

Interactive Redesign

Jonas Zhonghan Xie
jonasxie@umich.edu

Review of Static Design

Generally, I received positive feedback about my static design from both my discussion partner, Ruoyi Gao, and graders. Four charts are designed with clear domain goals: visualizing the intertwined housing segregation and school segregation in New York City. The data insight is well conveyed and integrated with the New York Times article. My partner seems to like the color encoding of my public school map and the proficiency-income scatter plot. The general edit suggestions are mainly about the readability of the dots and the labels. I list out some detailed suggestions from my partner and GSIs.

1. **Some axes should be well labelled.** For example, in the scatterplot of proficiency rate and neighborhood income, the x-axis should be labelled as something like neighborhood median income. My partner was confused by the plot at the first sight. I original thought is that the title has indicated that the chart is visualizing the bivariate relationship between neighborhood income and school proficiency rate. The readers can infer the meaning of axes from the title. But the graph reduces cognitive load.
2. **Some of the text is hard to read**, because of the size and the use of grey color. Also in the NYC public school map, some circles are darker red while some are light. That is because I also use the opacity encoding for the average proficiency rate. So that the high-performing schools are highlighted. I wanted to save some data-ink so I did not add a legend for the opacity encoding. Also, Ruoyi thought the circles are too small. It is somehow hard to read the map. It is always tricky to clearly visualize locations of DOIs. I will expand on how I dealt with it in my following revision.
3. **More flexibility:** it is specifically for the scatter plot. I set the top-performing and bottom schools as top 10% and bottom 10%. Ruoyi suggested that we should let user to make the decision on the threshold. It is reflected in my interactive version.
4. From GSI: **Fig 4. contains too many lines and the colors are difficult to see making it difficult to distinguish the jurisdictions.** I want to push it back a bit. I made the lines besides NYC and National average be lighter. The idea is that the audience is focused on placing NYC in other municipal jurisdiction in U.S.. They don't need to compare other jurisdictions besides NYC. Still, I can reduce some other jurisdictions and make the graph clearer.

Redesign

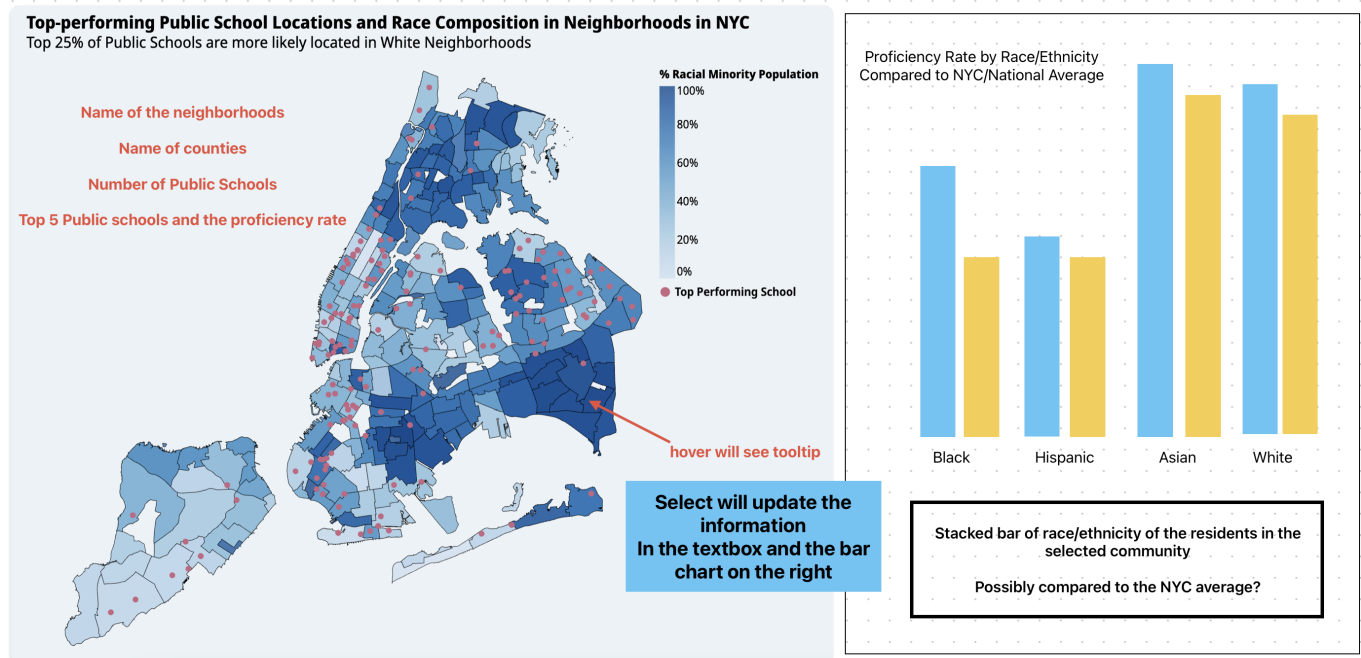
Based on the suggestions, I redesigned the New York City school map and the scatter plot.

