Capstone End of Semester Report - MTV Viacom

Dapo Adegbile, Jasmine Young, Pranav Manjunath

December 8th, 2021

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.

The questions we focused on this semester 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 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?
3. 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

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, proportion of racial students, and if it is a 2 Yr. or 4 Yr. Public/Private university.

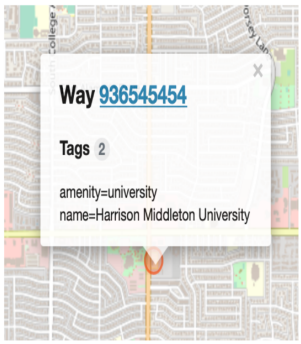
Geocoding is the process of transforming a description of a location, such as an address or a name of a place, to a location on the earth's surface. For our project, we geocoded the polling places addresses to obtain their respective latitude and longitude values.

Campus Campus Boundary Data

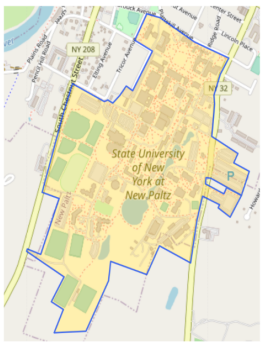
Source

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.

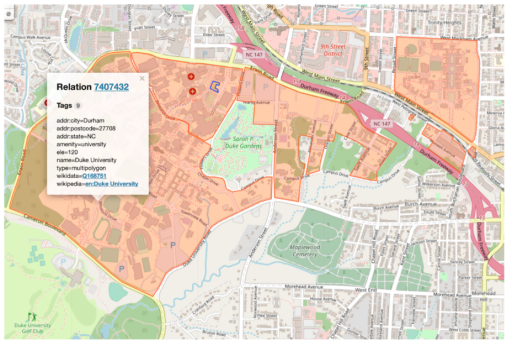
Point



Polygon



Multi Polygons



The main sources of College Campuses were from:

a) OverPass Turbo. A web-based data filtering tool for OpenStreetMap.
b) 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.

Cleaning

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.

Issues/Challenging

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.

Resolutions/Next Steps

Hence, in order to maintain uniformity, we removed entries from the dataset through string pattern matching techniques.

College Demographics Data

Source

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”, 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.

Cleaning

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.

Issues/Challenging

Similar to the HIFLD, there were rows that were not specifically college names.

Resolutions/Next Steps

Hence, in order to maintain uniformity, we removed rows with non-college names values from the dataset through string pattern matching techniques.

Merging College Campus Boundary Data & College Demographic Data

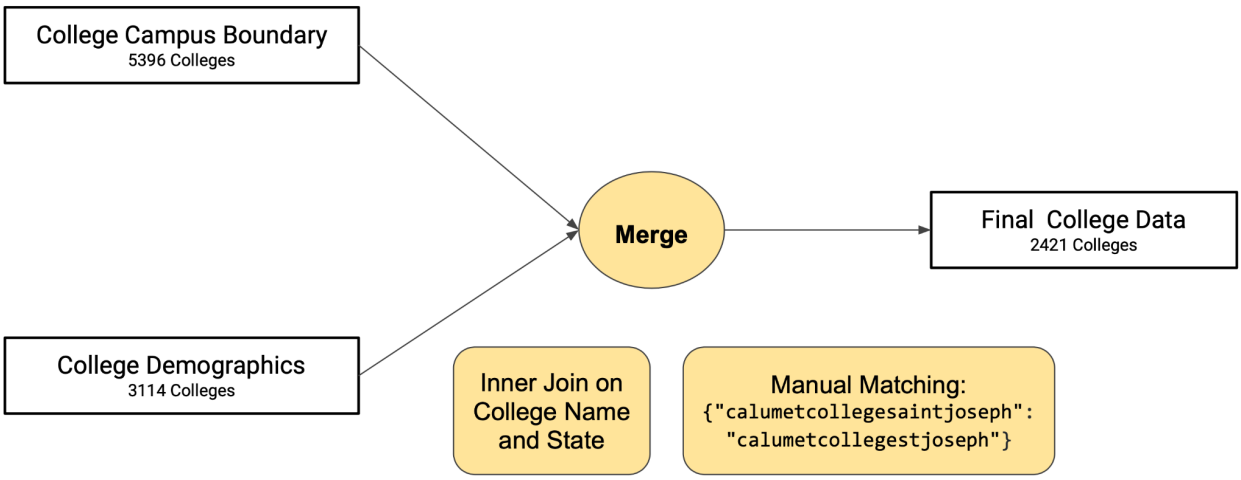
We performed a two step merge to help facilitate the merging of college campus boundary data with the college demographics data (ground truth data). The first step was conducting an inner join between the college polygon and college demographics by the college name and state.

