campuses in the United States.

This data was given to us in a structured csv format, hence we did not have to clean or modify the format of the data. This data consisted of 3114 colleges. Similar to the HIFLD, there were rows that contained names other than the college’s name. Hence, in order to maintain uniformity, we removed rows with non-college names values from the dataset through string pattern matching techniques.

Polling Place Data

Our polling place data comes from a few different sources. The 2020 polling data used in this analysis comes from the Center for Public Integrity and covers about 35 states. The 2020 Early Voting data used in our later analysis comes from Ballot Ready and covers around 45 states. We explored additional data options for 2020 Election Day that included Safegraph, a less reliable web scraping tool. Our 2018 Early Voting data also comes from Ballot Ready. The data we have for 2012, 2016, and 2018 comes from the Center for Public Integrity and also covers fewer than 35 states.

Some of our statistics and results rely on different samples of states. When looking at any election year individually, we used the maximum available data in that election year. We also looked at Early Voting and Dropbox voting separately at times. When we sought to compare changes over time, we used the combined dataset for 2012, 2016, 2018, and 2020. The combined dataset only contains colleges in states with polling data present for all years. The specifics of our samples and datasets are detailed below.

Election Covered	States Included	Number of Colleges	Source
2012 Election Day	27 States	1,189	Center for Public Integrity

2016 Election Day	26 States	1,148	Center for Public Integrity
2018 Election Day	30 States	1,310	Center for Public Integrity
2018 Dropbox Voting	6 States	198 Colleges	Ballot Ready
2018 Election Day & Dropbox Voting Combined	6 States	198 Colleges	Center for Public Integrity & Ballot Ready Combined
2020 Election Day	35 States	1,816	Center for Public Integrity
2020 Early Voting	35 States	1,896	Ballot Ready
2020 Dropbox Voting	48 States	2,241	Ballot Ready
2020 Election Day, Early Voting, & Dropbox Combined	32 States	1,727	Center for Public Integrity & Ballot Ready Combined
2012, 2016, 2018, 2020 Election Day Combined	18 States	902	Center for Public Integrity

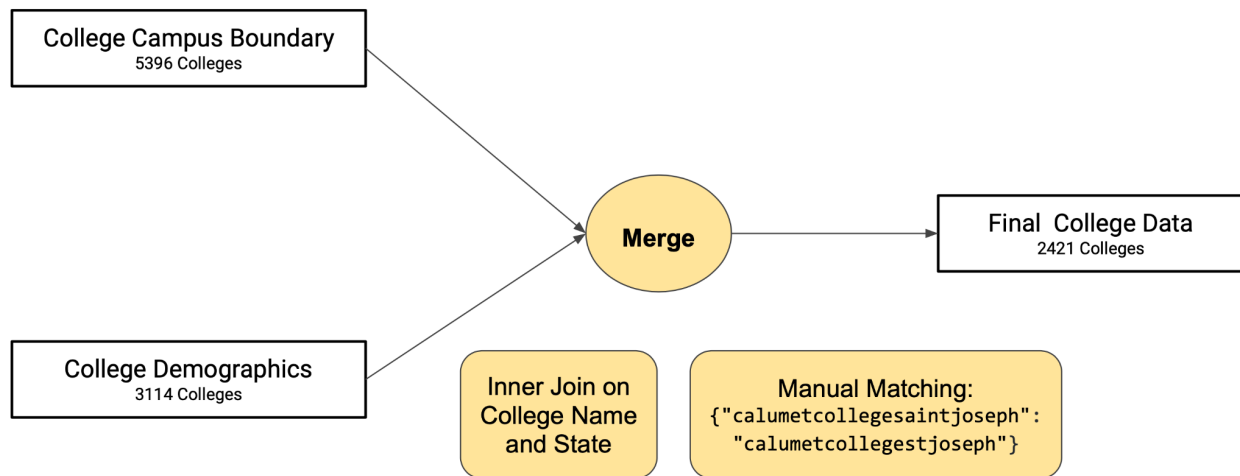
Data Methodologies

Merging College Campus Boundary Data & College Demographic Data

We performed a two step merge to help facilitate the merging of college campus boundary data with the college demographics data (ground truth data). The first step was conducting an inner join between the college polygon and college demographics by the college name and state. While this join was majorly successful, it was not able to include all colleges in the ground truth data. This was because there were several instances where the same college was spelled differently in both datasets and hence did not go through the inner join.

The second step was manually going through the colleges that were not included in the merge and checking if the college exists in both tables and including them as well in the final merged data. The final merged data resulted in a table of 2,421 colleges where each college consisted of its college campus boundary data as well as its college demographics. The colleges that are still missing from the ground truth table are those that are not present in the college campus boundaries dataset.

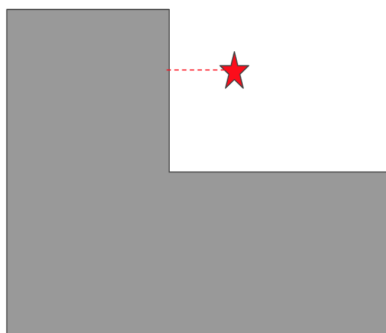
The below diagram visually represents the merging of the two datasets. It shows the two step matching process and also provides an example explaining why manual matching was required. Calumet College Saint Joseph was spelled as Calumet College St. Joseph and hence did not get matched in the first inner join. Manual matching helped us realize that these universities are the same and include them in the merge.



