Contextualizing The Impact of COVID-19 on U.S. K-12 Enrollment

Thanya Begum
Adviser: David Dobkin

Abstract

Across the United States, pandemic induced disruptions in kindergarten to 12th grade (K-12) schooling have drastically impacted public school enrollment. To contextualize these changes, this project examines shifts in enrollment in New Jersey as well as across the United States via visualizations on publicly available data such as state and national school enrollment rates. Since funding for public schools depends on the number of students enrolled, these visualizations and corresponding insights shed more light on the groups who may see less financial support for their education in the next academic year - perhaps even years to come.

1. Introduction

There is no doubt that education is a powerful means for upward mobility as well as a determinant for a person's future well-being. Yet, there still exist disparities in the availability of quality education based on a community's wealth, among other factors. The COVID-19 pandemic may exacerbate these differences further by disproportionately affecting some groups over others. In fact, recent policy research on the impact of COVID-19 on school funding predict that state funding cuts will disproportionately harm students from low-income backgrounds [3]. The reason being that schools with a large population of students in poverty depend largely on state funding, which has been influx due to COVID-19 shutdowns and other issues. This, combined with a plunge in U.S. enrollment as shown in Figure 1, leaves both school funding and student education vulnerable.

It, therefore, becomes necessary to contextualize recent disenrollment to better understand who will be most impacted. Figure 1 and other work on this topic are rightfully concerned about the policy implications of disenrollment, but their failure to breakdown the races/ethnicities, political

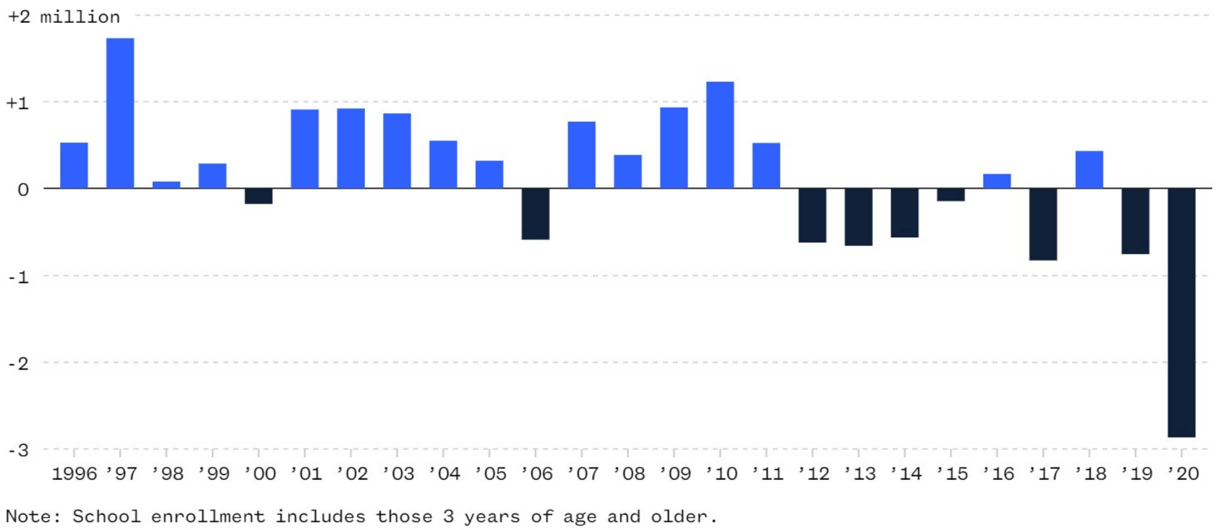

Figure 1: Annual change in the number of students enrolled in U.S. schools [2].

affiliation, and grades contributing to disenrollment leave many questions unanswered. Was a specific race disenrolling more than others? Was disenrollment higher in school districts with higher incomes? Were parents who were, say, Democratic more likely to disenroll their children? Did a particular grade lose more students than others? By answering these questions, this project draws a picture of the students who were unenrolled from school and where they went. Perhaps this information will even be useful for the long process of convincing these students to come back to public schools [5].

In contrast to other work on this topic, this project answers these questions through a data visualization driven approach in which analysis was conducted on visualizations rather than statistics. This approach involved gathering data and performing analysis on New Jersey's school enrollment first, and then applying a similar analysis on other states. Through a combination of Excel, R, Python, and geographic information system (GIS) programs, I produced multiple figures from maps to histograms. Although there were limitations in the available data and my own data analysis ability, these figures still provide interesting and at times surprising insight into disenrollment.

2. Problem Background and Related Work

One of the biggest limitations as well as advantages of this project is the recency of the subject. My goal is to analyze the impact of the COVID-19 pandemic on school enrollment, but the virus has not yet abated (especially at the time of writing) so its effects persist and, as a result, the relevant data is not always up-to-date. The pandemic also disrupted regular surveys conducted by the U.S. government such as the Private School Universe Survey, which last released data files for the 2017-2018 school year [1]. Although the topic's recency leaves room for this project to enter the scholarly conversation, it ultimately limits the available data and related work.