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 2421 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 Data

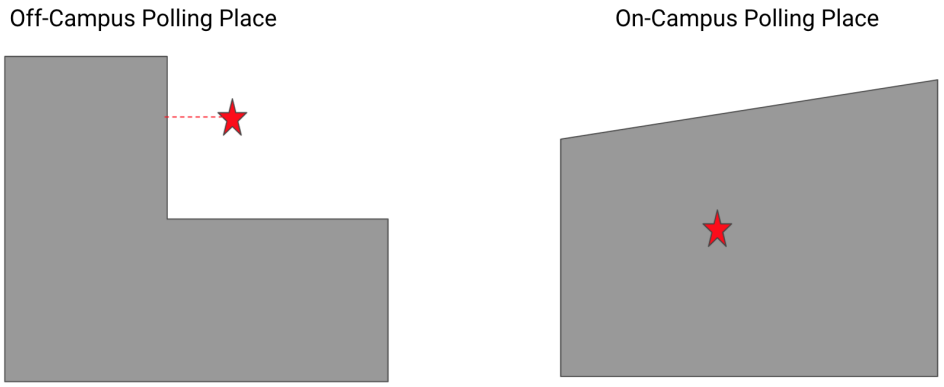
Our polling place data comes from a few different sources. The 2020 polling data used in this analysis comes from the Center for Public Integrity and covers about 35 states. The 2020 Early Voting data used in our later analysis comes from Ballot Ready and covers around 45 states. We explored additional data options for 2020 Election Day that included Safegraph, a less reliable web scraping tool, and we may be receiving election day data from Ballot Ready for 2020 in the spring. 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 about 35 states.

Year	In Person	Early Voting
2020	**CPI (~35 States)** • Safegraph (Less reliable web scraping) • Ballot Ready (Potential for Spring)	**Ballot Ready In-Person & Dropoff (45 States)**
2018	Center for Public Integrity (CPI) Data (~35 States)	Ballot Ready In-Person & Dropoff (~37 States)
2016	Center for Public Integrity (CPI) Data (~35 States)	
2012	Center for Public Integrity (CPI) Data (~35 States)	