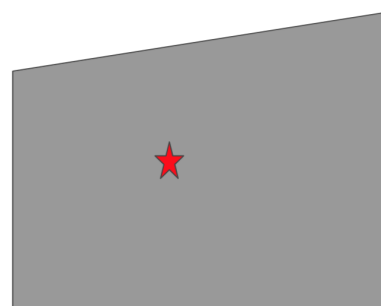
Polling Place and College Campus Distance Calculation

We used a point-to-polygon calculation in geopandas to measure the distance between campuses and their nearest polling place. If a polling place was within a campus' polygon boundary, it was labeled on campus and given a distance of zero. If a polling place was outside of a campus' polygon boundary, it was considered off-campus and given the distance to the nearest edge.

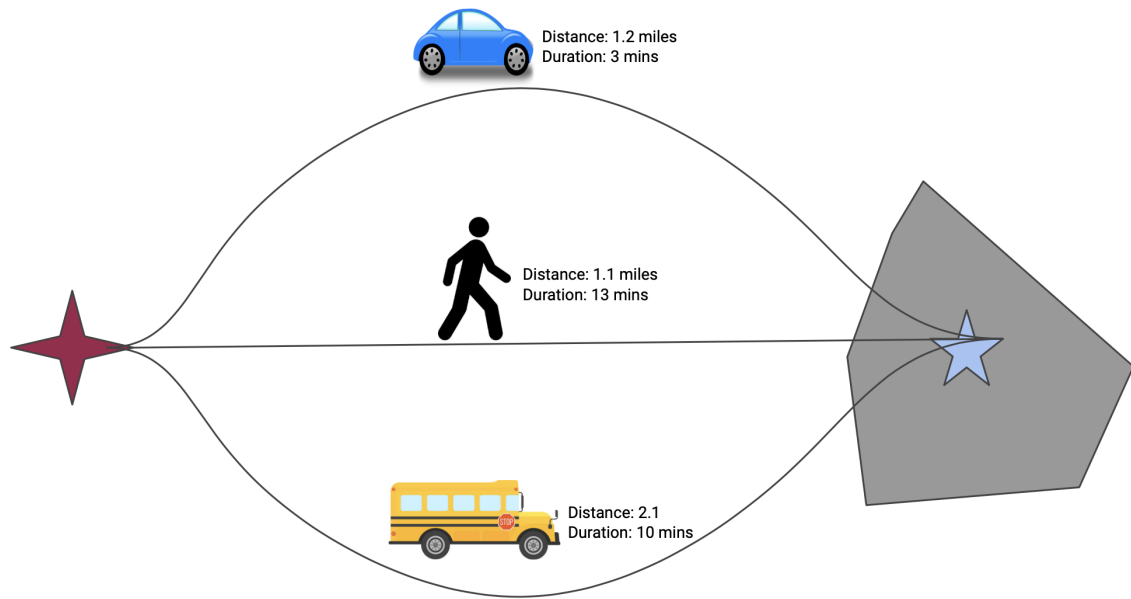
Off-Campus Polling Place



On-Campus Polling Place



We used the Google Maps API to calculate the distance (point to point) and duration between the polling place and the college campus centroid through various means of transportation, such as driving, walking, and public transit. This information is used to better understand the travel times, and by proxy access levels, students have by travel modalities from the centroid of the polygon to the polling place. Even if the polling place is located on campus, the Google API will still calculate the distances and durations.



Results & Analysis

In this section we will present our findings and analysis for each key question. In addition to the results contained in this report, we have provided our client a list of 17 focus colleges. For each of the 17 colleges we looked at their nearest polling places, travel routes, student housing, and any other unique factors that describe a student's journey to vote.

How many and which schools across the country have voting sites on their college campuses?

In 2020, we found that 76% of colleges did not have an election day polling place on campus and 90% of colleges did not have an early voting polling place on campus. Additionally, in 2020 7% of colleges had a voting dropbox on campus. Overall, 73% of colleges did not have a voting option (election day, early voting, or dropbox) on campus for the 2020 election and the nearest

voting option was more than 1 mile from the edge of campus for 14.0 % of colleges. Looking more closely at 2-Year and 4-Year colleges - the nearest election day voting option is more than 1 mile from the edge of campus for 20.5% of 2-year colleges. The nearest election day voting option is more than 1 mile from the edge of campus for 8.6% of 4-year colleges and universities. We see 2-Year colleges are typically farther from their nearest polling place than 4-Year colleges.

In 2018, we found that 84.0 % of colleges do not have an election day polling place on campus. The nearest election day voting option was more than 1 mile from the edge of campus for 23.0 % of colleges in 2018.

In 2016, we found that 85.0 % of colleges do not have an election day polling place on campus. The nearest election day voting option was more than 1 mile from the edge of campus for 23.0 % of colleges in 2016.

In 2012, we found that 81.0 % of colleges do not have an election day polling place on campus. The nearest election day voting option was more than 1 mile from the edge of campus for 14.0 % of colleges in 2012.

How has this changed, if at all, from 2012 - 2020?

We were able to compare data for 902 colleges across 18 states in 2012, 2016, 2018, and 2020. Figure 1 shows the percentage of colleges that had an election day polling place on campus in each year. We see the percentage of colleges with on-campus election day polling drops in 2016 and 2018 before rising again in 2020. In Figure 2 we see that the proportion of colleges that had an election day polling place within 1 mile of campus dropped in 2016 and 2018. Consequently, the percentage of colleges that were 1-3 miles and over 3 miles from their nearest election day polling increased in 2016 and 2018. We take this to mean that college voting access worsened in 2016 and 2018, before improving again in 2020.

