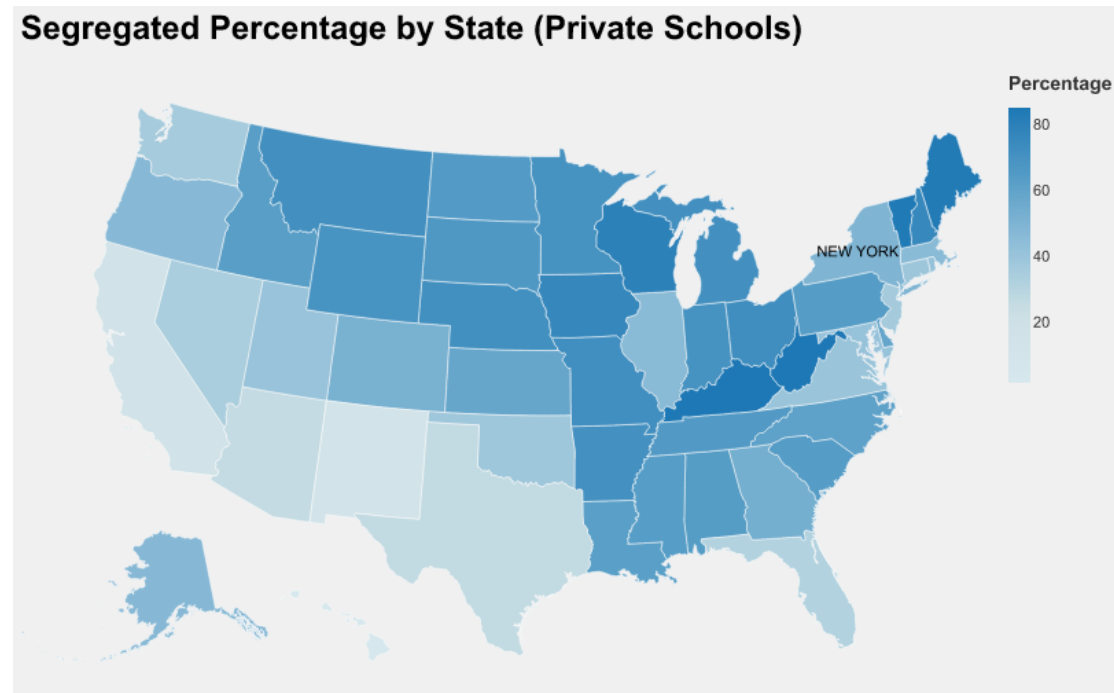


Graph1:



1. Tasks:

This visualization is designed to support comparisons between different U.S. states regarding the percentage of highly segregated private schools. The map helps the viewer to identify states where private school segregation is most severe and compare the segregation rates across states and regions.

Also, users can observe regional patterns and clusters of high or low segregation. Especially, I highlight the result of New York States, where the article background takes place. Readers can compare New York with other states on the segregated school proportion among private school.

The key takeaway for the viewer is to understand how school segregation varies across the U.S. and recognize which states exhibit the highest concentration of segregated private schools. This visualization can inform the readers whether New York is a seriously segregated states in the United States, which is mentioned in the article by the author. Additionally, it can provide the background knowledge of the situation among the United States. We can find that the South and East parts of the country is having a more segregated education environment, which can help users understand as well as thinking about the following content in the article in the aspects of economy and race, etc.

2. Data:

The data used for this visualization was sourced from the project document, which is the NCES Private School Dataset (2015-16), listed in the first one of Datasets. The dataset provides lots of information including the private school's name, id, state, county and other basic information of United States Private School in this year. Besides, the data contains information about school level, school type and community type. Furthermore, there are columns about students' proportion such as the counts and percentage of different grades as well as different races. Finally, the last several columns are about the school's education resources situation.

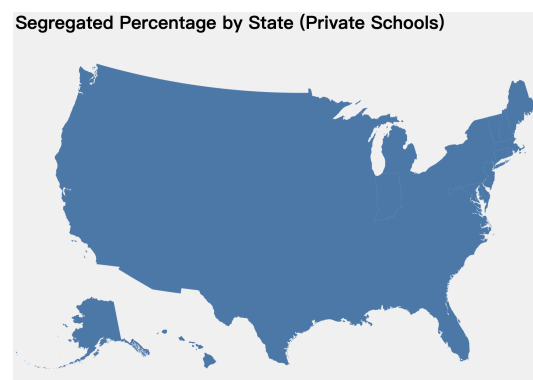
Data Processing and Cleaning:

Only effective columns are preserved in the first step, including location, percentage of races.