Polling Place and College Campus Distance Calculation

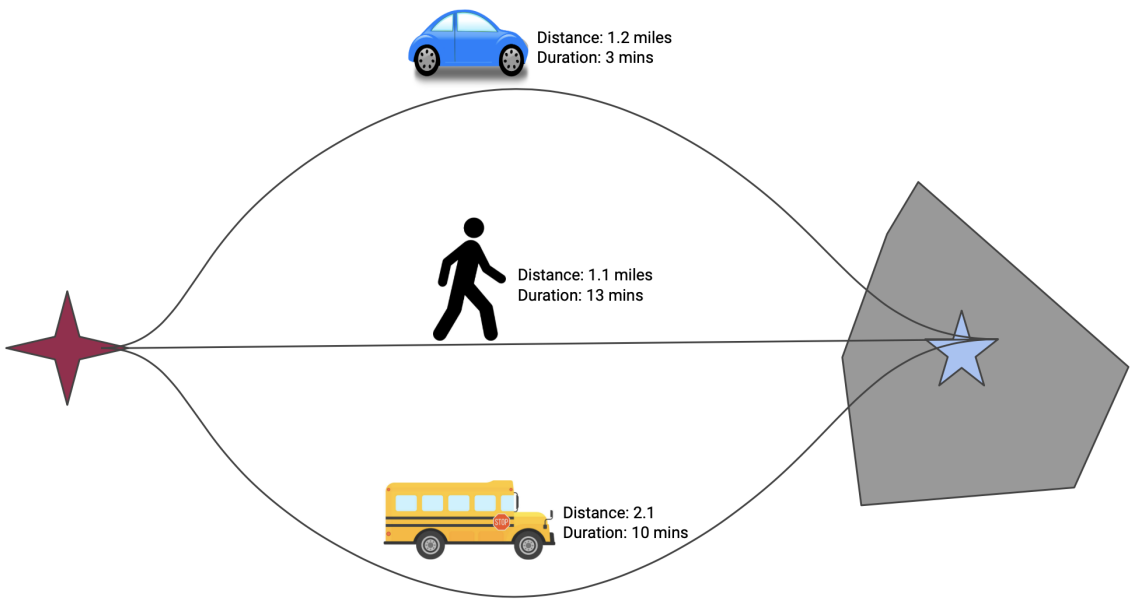
i) Polling Place to Campus Calculation

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.



ii) Google Maps API

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.



Preliminary Results & Analysis

In this section we will discuss some of our preliminary results and initial analysis. These results will focus on 2020 Early Voting and 2020 Election Day Voting for 31 States[1]. The visualizations and tables will accompany targeted analysis questions in the style we intend to mimic moving forward. Additionally, we introduce Regions (South, West, Mid-West, and North-East) and Urban/Rural census designations to our data for analysis purposes.

What percentage of schools have on-campus polling places?

Percentage of Schools With On-Campus Election Day Polling Places	24.26%
Percentage of Schools With On-Campus Early Voting Sites	9.0%
Percentage of Schools With On-Campus Voting Options (Any)	25.89%

About 24% of colleges have on-campus election day polling places, and about 9% of colleges have on-campus early voting sites. This leaves about 75% of colleges without on-campus election day voting options, and 91% of colleges without on-campus early voting options.

[1] *'AL', 'TX', 'ME', 'MA', 'IN', 'AK', 'IA', 'NC', 'CA', 'AZ', 'AR', 'MN', 'MI', 'TN', 'NH', 'WI', 'MT', 'NE', 'WV', 'KY', 'KS', 'CO','IL', 'OK', 'DC', 'FL', 'SD', 'GA', 'ID', 'HI', 'UT', 'NY', 'LA','MD', 'NV', 'NM', 'SC', 'PA', 'OH', 'RI', 'WY', 'VA', 'ND', 'VT', 'WA'

What percentage of schools are more than 1 mile from their nearest voting site?

Percentage of Schools With Nearest Election Day Polling Place More Than 1 Mile Away From Center	41.32%
Percentage of Schools With Nearest Early Voting Site More Than 1 Mile Away From Center	76.32%
Percentage of Schools With Nearest Voting Option (Any) More Than 1 Mile Away From Center	38.83%

About 41% of colleges have their nearest election day polling place more than one mile from the campus' center. About 76% of colleges have their nearest early voting site more than one mile from the campus' center. About 39% of college campus centers are more than one mile from their nearest election day and early voting option.

Percentage of Schools With Nearest Election Day Polling Place More Than 1 Mile Away From Campus	75.62%
Percentage of Schools With Nearest Early Voting Site More Than 1 Mile Away From Campus	91.0%
Percentage of Schools With Nearest Voting Option (Any) More Than 1 Mile Away From Campus	74.0%

About 76% of colleges have their nearest election day polling place more than one mile from campus. About 91% of colleges have their nearest early voting site more than one mile from the campus. About 74% of colleges are more than one mile from their nearest election day and early voting option.

Of the on-campus polling sites, what percentage are on PWIs vs. MSIs?