Of the available relevant literature, the methodological inspiration for this project came from a special analysis conducted by the National Student Clearinghouse Research Center (NSCRC) on national college progression rates [4]. College progression rate refers to the rates at which high school seniors enroll, or progress, into college. The NSCRC used a subset of U.S. high schools and colleges to examine the college progression rate across different high school characteristics (e.g. poverty, minority). One of their key findings from this approach was that "the pandemic disproportionately affected graduates of low income, high poverty, and high minority high schools, with their enrollments dropping more steeply than their more advantaged counterparts" [4]. This finding, along with their others, concerns enrollment into college specifically, but that begs the question: what could be learned by applying similar analysis to K-12 enrollment data?

There are also other opportunities to adapt the NSCRC's methods to achieve the goal of this project. The NSCRC used college progression rate data from a subset of all U.S. high schools and colleges, leading to the cautionary note: "Given that the participating high schools represent no more than 14 percent of the nation's high schools, and low income high schools are slightly overrepresented in the data, our results are not intended to be national estimates of immediate college going rates" [4]. Thanks to nationwide data collected by the National Center for Education Statistics (NCES), this project has access to information on almost all public schools servicing K-12 students. Although the NCES has not yet released data on individual school enrollments for the

2020-2021 academic year, there is still district level data to analyze. Thus, this project *can* make national estimates.

Another work of interest to this project is a brief by the Center on Reinventing Public Education (CRPE) assessing Washington state enrollment data across grade levels and among public schools—both district and charter [6]. In drawing comparisons between public and charter district enrollment, the CRPE's brief explores how students may have moved between the two types of schools based on the modality or quality of instructions. Although the brief concludes that low enrollment in public schools is not fully explained by movement into charter schools, especially since both types started the 2020-2021 academic year virtually in Washington state, it provides a starting point to apply similar analysis nationwide in case there were differences between states [6].

3. Approach

The advantage of a data visualization driven approach is the accessibility of the tools and results. The creation of insightful visualizations does not require a person to be an expert in statistics or data science. Hence, a project using this approach is a perfect starting point for someone with amateur data analysis skills. The resulting visualizations, too, make it easier to discern trends and outliers in the data. It also does not hurt that visualizations are easy to discuss with the average person, whether that's a friend or family member.

Although the eventual goal of this project is to contextualize disenrollment across the United States, it is helpful to attempt analysis on a single state before moving on to others. The benefit is that the number of school districts in a single state is much less than the total number of districts nationwide. By working on a smaller amount of school districts, especially a familiar one like New Jersey, it is easier to process and analyze the data, identify school districts of interest (e.g. Trenton, Princeton, Newark), and resolve any errors that arise as the data is manipulated. Working through New Jersey's enrollment data first also gives a framework for approaching the data of other states as well as trends to check against.

In broadening this project's focus from New Jersey to the entirety of the United States, the approach changes to involve first looking at the distribution of the percent change in enrollment across all states by a specific category (either grade or race), and then identifying outlier states to examine more closely. For example, one of the most interesting visualizations on New Jersey's enrollment data was Figure 6, so I wanted to see how grade level enrollment changed for all states. It would be too tedious to create an individual graph for each and every state. Instead, I created a plot of multiple histograms displaying the distribution of states' percent change in enrollment, as demonstrated in Figure 7. Each histogram is the distribution of percent changes for a specific grade level - later this becomes a race/ethnicity. The dashed line at zero allows for easily distinguishing how much of the distribution is a negative or positive percentage change (decrease or increase in enrollment, respectively). Using this enormous plot, I could identify outliers in the distribution and pinpoint them by sorting and filtering the underlying dataset.

4. Implementation

4.1. Data Acquisition and Preparation

As in most data focused projects, the greatest challenge was the acquisition of relevant data. Disruptions to survey collection as well as the ongoing pandemic limited the data that was available on this project's topic. For example, the NCES has not yet released 2020-2021 national enrollment data at the individual school level, so it was not possible to compute the nationwide change in enrollment for low income students since the percent of students on free or reduced lunch is available at the individual school level, not the district level. Even NCES's district level preliminary enrollment data for the 2020-2021 academic year does not break down enrollment by race/ethnicity, grade, or the number of students who are on free or reduced lunch. As a result, the nationwide enrollment data analysis conducted in this project is limited to the state level.

New Jersey's Department of Education's enrollment data, however, did provide a breakdown by race/ethnicity, grade, etc. for the 2020-2021 academic year as well as a decade going back. Other states seemed to provide the same data, so it is possible to compute the breakdown missing in the NCES nationwide data, but it proved to be too time consuming of an endeavor. In fact, it took more than a day to join New Jersey's enrollment data going back a decade in order to see how enrollment changes during the pandemic compare to changes over previous years - imagine doing the same for the rest of the fifty states.