NYC School Map

The basic idea of the NYC school map is visualizing the distribution of quality public educational resources in the city. Based on the static version, we can find that most of the high-performing public schools are located in White neighborhoods, suggesting a unequal access to public education across different race/ethnicity. One of the evident drawbacks of the static map is that it **only provides an overview at the city level**. The audience may need to know more about the detail within every neighborhood. Also, there is variance across

different neighborhoods. There are also some high-performing schools located in neighborhoods with more racial minorities. Therefore, we need to use some interaction techniques to combine the high-level overview and neighborhood-level details.

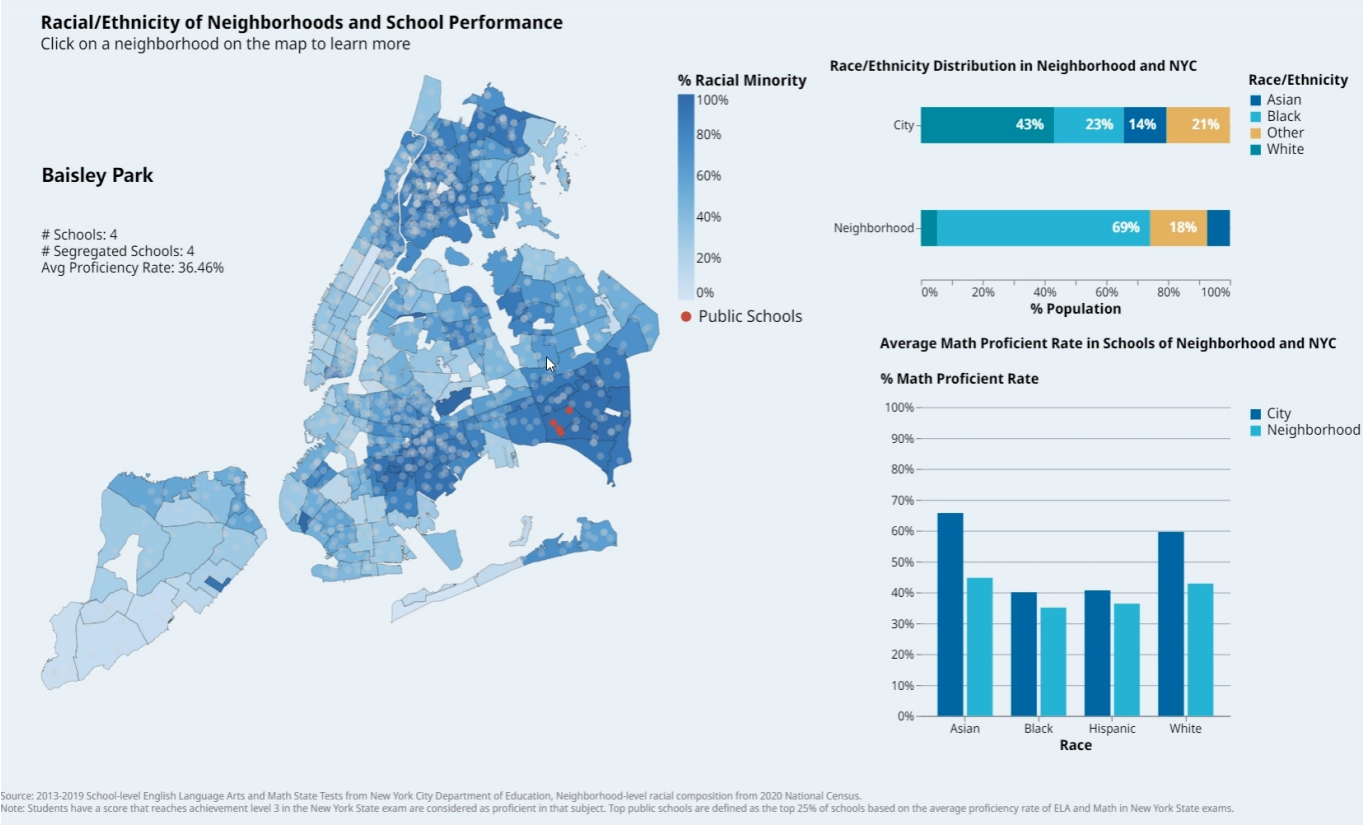
The main approach I used in the interactive map is similar to the **zoom in and zoom out** technique. The interactive map consists of two parts: NYC map on the left and the neighborhood details on the right.