We have grouped colleges into Predominantly White Institutions (PWI's), Historically Black Colleges & Universities/Predominantly Black Institutions (HBCU/PBI), Hispanic Serving Institutions (HSI), and Other. The Other category holds MSI's with particularly small numbers - Asian American and Native American Pacific Islander Serving Institutions (AANAPISI), Alaska Native and Native Hawaiian Serving Institutions (ANNH), Native-American Serving Nontribal Institutions (NASNTI), and Tribal Colleges & Universities (TCU).

PWI Statistics	
Percentage of Urban Schools With On-Campus Voting Options (Any)	24.66%
Percentage of Rural Schools With On-Campus Voting Options (Any)	15.41%
Percentage of Schools With On-Campus Voting Options (Any)	22.45%

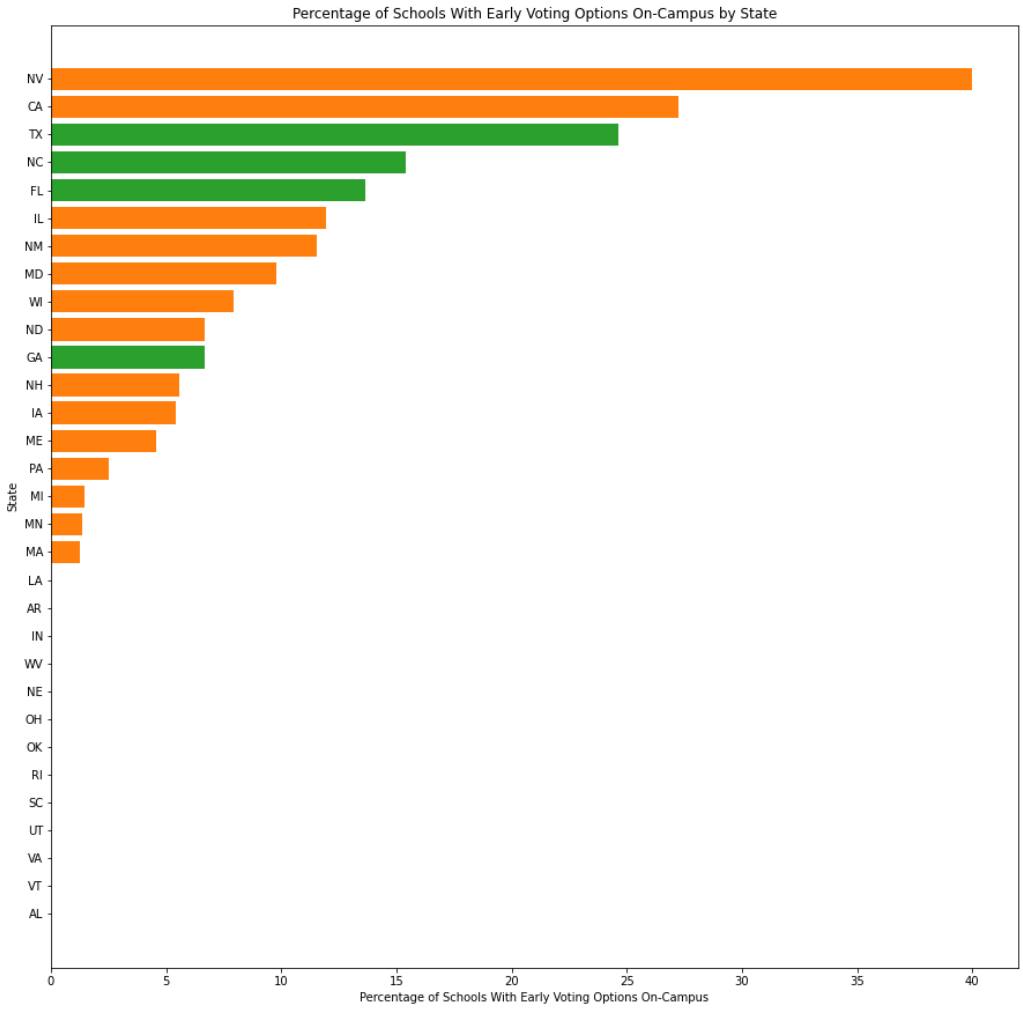
HBCU/PBI Statistics	
Percentage of Urban Schools With On-Campus Voting Options (Any)	30.93%
Percentage of Rural Schools With On-Campus Voting Options (Any)	30.0%
Percentage of Schools With On-Campus Voting Options (Any)	30.71%