The process first involved using Excel to quickly rename columns from the ten workbooks to be the same so as to simplify row binding them together later in RStudio. Then, once the workbooks were converted to csv format and imported into RStudio, the data was further cleaned via an all-in-one function that combined, renamed, relocated, and dropped columns to make each year's enrollment data the same. For instance, before the 2019-2020 school year when New Jersey started dedicating individual Excel workbook sheets to district and school level data, New Jersey's enrollment data had subtotals all in one file as well as counts for each sex. The cleaning function, therefore, had to filter out subtotal rows and add together columns for each sex to obtain the count of students of a specific race enrolled in school. Afterwards, another function appended the appropriate year to each dataframe and pivoted them longer such that race/ethnicity was a single column of categories. Finally, the dataframes were row binded together to create a single, massive table of New Jersey enrollment counts over the past decade.

Although creating that massive table was a time-consuming undertaking, it still beats out having to work with data tables presented by a PDF, image, or digital map. State governments are still learning to collect and provide data such that it is accessible for others to work with, as evidenced by the changing column names in New Jersey's decade worth of enrollment data. This curve was especially evident in New Jersey's voter registration data as well as vaccination rate maps. The former was presented as a PDF of a data table (probably exported from an Excel workbook), which necessitated learning to use Python to scrape the table from the file. This became quite straightforward after reading through web scraping tutorials, but grabbing data from its counterpart - an interactive, digital map of vaccination rates across New Jersey's municipalities - was not. It ultimately became too time-consuming to manually create a spreadsheet by writing down the vaccination rate for each municipality. Even the API behind the digital map could not load the large amount of data composing the map when calling the endpoint via the browser.

Fortunately, the data acquisition process of this project was not all met with difficulty. With the help of Princeton University's Data Librarians, acquiring Census population and demographic data become a breeze. The NCES nationwide enrollment data, though missing information on Illinois because they missed the enrollment data submission date, as well as geographic boundaries for school districts - for the purpose of creating map data visualizations - were also just a click away.

4.2. Analysis and Visualization

The prepared data was then imported into RStudio for analysis and visualization, primarily through the ggplot2 package. The ggplot2 package was easier to use and customize than base R plots and had a lot of tutorials available online. Using ggplot2, I created a set of functions that plotted line graphs presenting total enrollment in a given New Jersey county or district over the course of the 2010-2011 to 2020-2021 academic years. The line graphs also broke down enrollment by race/ethnicity so as to observe trends for a specific race/ethnicity. These functions came about from code used to plot New Jersey's statewide enrollment over the past decade (see Figure 5). Similarly, when analyzing nationwide enrollment data, I implemented a function that created a bar graph of the percent change in enrollment across grades for a given state. These functions complemented one-off code creating tables and visualizations to answer questions as they arose.

Despite the ease and power of ggplot2, the package is, unfortunately, not the appropriate tool for creating map data visualizations. For instance, one potential correlating factor to disenrollment was location. Rather than adding a column to the data identifying if a school, district, or state was northern, southern, etc., it was far easier and more insightful to apply that data to a map by associating the color and saturation of a school district polygon to the change in enrollment. When exploring other potential correlating factors, as well, it was no trouble overlaying an additional layer of a different style on the original figure to see if perhaps there were more purple school districts in red counties (i.e. more disenrollment in counties where a majority of registered voters are Republican). These map visualizations ultimately became a crucial part of this project because they not only looked interesting, but they were not part of the approaches taken in related work.

When it seemed that a map would be better than, say, a bar graph for visualizing some data, I used either CARTO or ArcGIS to create the figure. CARTO and ArcGIS are both GIS programs that enable the creation of geographic visualizations. I initially used only CARTO for making the map visualizations since it had an extremely accessible and appealing user interface. ArcGIS, in contrast, necessitated a tutorial from a Princeton University GIS Librarian to even begin creating with the platform. But CARTO's accessibility did not outdo ArcGIS's powerful capabilities, including access to geographic boundaries created by others and unlimited data storage afforded by the University.

The process for creating a map through ArcGIS broadly involved importing already prepared data, creating an analysis, and finally styling the map appropriately. For example, creating Figure 3 required a pipeline utilizing Python, R, and ArcGIS in that order. Python came into play when converting a PDF table of New Jersey voter registration data to a CSV in order to obtain counts of Democrats and Republicans in each county. Next, I created an R script calculating the total and percent change in enrollment between the 2019-2020 and 2020-2021 academic years. The difference between these two years was crucial for this project because the 2019-2020 academic year partially preceded the pandemic, while 2020-2021 occurred after the pandemic's start. After bringing both tables into ArcGIS, I joined the district level enrollment table to school district geographic boundaries so that the percent enrollment changes would be associated with the correct polygon on the map. These school district polygons were then styled according to the value of the percent change such that a greener school district polygon indicated a large increase in enrollment, while a purpler polygon indicated a large decrease. By adding another layer in ArcGIS, I was able to overlay a map based on the voter registration data in which blue county polygons identified a majority Democratic county and a red polygon identified a majority Republican county. Together, these layers display the distribution of disenrollment by political affiliation in a way tables or bar graphs would not be able to.