Figure 1: Election Day On-Campus Voting By Year

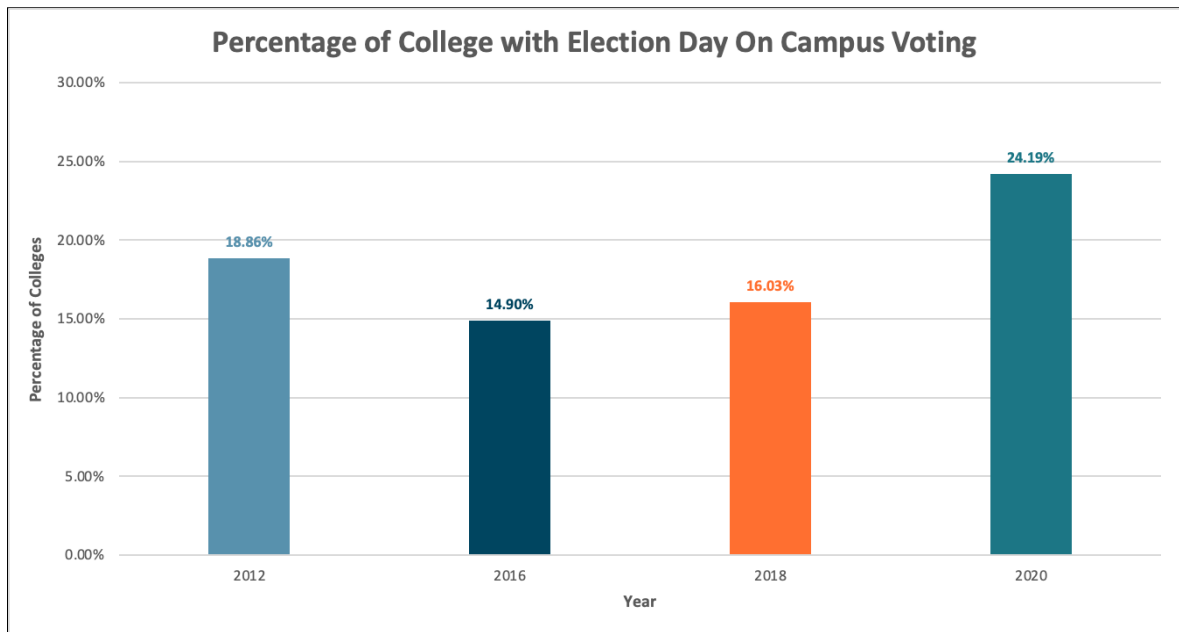
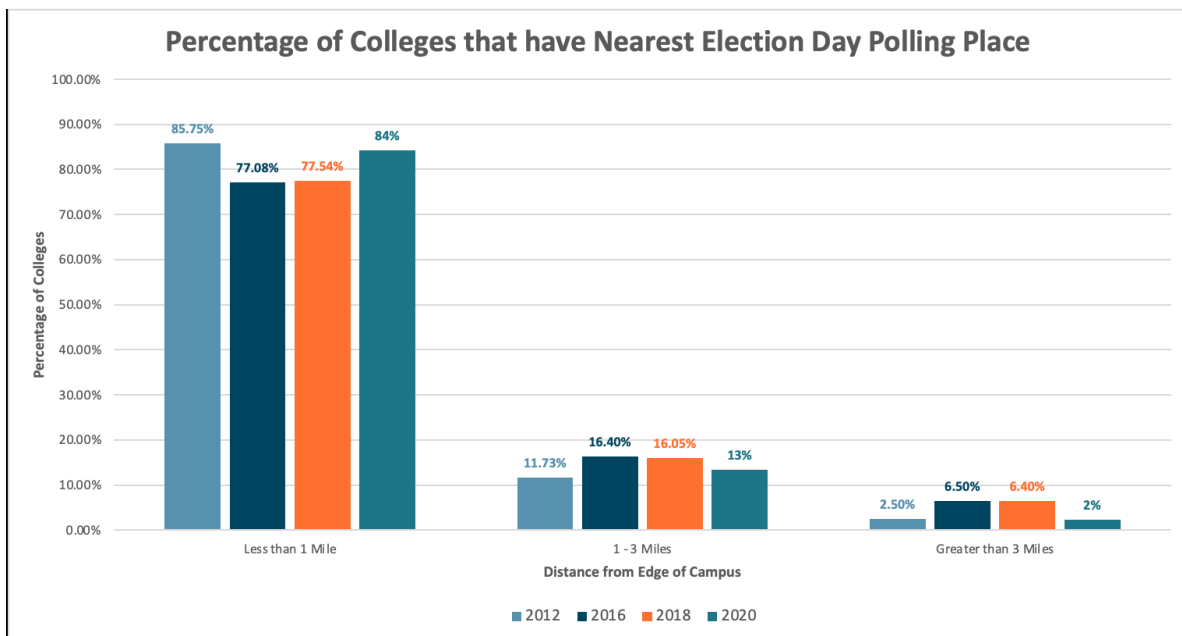


Figure 2: Nearest Election Day Polling Place Distance By Year



After seeing a drop in voting access in 2016 and 2018, we looked more closely at how voting access changed for colleges in each Region of the United States. Figure 3 shows that the percentage of colleges with on-campus election day polling places decreased sharply in the North East and Southern regions in 2016 and 2018 before rising again in 2020. Though the Midwest also experienced a drop in the percentage of colleges, it was not as steep as the drop in the North East and South. The West experienced a drop from 2016 to 2018 and did not improve again in 2020. In Figure 4 we notice something similar, the proportion of schools over 1 mile

from their election day polling place increased more sharply in the North East and Southern regions. Our findings reiterate that voting access worsened in 2016 and 2018, and this drop in access was the most pronounced in the North East and Southern regions.