- NaN values were removed to ensure accurate calculations as the second step.
- The percentage columns were converted to numeric values due to the wrong format after reading from csv, so that we can later use the numeric values to color the graph.
- Aggregation by state: The dataset was grouped by state to compute the total number of private schools and the number of highly segregated private schools.
- Defining segregation: A school was classified as "highly segregated" if either $\geq 80\%$ Black or $\geq 80\%$ White students were present. Add a new column representing the percentage of segregated schools in this state.
- State code mapping: Due to the limitation of creating a map in Altair, a dictionary was used to assign appropriate numeric state IDs. And the dataframe is transformed to a proper format which can be a valid input for map drawing.
- Also import from vega_datasets to gain the necessary data for map drawing.

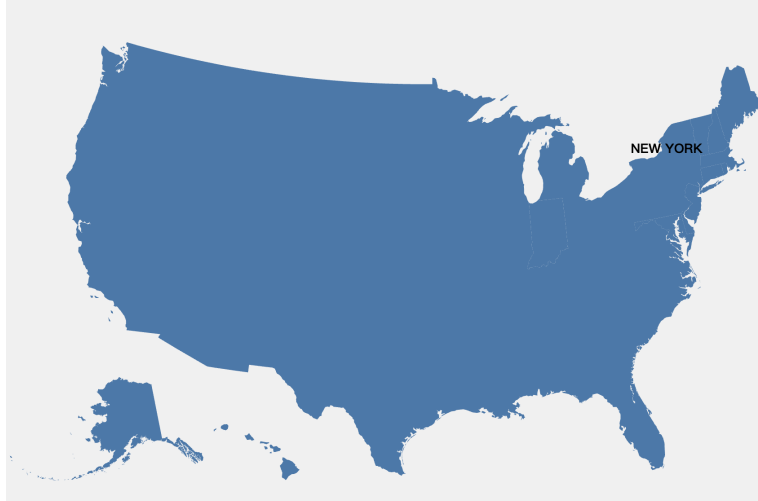
3. Design Process:

Initially, I try to draw a map of United States without any input color method and details, and get:



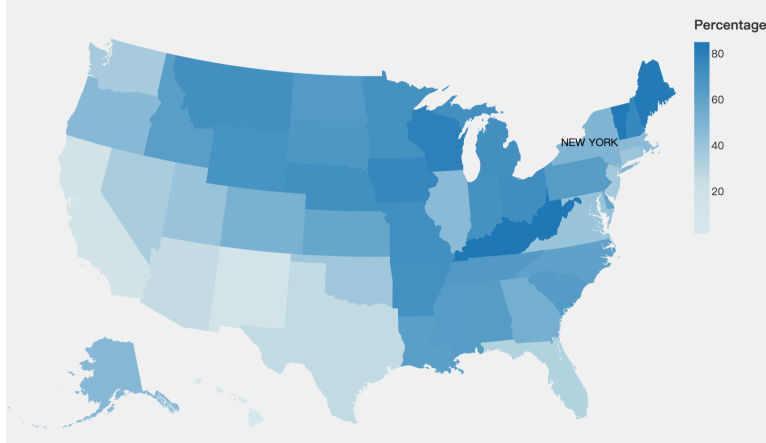
Then I try to emphasize the location of New York by another chart layer of text.

Segregated Percentage by State (Private Schools)



After that, I add the color which represents the percentage of segregated schools.