5. Results and Evaluation

5.1. New Jersey

The geographic distribution of enrollment change across New Jersey's school districts, as shown in Figure 2, provides insight into the potential correlation of location with enrollment change. As a reminder, a green polygon indicates an increase in enrollment while a purple polygon indicates a decrease. The more saturated a polygon is the higher the absolute value of that change. Notice that more districts are purple than green, demonstrating that most of New Jersey's school districts experienced a decrease in enrollment. This observation not only corroborates a larger, nationwide trend of enrollment decline, but also implies that the location of a district was not a correlating factor with change in enrollment (e.g. southern districts did experience disenrollment more than northern districts).

Surprisingly, however, there is not a large amount of disenrollment in the districts near the early epicenter of the pandemic, New York City. One would imagine that parents would be more hesitant to send their kids to school when COVID-19 cases are high in the immediate area for fear of

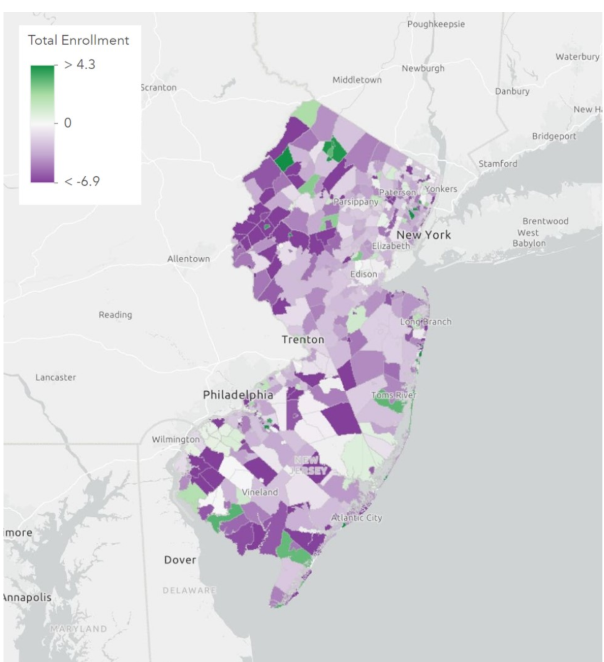

Figure 2: Percent change in enrollment across NJ school districts.

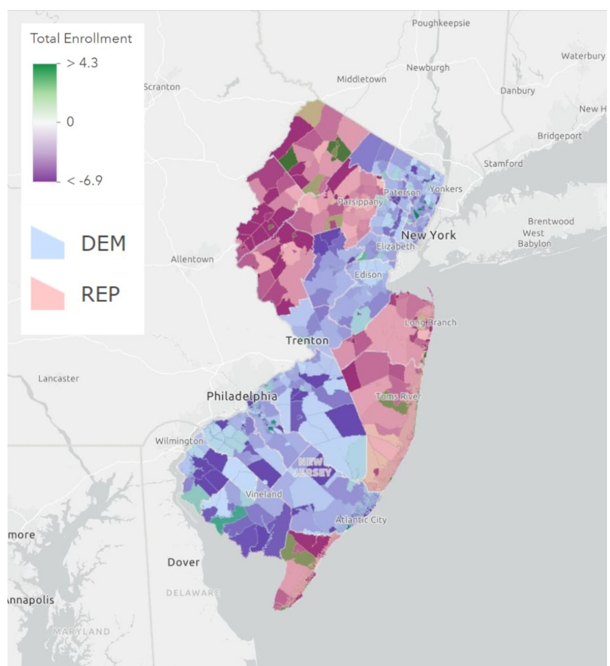

Figure 3: Percent change in enrollment across NJ school districts overlayed by county party.

infection. Yet, there is a dearth of dark purple polygons in northeastern New Jersey, nearby New York City. Compare this with the northwestern and western side of New Jersey where the dark purple polygons are more prevalent. The contrast suggests that perhaps high case loads were not a cause for parents to unenroll their children from school.

Figure 3 depicts the same geographic distribution of enrollment changes as Figure 2 with the addition of a political affiliation layer. The new layer colors a county red if a majority of registered voters are Republican and blue if a majority are Democratic. If political affiliation did correlate with enrollment changes, then one might see more green polygons below red overlays, which would indicate that districts in Republican counties generally experienced increases in enrollment, unlike Democratic counties perhaps. However, there seem to be almost equal bursts of green and purple in the separate overlays, so a correlation between political affiliation and change in enrollment are unlikely.