Figure 3: Election Day On-Campus Voting by Year & Region

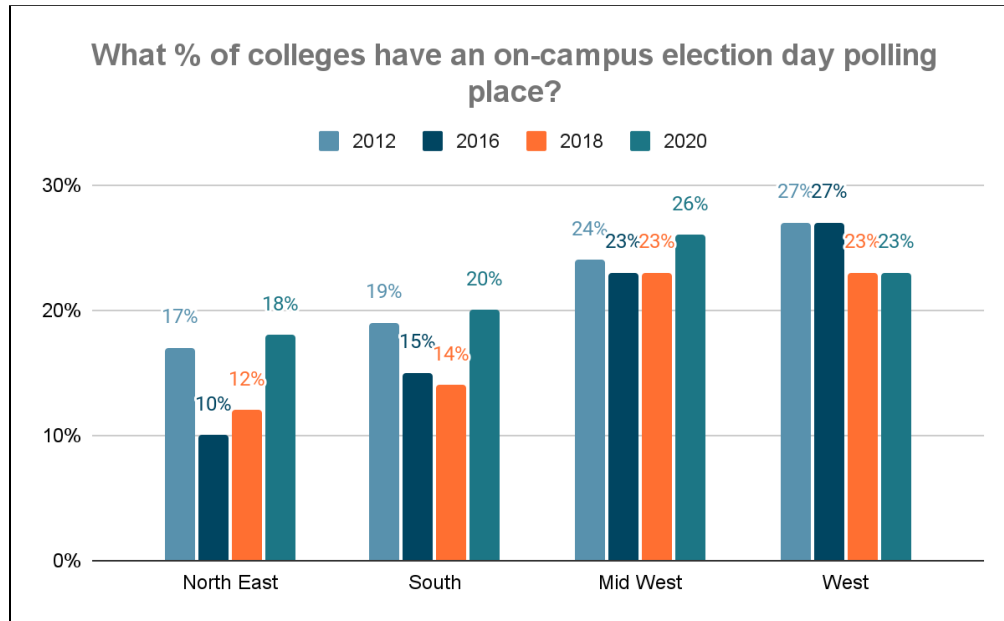
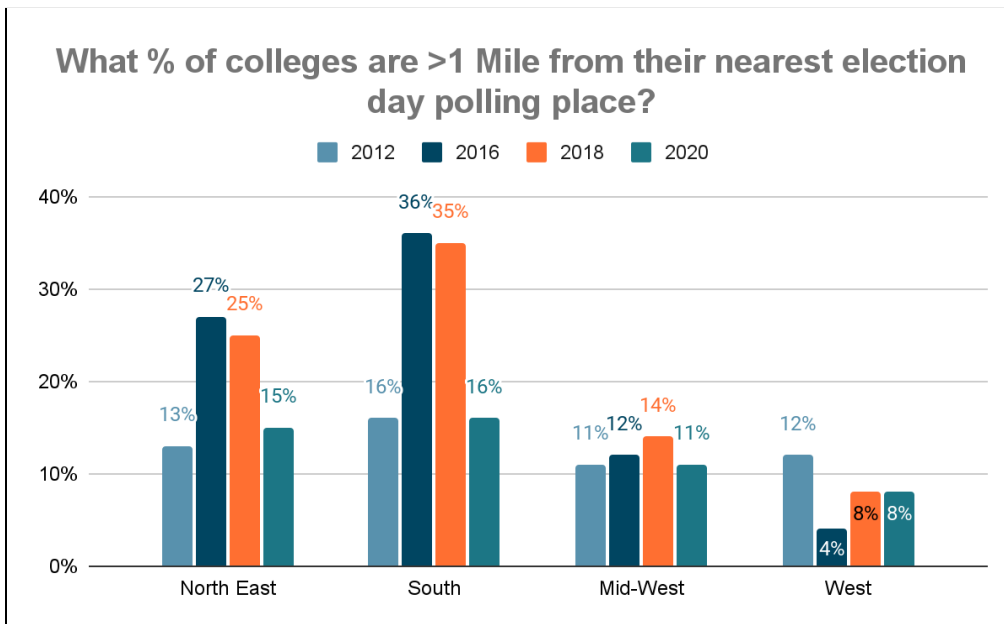


Figure 4: Colleges Over 1 Mile From Their Nearest ED Polling Place by Region & Year