HSI Statistics	
Percentage of Urban Schools With On-Campus Voting Options (Any)	39.04%
Percentage of Rural Schools With On-Campus Voting Options (Any)	42.86%
Percentage of Schools With On-Campus Voting Options (Any)	39.18%

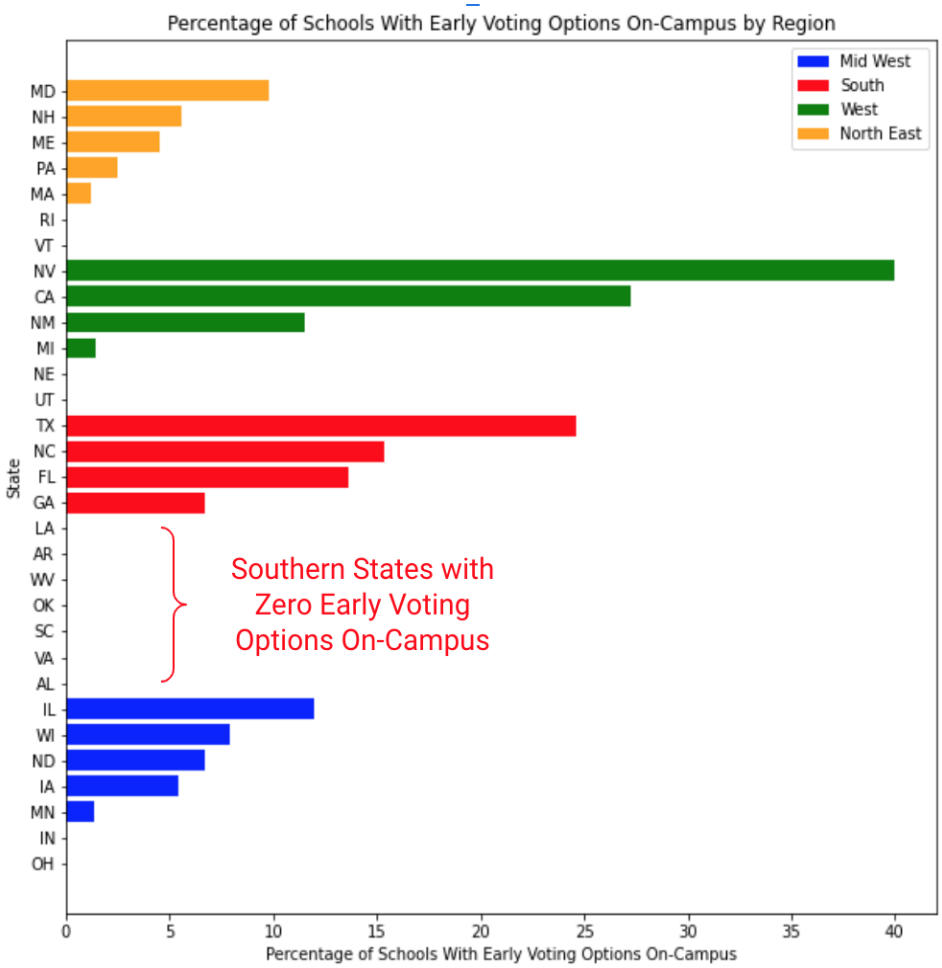
Other Statistics	
Percentage of Urban Schools With On-Campus Voting Options (Any)	39.25%
Percentage of Rural Schools With On-Campus Voting Options (Any)	11.76%
Percentage of Schools With On-Campus Voting Options (Any)	35.48%

For PWI’s about 25% of urban schools have on-campus voting options, and about 15% of rural schools have on-campus voting options. For our HBCU’s, PBI’s, and HSI’s the amount of urban schools with on-campus voting options is very similar to the amount of rural schools with on-campus voting options. Overall, MSI designated institutions seem to have more on-campus voting options than PWI’s.