Line plots of New Jersey school enrollment over the past decade reveal how enrollment changes during the pandemic fit within larger enrollment and population trends. For example, Figure 4 appears to demonstrate that enrollment tallies between the 2019-2020 and 2020-2021 academic years fall in line with trends over the past decade. But if one looks closely, it looks as though the enrollment of white students saw a particularly steep drop in comparison to the relative steadiness in enrollment of the other races/ethnicities. Still, the trends in Figure 5 suggest that this steep drop may be explained by the decreasing number of school-aged white students in New Jersey and not

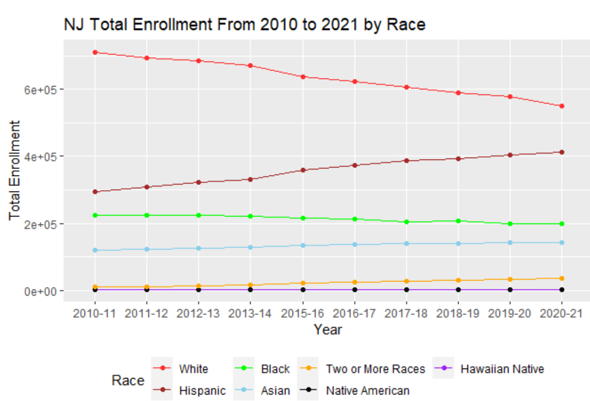

Figure 4: NJ total school enrollment over time by race.

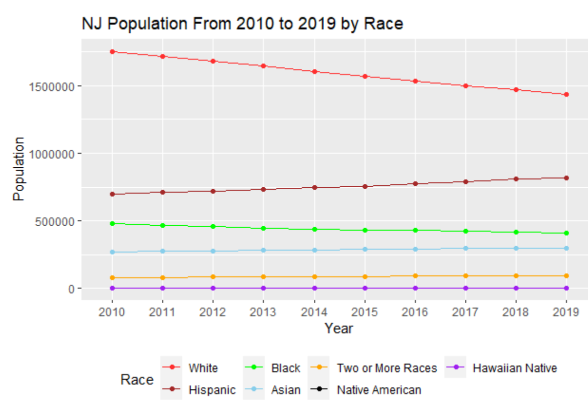

Figure 5: NJ school-aged population over time by race.

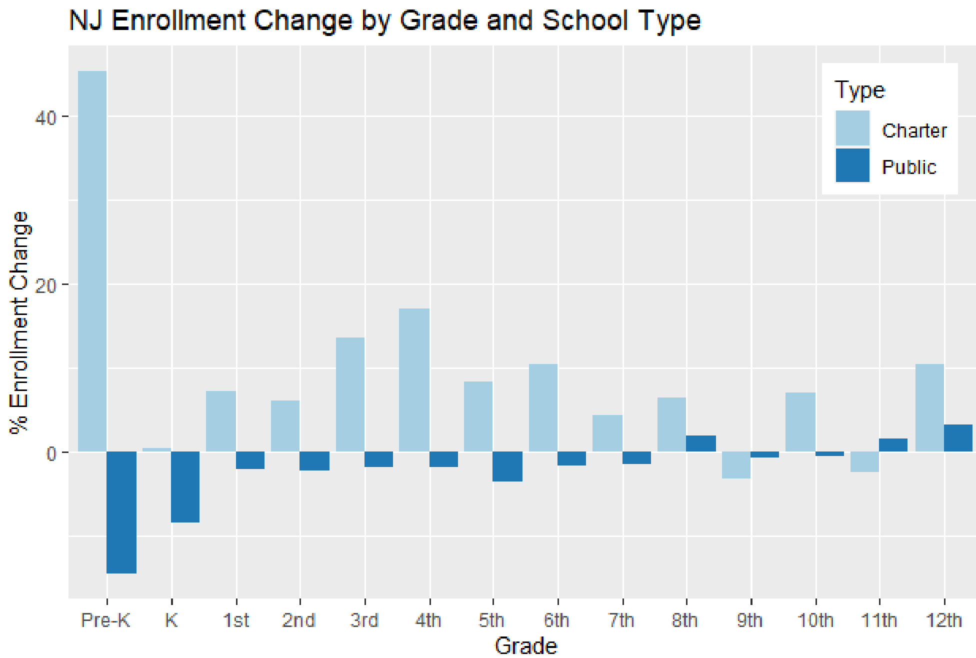

Figure 6: Percent change in NJ enrollment by grade and school type.

by disruptions by the pandemic.

Bar graphs of enrollment by grade level, as well, are valuable in learning which grades experienced more disenrollment. Figure 6 demonstrates that in public school districts, most grade levels saw some decline in enrollment, but pre-kindergarten and kindergarten were hit the hardest. A potential explanation is that since Pre-K is usually optional, some parents may have opted to skip it altogether. Kindergarten, too, marks the start of a child's public school education and, as a result, parents could afford to hold off on starting their kids in school. The reasons for holding off could range from the modality of the education (i.e. virtual vs. remote) to the risk of infection. Interestingly, terminal grades like eighth grade or twelfth grade had an increase in enrollment, as if students chose to stick it out until they graduated. And eleventh grade presumably experienced an increase in enrollment because students wanted to stay and finish requirements pivotal for college applications.