Segregated Percentage by State (Private Schools)



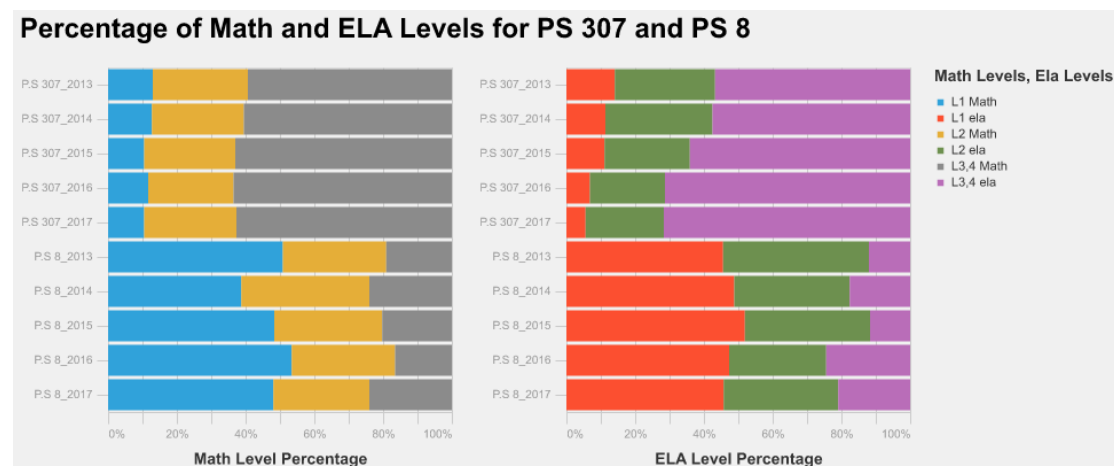
Finally, I do some detail work, such as adjust the stroke, reaching the final results.

For the initial version of the graph, I lost the picture, but it is kind of ugly and improper in the aspect of color. It uses a scale from green to red instead of using the legend shown in the final work. Also, it focus on the trend of US, ignoring the detail situation of New York, which should be a crucial point for this chart in the article.

4. Qualitative Self-Evaluation:

Overall, the visualization is effective in communicating the extent of segregation in private schools across the country. The use of color helps convey the severity of segregation in different states, and the geographic layout ensures that patterns are easily identifiable. I think the color used here is well aligned with using a saturation color scale for the quantitative values. However, there are areas for further improvement. A comparison with public schools would provide a more comprehensive understanding of segregation in the education system. And I think there can be a partial focus on New York States and surrounding states to help readers focus on the background environment of the article.

Graph 2:



1.Tasks:

This visualization compares student performance at P.S. 307 Daniel Hale Williams and P.S. 8 Robert Fulton from 2013 to 2017, highlighting disparities in Math and ELA proficiency levels. The two stacked bar charts illustrate how student performance is distributed across three levels: Level 1, Level 2, and Levels 3 & 4, where 1 represents the best performance and 3 or 4 are the worst levels. By normalizing the percentages, this visualization allows direct comparison between the two schools. Also, there are comparisons of each school's level distribution among recent years, showing the trend of students' level variation.

The key takeaway for the reader from this graph is the clear achievement gap between the two schools. While P.S. 8 shows a consistently higher percentage of students in Level 1, indicating strong academic performance, P.S. 307 has a much larger share of students in Level 3 and 4, struggling to meet basic proficiency standards. This aligns with the segregation results in unequal access to quality education, experienced teachers, and advanced coursework. The readers can easily find that P.S. 8 students have a much higher academic performance in both math and language. Furthermore, the visualization reveals that this disparity persists over multiple years, underscoring the long-term consequences of segregation. Though the difference between two schools remains stable, readers can still focus on the information of year and get informed of the detail change of a specific schools' performance.

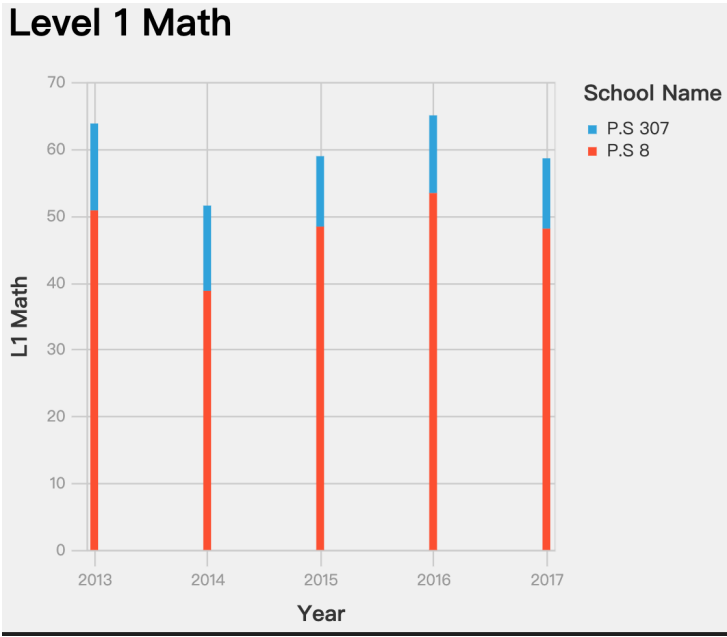
2. Data:

The data for this visualization was sourced from Historical New York school test scores (math and language), which is provided in the datasets list in the project documents. It records test performance for all public schools in the state in both math and language. The dataset includes variables such as school name, year, mean scale score, level percentage and level counts. Since the focus of this visualization is on school-wide performance rather than individual grade levels, only the data for "All Grades" was selected as the first step of data preprocessing. Besides, the dataset was filtered to retain only P.S. 307 and P.S. 8, ensuring that only relevant data was included. Next, the year column was formatted properly, and non-numeric values in the percentage columns were converted to numerical format for accurate

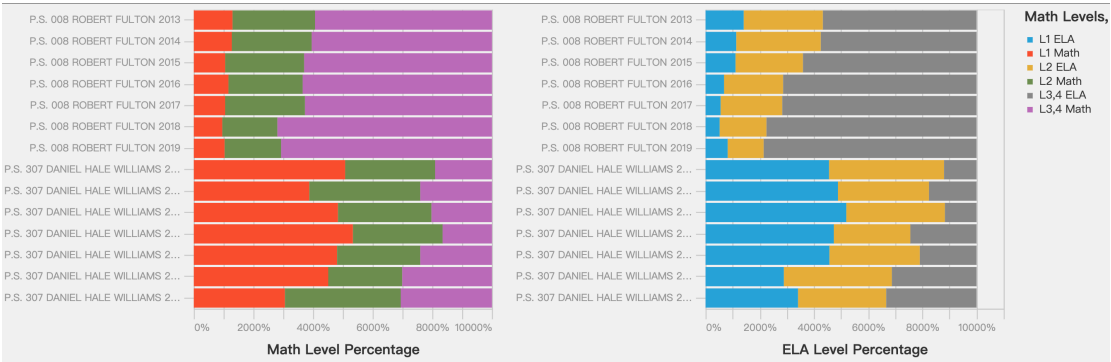
calculations. Finally, to create the stacked bar charts, the dataset was transformed using the melt() function, which reshaped the data into a long format, making it easier to plot stacked proportions.

3. Design Process:

For the very first try, I used simple bar chart without stack the data, but this is obviously not a proper one to show the statistics of student level distribution, it can only show few level data and is not clear for comparison.



Then I move to the stack bar which can shows in percentage all the levels percentage, and also make the y axis a combined information of school as well as year.



And finally, I adjust the details including the tooltips and titles, and the y axis name are also adjusted to a more simplified version while maintaining showing enough information.

As for me, I prefer the last version to my initial design since it can show more information to the users, the first one is ugly and less informative.

4. Qualitative self-evaluation:

This visualization is effective in communicating the long-term consequences of educational

segregation. The side-by-side structure makes it easy to see that P.S. 8 students do better than P.S. 307 in both Math and ELA, reinforcing the idea that access to resources and integration plays a crucial role in academic success. The color scheme was chosen properly to ensure clarity, with distinct colors representing each level. The design also aligns well with using stacked bars instead of raw scores, which is a better form of data visualization for percentage data.

However, there are areas for improvement. Adding annotations of related background environment changes such as policy can provide further context for reader understanding. Additionally, the color encoding can be reconsidered, such as trying to label two schools in different scale, helping in better recognition. And the sorting of y axis can be improved based on year prior than school, but this point deserves further discussion. Last but not least, a broader comparison with other schools in New York could also provide additional insight into whether these patterns are unique to P.S. 307 and P.S. 8 or reflective of a larger systemic issue.