What percentage of schools in GA, NC, TX, and FL respectively have early voting options on campus? How does this compare to other states in the country?

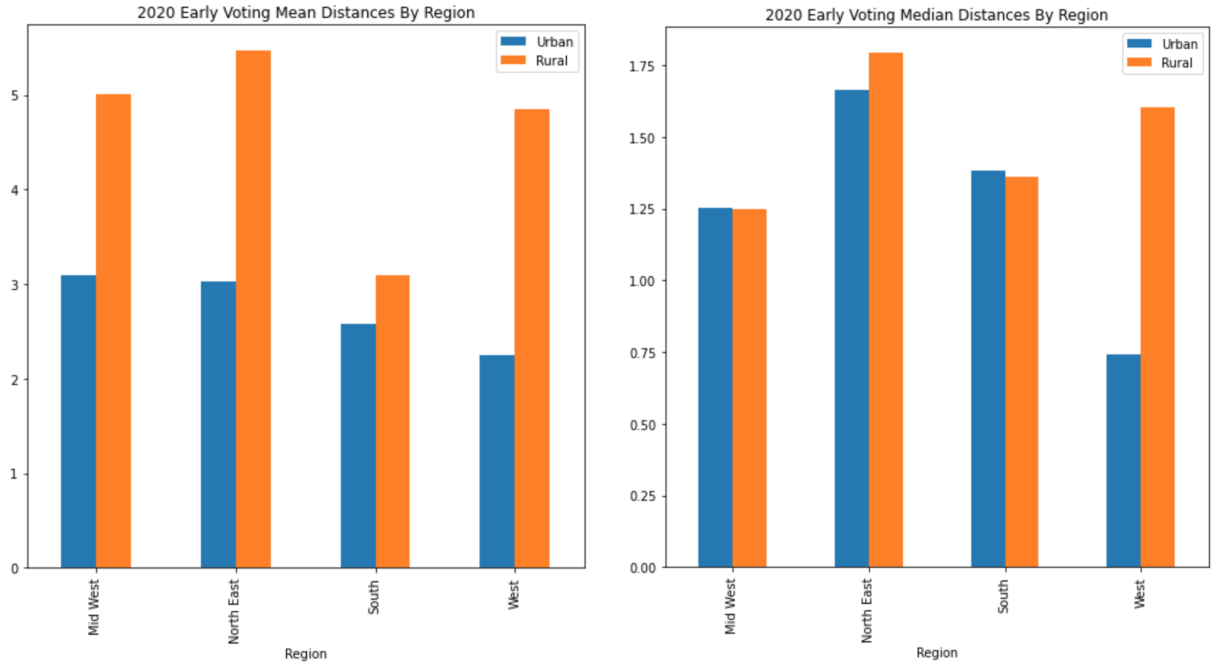


Our client was interested in four particular states as potential focus points for their campaign - North Carolina, Georgia, Texas, and Florida. In Georgia about 7% of colleges have early voting options on campus. In Florida about 14% of colleges have early voting options on campus. In North Carolina about 16% of colleges have early voting options on campus. In Texas about 25% of colleges have early voting options on campus. Texas has the 3rd highest percentage of early voting options on campus in the US, coming after Nevada and California. Though these states are relatively high up compared to others - most of our states sit below 20% of colleges having early voting options on campus.