In contrast to the drops in public school district enrollment presented by Figure 6, charter schools experienced a consistent increase in enrollment. This graph is deceptive, however, because charter schools have a small student population so even an increase of a few students equates to a large percentage change in enrollment. Nonetheless, charter schools may have been an educational alternative for parents who disenrolled their children from public school districts. Some advantages of charter schools could be their maintenance of in-person learning or greater experience in delivering a virtual education. Yet, both public and charter school districts lost students in the ninth grade. Perhaps like kindergarten, parents chose to hold off on high school until the pandemic passed.

5.2. Nationwide

Recall that in Figures 7 and 8, each histogram is the distribution of percent changes for a specific grade level or race/ethnicity. And the dashed line at zero allows for discerning how much of the distribution is a negative or positive percentage change (i.e. how many states experienced a decrease or increase in enrollment). The advantage of these enormous plots is in identifying outliers in the distributions and then pinpointing the state represented by the outlier by sorting and filtering through the underlying dataset.

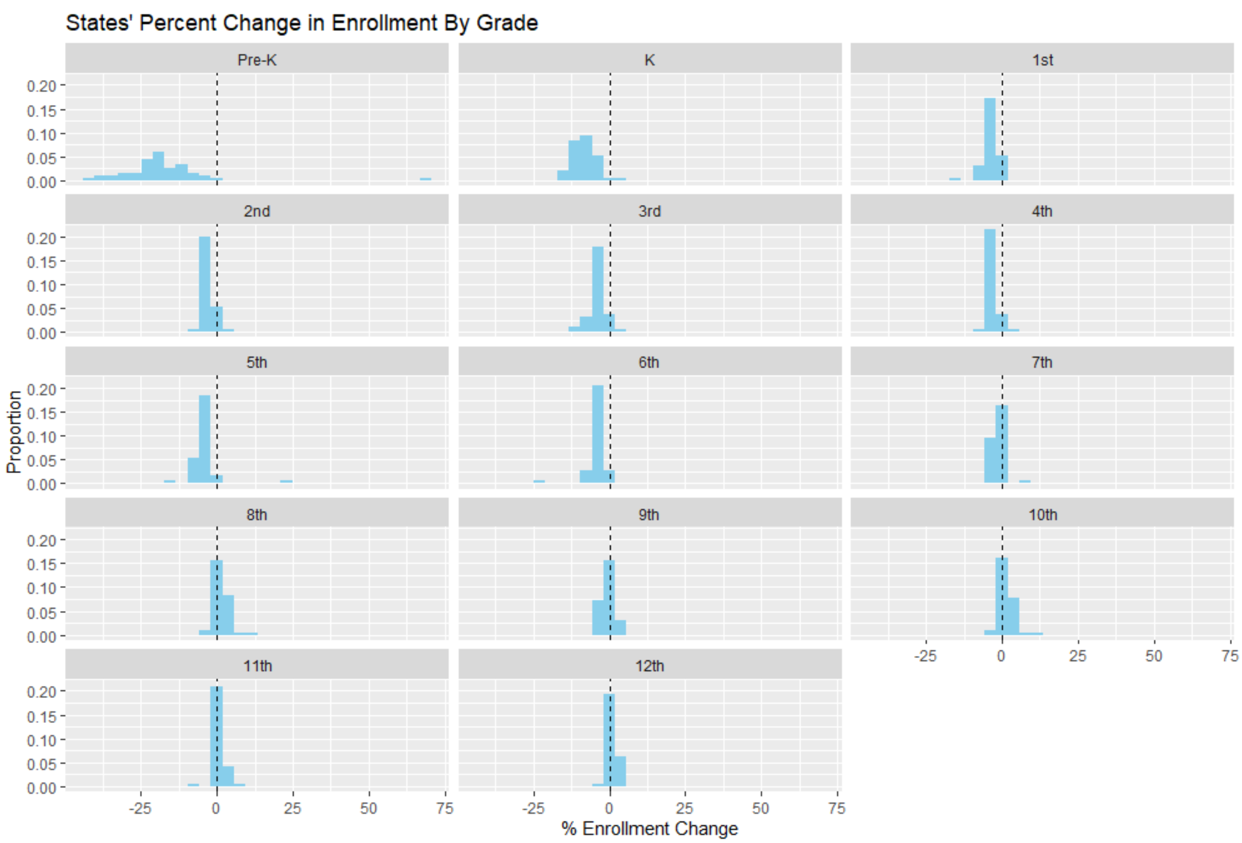

Figure 7: Distribution of states' percent change in enrollment by grade.