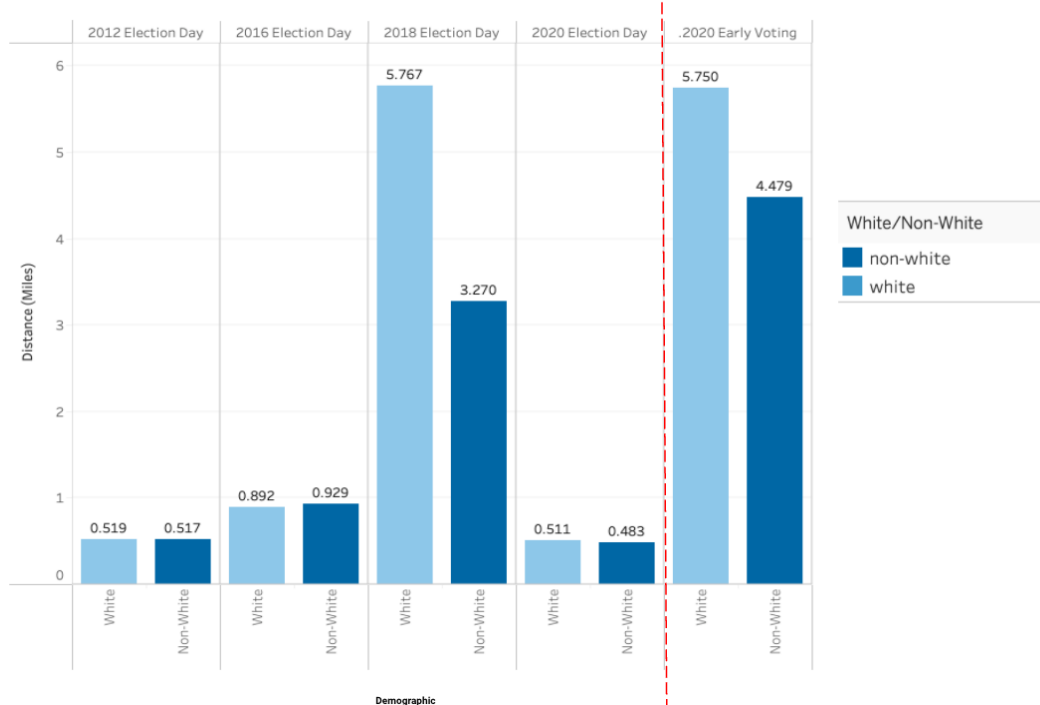


How does the presence of a polling place on or near campus differ depending on the type of school and student demographics?

For this analysis we decided to group schools by the percentage of each racial demographic, instead of their MSI designation. This grouping adds another layer of analysis, giving us a gauge of how many students are being impacted by the polling place locations. More specifically, if greater than 50% of a school population was white, it was labeled as predominantly white. Conversely, if the minority population was greater than 50%, we labeled it as predominantly non white. After doing this we were left with 1025 predominantly white schools and 173 schools that were predominantly non white. A plot of campus polling place distance by student demographics can be seen below.

Figure 5: Campus Polling Place and Student Demographic Breakdown

Campus Polling Places & Student Demographic Breakdown

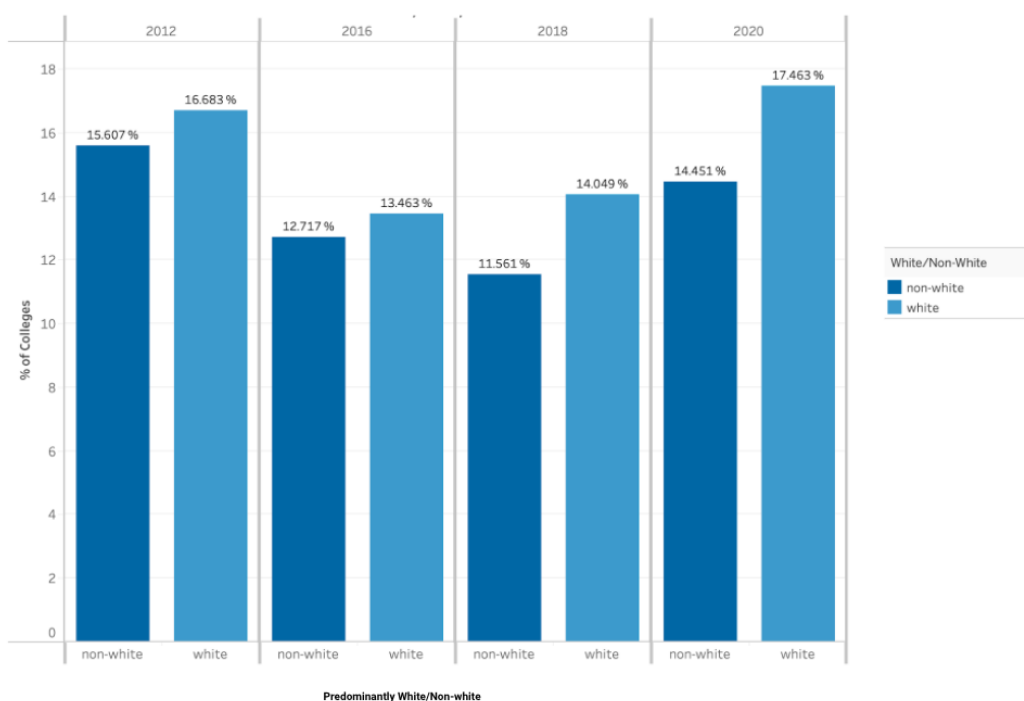


From the plot we can see that in election years (2012, 2016, & 2018) the distance from white and non white schools to their nearest polling place is less than one mile on average. Additionally, in these years, on average white and non white schools are roughly the same distance away from their nearest polling place. In 2018, predominantly white schools are roughly 5.8 miles away from their nearest polling place contrasted with the predominantly non white schools, on average, are roughly 3.2 miles away from their nearest polling place. We believe that some of this can be attributed to the fact that schools in rural or less population dense areas are more likely to be predominantly white.

In an attempt to get a more granular look at our data, we took a look at the percentage of white and non white schools that have a polling place on their campus.

Figure 5: Percentage of White and Non White Schools with Polling Place on Campus

Percentage of White/Non-white Schools with Polling Place on Campus



In each year we can see that compared to non white schools, predominantly white schools have a higher percentage of polling places on their campus.

For those schools that do not have on-campus election day voting options, what is the average travel distance to their nearest voting option by foot, public transportation, or car?