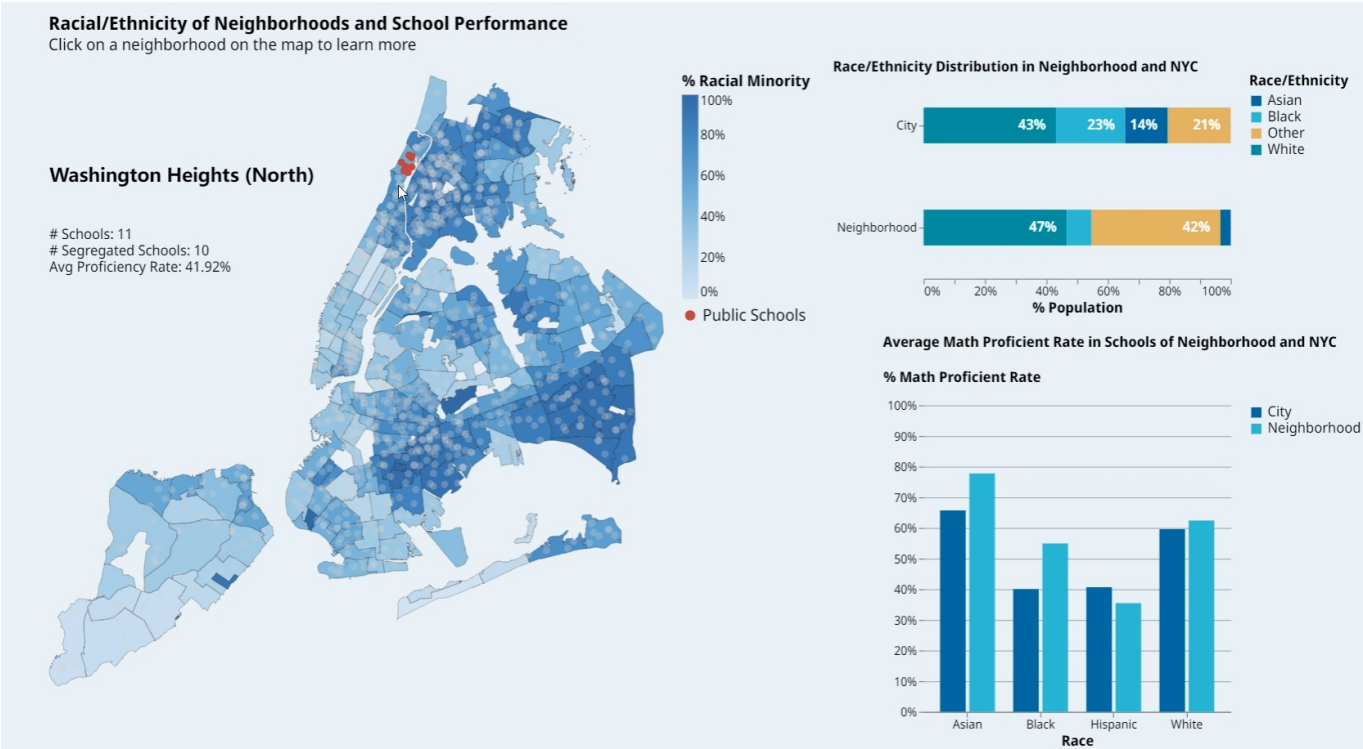
For the map part, I mainly reused the static version of the NYC map. But this time I will show the locations of all the schools. Users are able to select the neighborhood which they are interested in. The locations of the schools in that neighborhood will be colored as red. The rest of the schools will be colored lightgrey, highlighting the selected neighborhood. To better use the space on the upper-left part of the map, I created a text layer to show the name of the selected neighborhood, number of public schools and number of segregated public schools. I originally planned to show the Top 5 schools. But it is kind of messy to achieve it in `altair` as the `transform` functions of `altair` is not so accessible for me. So I kept it simple for this assignment.

The newly-added right part is showing the details of schools in that neighborhood. I planned to visualize the demographic distribution of that selected neighborhood and the heterogeneity of performance across different race/ethnicity. The bar chart compares the neighborhood's detail to NYC averages, showing some variance across neighborhoods.

The key to connecting both parts is selecting the neighborhood on the map. The way I did it is I added a `selection_point()` to the school location layer and select the NTA (Neighborhood Tabulation Area) that is `nearest` to user's selection. The simple reason why I did not add the selection to the map layer is that I could not find documentation for that and selecting nearest school point is easier. The right part used `transform_filter()` to limit the schools to that NTA and used `transform_aggregate` to calculate the averages.