Figure 7 corroborates the changes in enrollment seen across grade levels of New Jersey's public school districts. Unsurprisingly, grades eight, ten, eleven, and twelve have a larger portion of their distributions to the right of the zero line, indicating an overall increase in enrollment. In fact, these four grade levels had the only positive average percentage change in enrollment across the country - 1.39%, 1%, 0.96%, and 0.92% respectively. It also appears that most, if not all, states witnessed drops in Pre-K and kindergarten enrollment because the distributions of those two grades are almost entirely to the left of the zero line. However, there is an outlier in the Pre-K histogram. That outlier is Wyoming, which actually had a 67.4% increase in its Pre-K enrollment between the 2019-2020 and 2020-2021 academic years. Perhaps there is something to be learned from how Wyoming avoided the plunges in enrollment seen by other states.

Figure 8 also corroborates the changes in enrollment seen across races/ethnicities of New Jersey's public school districts. Most races/ethnicities have a larger portion of their distributions to the left of the zero line, indicating an overall decrease in enrollment. However, enrollment of students

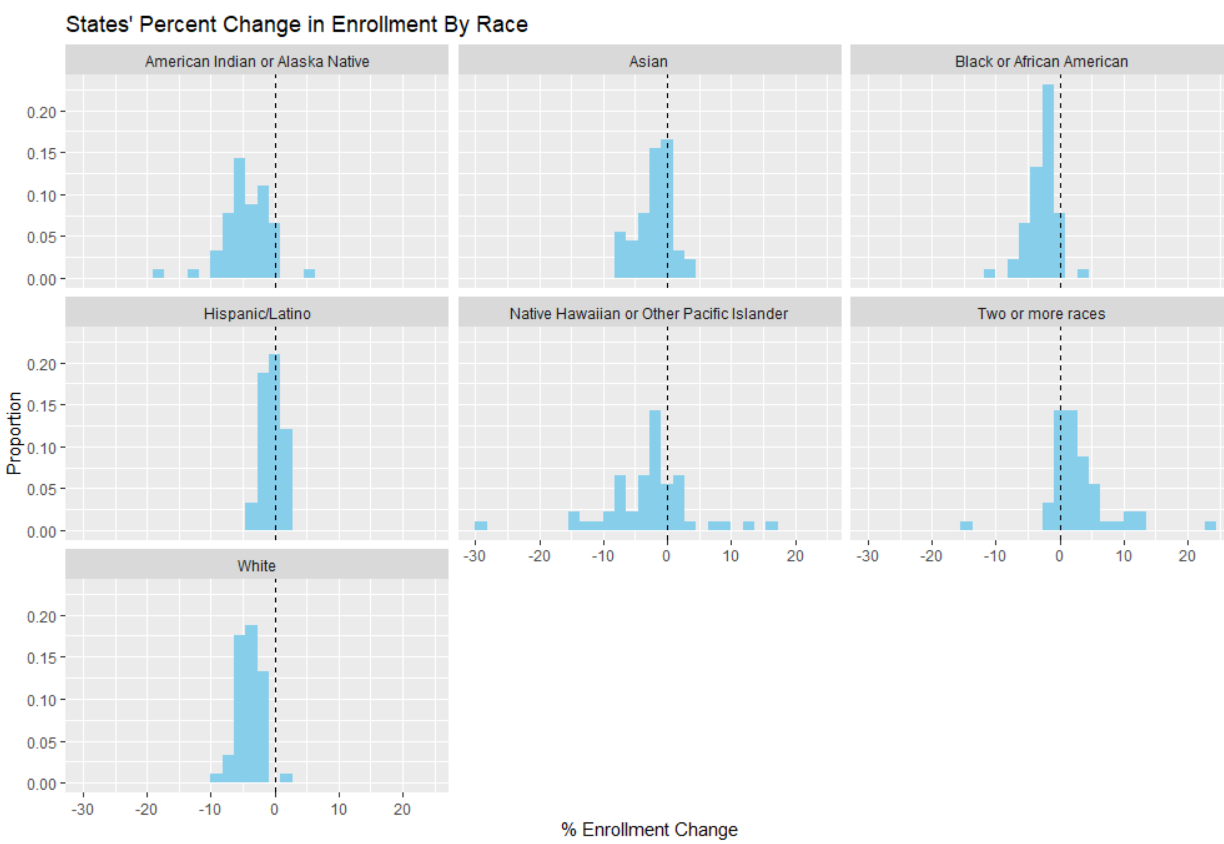

Figure 8: Distribution of states' percent change in enrollment by race.