This question has been answered through the distances and durations calculated from the google maps API which calculates the distance and duration between the college centroid and the polling place. The distances are calculated in miles while the durations are calculated in minutes.

Figure 6 shows the average distances between Election Day polling place and the college centroid between 2012 to 2020 by walking (blue), driving (orange) and transit (gray). Interestingly, it can be seen that the distances across the three modes of transportation seem to be similar during the years 2012 and 2020. In all three modes of transportation, the distances seem to be higher during 2016 and 2018 when compared to 2012 and 2020. This indicates that the voting access has been better in 2020 when compared to 2018 and 2016 as the distance from college centroid to polling place reduces by walk, car, and public transportation.

Figure 6: Average Distances from Election Day Polling Place to College Centroid Across Years

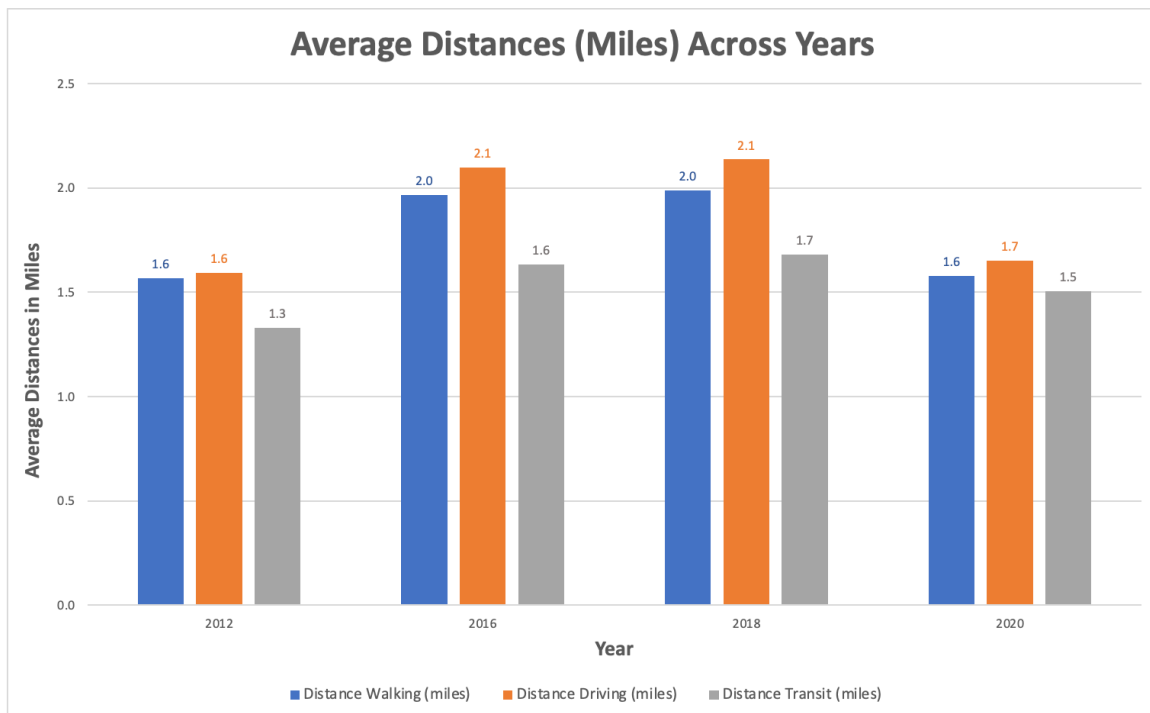


Figure 7 indicates the duration in minutes from the college centroid to the nearest election day polling place by walk, car and public transportation. It can be seen that the trend is similar to Figure 6. It can be seen that the average time it takes to walk in 2020 is 29.8 minutes, similar to the duration in 2012 but lower than the duration in 2018 and 2016. This is supported as since the distances decrease in 2020 when compared to 2018 and 2016, the duration would also decrease. Indicating that voting access improves in 2020.

Figure 7: Average Travel Time from Election Day Polling Place to Campus Centroid Across Years