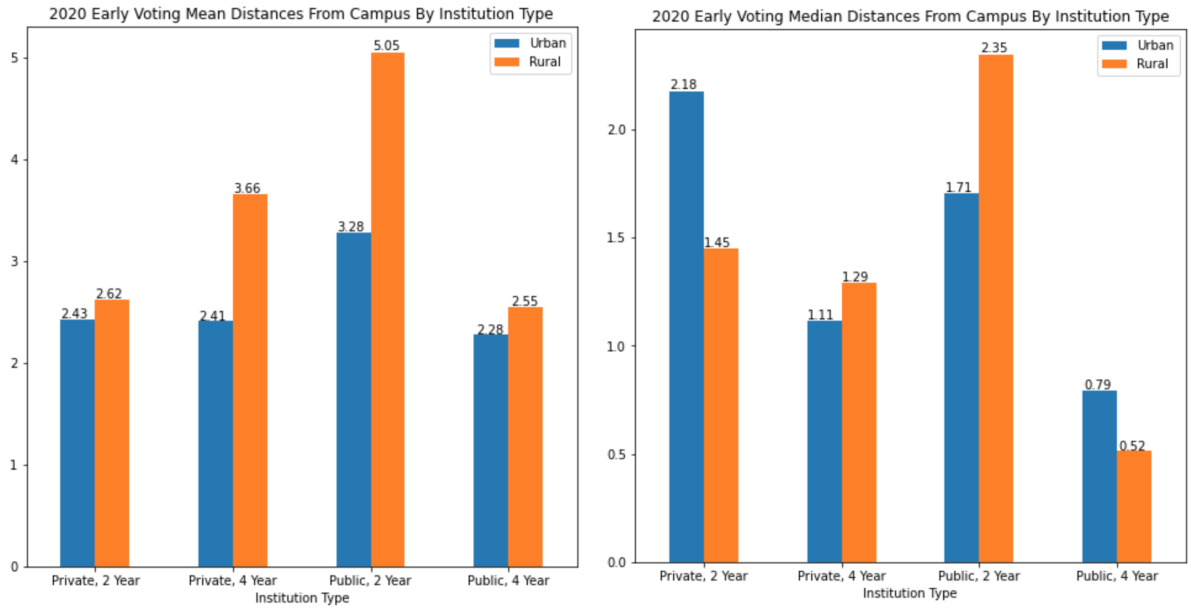


While the West and South have some states with the highest percentage of schools with early voting options on campus, the South has the most states with no early voting options on campus. The South may be a region worth targeting because of the lack of early voting options in many of its states.

How do Urban/Rural Colleges Compare Across Regions & Institution Type?



Most regions have larger distances for their rural colleges than they do for their urban colleges. The North East ranks consistently higher than other regions in terms of distance to polling places, and the West has much larger distances for rural colleges. The South is unique in that both the mean and median distances are similar for urban vs. rural colleges.



When we look at early voting distances by Institution Type we see that Public 4-Year colleges have the shortest mean and median distances. This means that there may be reason to focus on Private 2-Year, Private 4-Year, and Public 2-Year colleges for voting access. Their mean and median distances are consistently higher than those for Public 4-Year colleges.

Conclusions & Next Steps

In this section we will discuss the main conclusions from our work thus far and our focus moving into next semester. We have continued work to do in terms of data cleaning/incorporation and framing 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 ~75% of colleges do not have election day polling on campus, and ~91% of colleges do not have early voting on campus. Many Southern states have zero on-campus early voting options according to our current data - and the South may be an area of focus moving forward. Additionally, Private 2-Year, Private 4-Year, and Public 2-Year colleges have much larger distances to polling places than Public 4-Year colleges. Our Urban/Rural designation also revealed that colleges in rural counties continually have higher average distances to their polling places.

Next Steps

On the data front, our next steps include accessing and incorporating additional datasets. We may receive Ballot Ready's 2020 Election Day data - getting us closer to covering all 50 states in 2020. Additionally, we will continue to work on the manual cleaning and inspection necessary to improve our 2012-2018 data. After bringing that data up to par we intend to evaluate trends over time, especially in relation to the Shelby v. Holder decision (2013). And finally, we can continue to improve the coverage of our early voting data by incorporating drop-off early voting sites in addition to early voting polling places.

In terms of our future analysis, we're looking to break down the data in more meaningful ways for our client. We intend to look at college populations to understand how many students are served by an on-campus polling place. We'll also pull in college demographic data to see if this changes with the racial breakdown of the school - instead of only focusing on MSI status. Additionally, we'll focus on levels of access for individual colleges to provide target universities that may benefit the most from additional voting access. Our most imminent work will be verifying state-to-state policies on early voting (to ensure we're only looking where early voting is possible), separating community colleges from 4-Year public universities for more accurate comparisons, and re-grouping for Viacom's January press release.