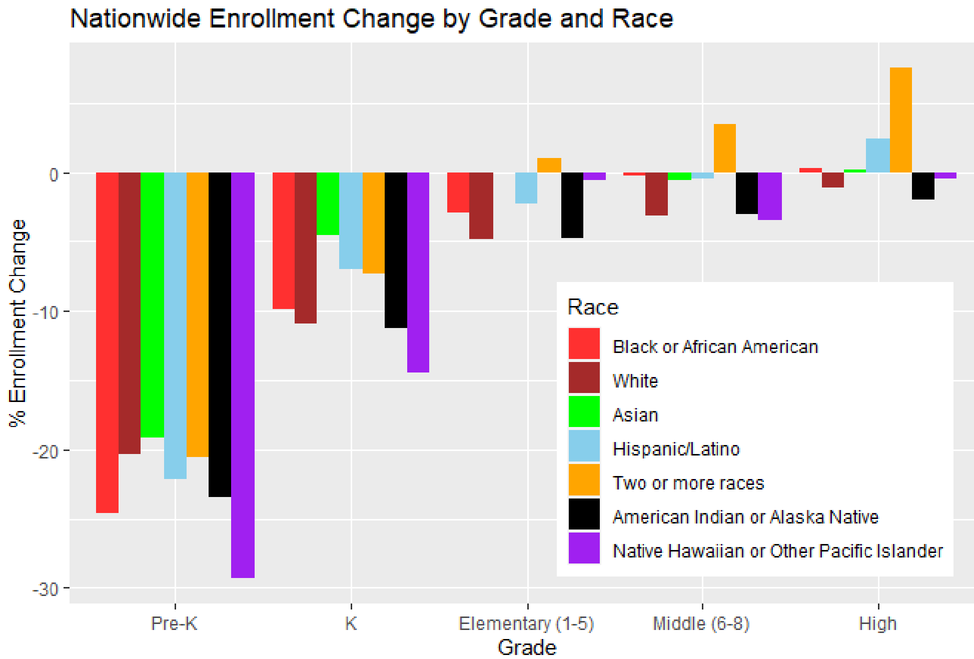

Figure 9: Percent change in nationwide enrollment by grade and race.

identifying as two or more races appears to be the only race/ethnicity that saw an increase during the pandemic. In fact, across the United States, those of two or more races had an average increase in enrollment of about 3%.

Finally, Figure 9 breaks down enrollment change by grade and race/ethnicity so that enrollment changes across both those categories can be compared. Notice that all races/ethnicities exhibited a drop in enrollment for both Pre-K and kindergarten, implying that parents across all races/ethnicites shared the same sentiment of holding off on schooling. However, the unanimity stops after kindergarten with enrollment for students of two or more races showing a surprising increase at the elementary, middle, *and* high school levels.

6. Conclusions and Future Work

Using a data visualization driven approach ultimately enabled me to make a some conclusions contextualizing disenrollment during the pandemic. First, no one region, political party, or race seemed to be responsible for the plunge in school enrollment in the United States. Over and over, it appeared as though parents across the political spectrum and of different races alike disenrolled their children. Some reasons could be to avoid remote schooling or risk of COVID-19 infection.

Additionally, grade levels that could afford to be pushed off were, such as Pre-K, kindergarten, and ninth grade. In contrast, grades pivotal for college (i.e. grades 10-12) saw retention, even increases, in enrollment, suggesting some students stuck it out to secure their future. Finally, one possible path out of public school districts were charter schools, which may have been desirable to some parents for their in-person instruction or experience with virtual education.

Given the opportunity to conduct more work on this topic, I would create additional map visualizations as well as incorporate vaccine and caseload data. Due to the unavailability of district and school level NCES enrollment data, I was limited to creating maps for the state of New Jersey, but this project would be further enriched by having the same visualization for each state in one large map of the United States. I would also include vaccine and caseload data to see if they were correlating factors with changes in enrollment. For instance, perhaps districts that had a greater rate of vaccination had less disenrollment.

7. Acknowledgments

I would first like to thank Professor David Dobkin and my peers in COS IW08 for their advice and feedback during our class meetings that helped shape this project. I would also like to thank the data librarians who were pivotal in my search for datasets as well as journey learning to create geographic visualizations. Finally, I would like to thank my friends and family for their questions and feedback during discussions of my research.

8. Honor Code

This paper represents my own work in accordance with university regulations.

/s/ Thanya Begum

References

[1] "Private school universe survey (pss) - data and documentation." [Online]. Available: https://nces.ed.gov/surveys/pss/pssdata.asp

[2] "School enrollment plummeted in 2020, data shows." [Online]. Available: https://www.nbcnews.com/news/us-news/school-enrollment-plummeted-2020-data-shows-n1282017

[3] K. Blagg, E. Gutierrez, and V. Lee, "How covid-19-induced changes to k–12 enrollment and poverty might affect school funding," 2021. Available: https://eric.ed.gov/?id=ED613234

[4] J. Causey *et al.*, "A covid-19 special analysis update for high school benchmarks," 2021. Available: https://eric.ed.gov/?id=ED613158

[5] A. Kamenetz, C. Turner, and M. Khurana, "Where are the students? for a second straight year, school enrollment is dropping." Available: https://www.npr.org/2021/12/15/1062999168/school-enrollment-drops-for-second-straight-year

[6] S. Tuchman and G. Heyward, "How is school enrollment in washington state shifting during covid-19?" 2021. Available: https://eric.ed.gov/?id=ED610650