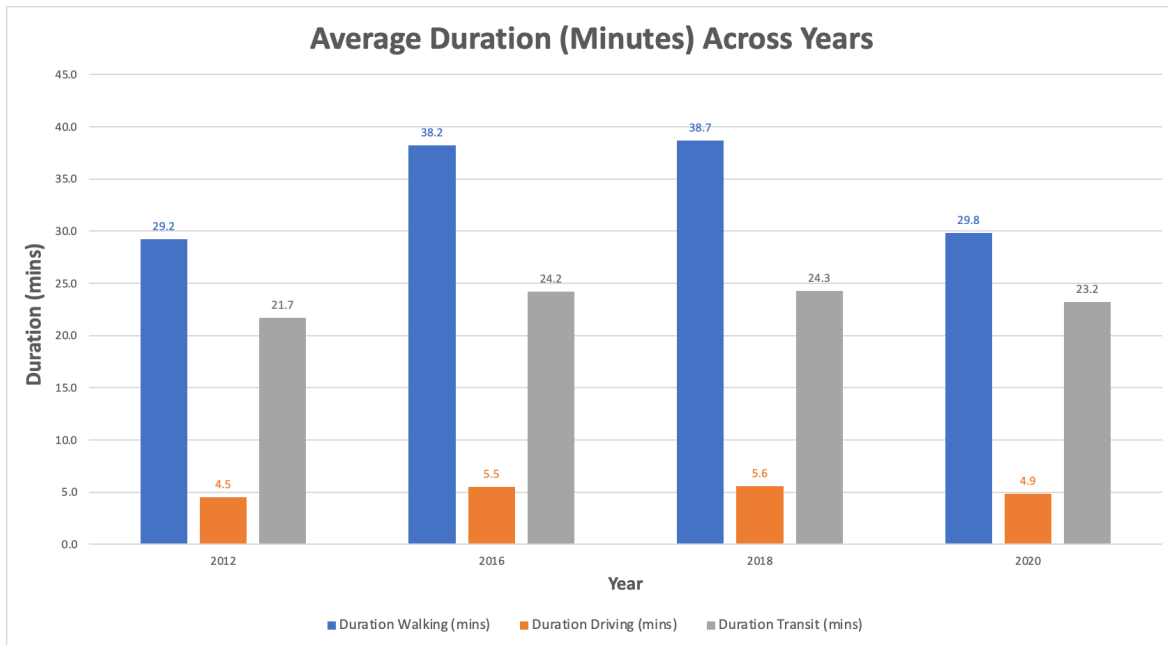
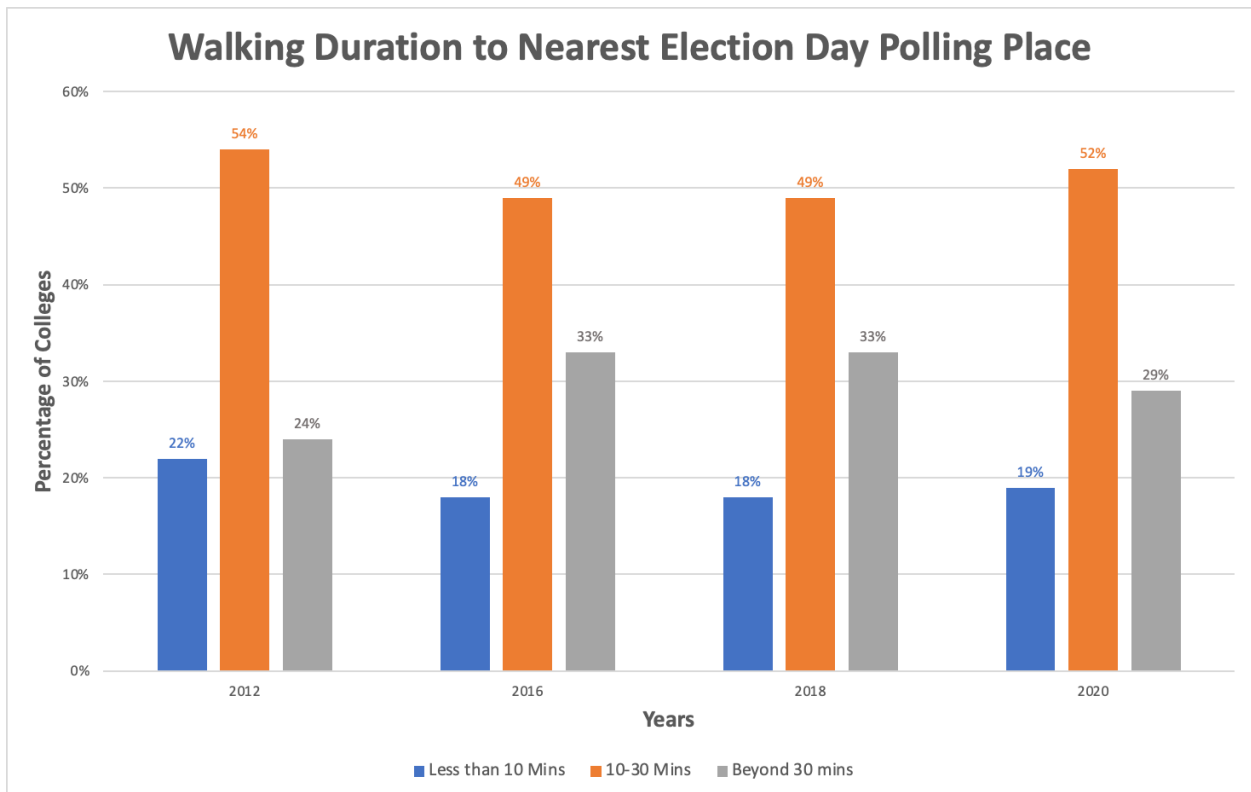


Figure 8: Walking Duration to Nearest Election Day Polling Place from College Centroid



Intuitively, analysis in terms of duration can help depict voting access in colleges, as depicted in Figure 8. It can be seen that from the colleges that do not have any on campus election day voting option, students in 52% of colleges have to walk anywhere between 10-30 mins to reach their nearest election day polling place in 2020. This is higher when compared to 2018 where students in 49% of colleges had to walk anywhere between 10-30 mins. We also can see that in 2020, only 19% of colleges have its nearest election day polling place within a 10 minute walk from college centroid, a slight increase when compared to 2018 and 2016. Interestingly, in 2020, nearly 30% of colleges have its nearest election day polling 30 mins away from college centroid, which is a decrease from 2018 and 2016.

Figure 9: Walking Duration to Nearest Election Day Polling Place from College Centroid

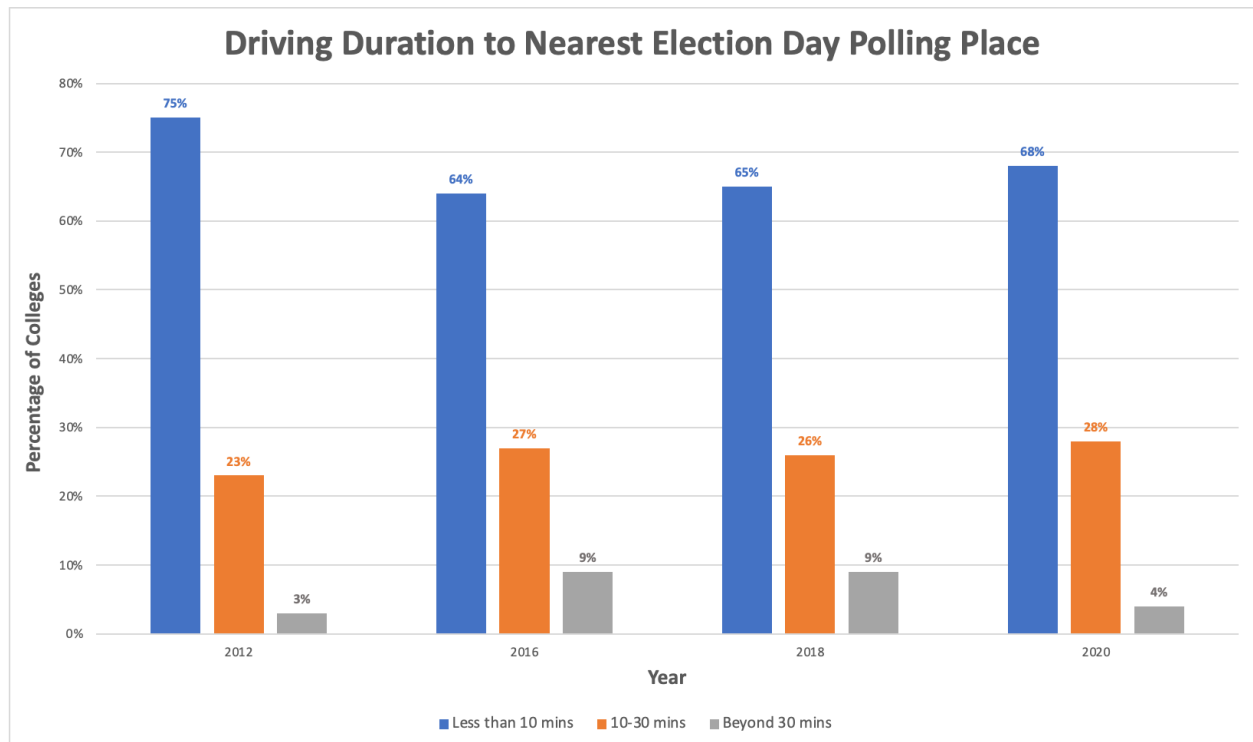


Figure 9 represents the driving duration of colleges without on campus election day voting options between the years of 2012 and 2020. It can be seen that in 2020, 68% colleges have their nearest election day polling place within a 10 minutes drive. Only 4% of colleges have their nearest election day polling place beyond a 30 minute drive, less than 2018 and 2016. 2012 had the best voting access as 75% of colleges had their nearest election day polling place within a 10 minute drive.

Conclusions & Limitations

In this section we will discuss the main conclusions from our work thus far and the limitations of our analysis.

Conclusions

We acknowledge that understanding college access to polling places can be layered and very nuanced. With that in mind, of the data we were able to collect, we currently see that in 2020, 76% of colleges did not have election day polling on campus, and 90% of colleges did not have early voting on campus. In 2018, 2016 and 2012 81% - 85% of colleges did not have polling places on campus. Additionally, 2-Year colleges have larger distances to their nearest polling places than 4-Year colleges.

Through this study, we also noticed voting access decreased from 2012 to 2016 and has increased in 2020. The percentage of on campus election day voting sites increased from 2016 to 2020 indicating that the voting access improved in the most recent election. We also observed that a higher percentage of colleges had their nearest election day polling place between 1-3 miles in 2016 and 2018 when compared to 2012 and 2020, meaning college voting access worsened in 2016 and 2018. When looking at regions - the NorthEast and South have less on campus election day polling places, especially in 2016 and 2018.

When looking at the demographic makeup of schools - we see that white and non-white schools have very similar average distances to their nearest polling place in all election years all under one mile. In terms of the proportion of schools with an on-campus polling place, predominantly white schools have a higher percentage of polling places on their campus compared to non-white schools.

In terms of travel time, the average time it takes for students to walk to their nearest election day polling place in 2020 is about 30 mins by walk, a significant decrease when compared to 2016 and 2018. When looking at the colleges without on campus voting - 68% of colleges in 2020 have their nearest election day polling place less than 10 mins by car, higher than 2016 and 2018. Voting access has improved from the last two elections and hopefully, our research and work can be used to spread awareness and improve voting access in places that require attention.

Limitations & Future Directions

Using distance as a proxy for access was the primary limitation of our analysis. We assumed that colleges closer to polling places had more access than colleges that were farther from polling places. In actuality, understanding voter access for college students is much more complex. Our analysis could be enhanced and built upon if there were a way to truly evaluate the quality of polling places. Data on the amount of people served at a polling place, or the typical wait times a voter could expect would further the study of voter access for college students. There are also access issues at other stages of voting - for example, registering to vote and requesting an absentee ballot are other important voting access points worth studying.

Additionally, we were limited by the states, years, and colleges for which we had data. Our combined dataset from 2012 - 2020 contained 18 states. Gathering polling data for additional states in the years to come could validate our analysis for a larger part of the country. There are still many avenues of voter access to be studied to improve voting for college students, young people, and voters across the United States.