Baisley Park schools are performing worse than NYC for all race/ethnicities

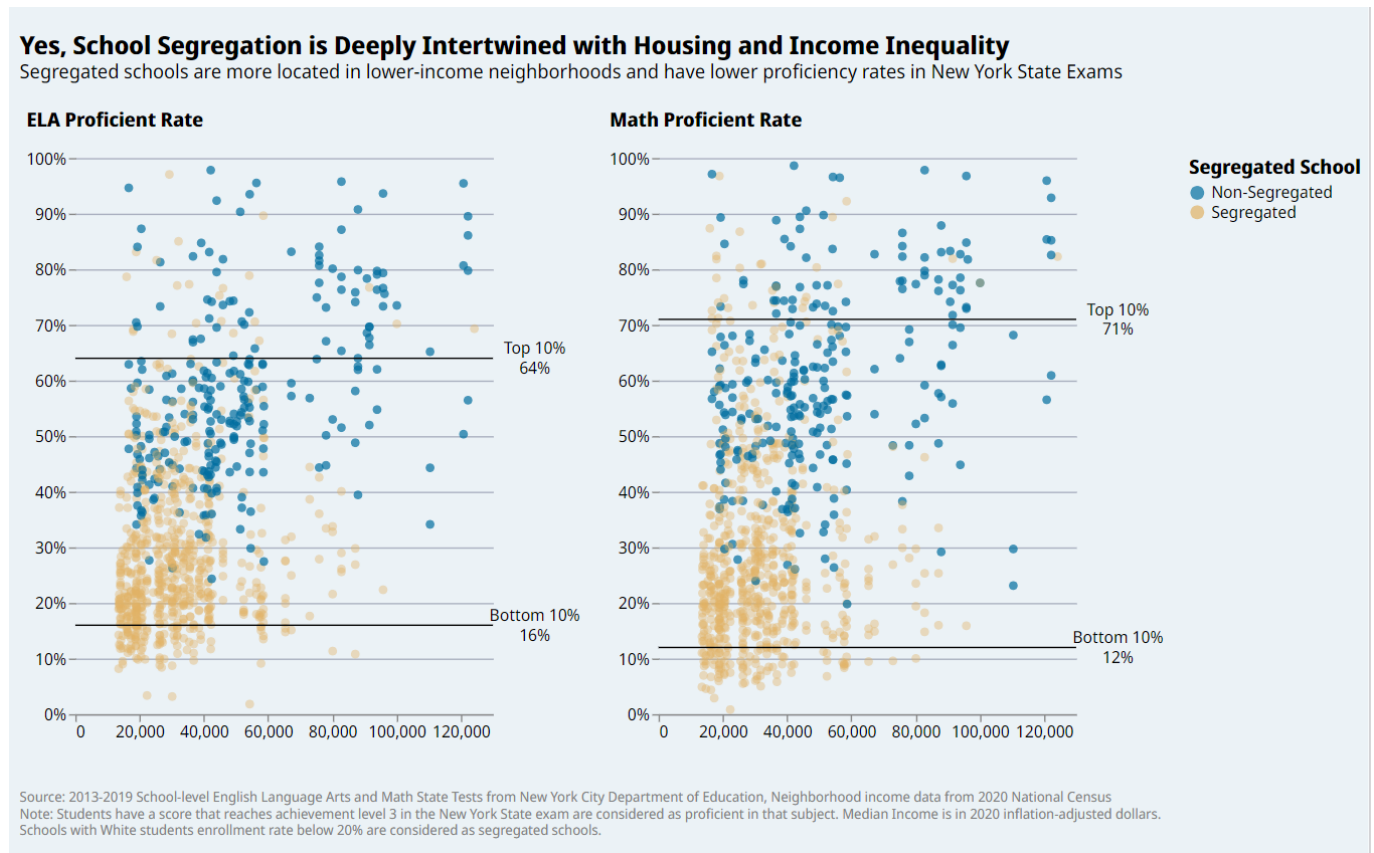


Schools in Washington Heights perform well on average. We find large Hispanic population in that NTA though.

With this zoom in zoom out interactivity, I did find some variance across the city. For Baisley Park (NTA in southern Queens), all public schools are racially segregated and perform worse than NYC averages across all race/ethnicity. It is not surprising according to the news article. However, Washington Heights seems to tell a different story. Most of the public schools are segregated schools. We find large proportion of residents are not White. However, public schools perform somehow better than NYC averages for White, Asian and Black students. Audience can use this map to explore heterogeneity across neighborhoods.

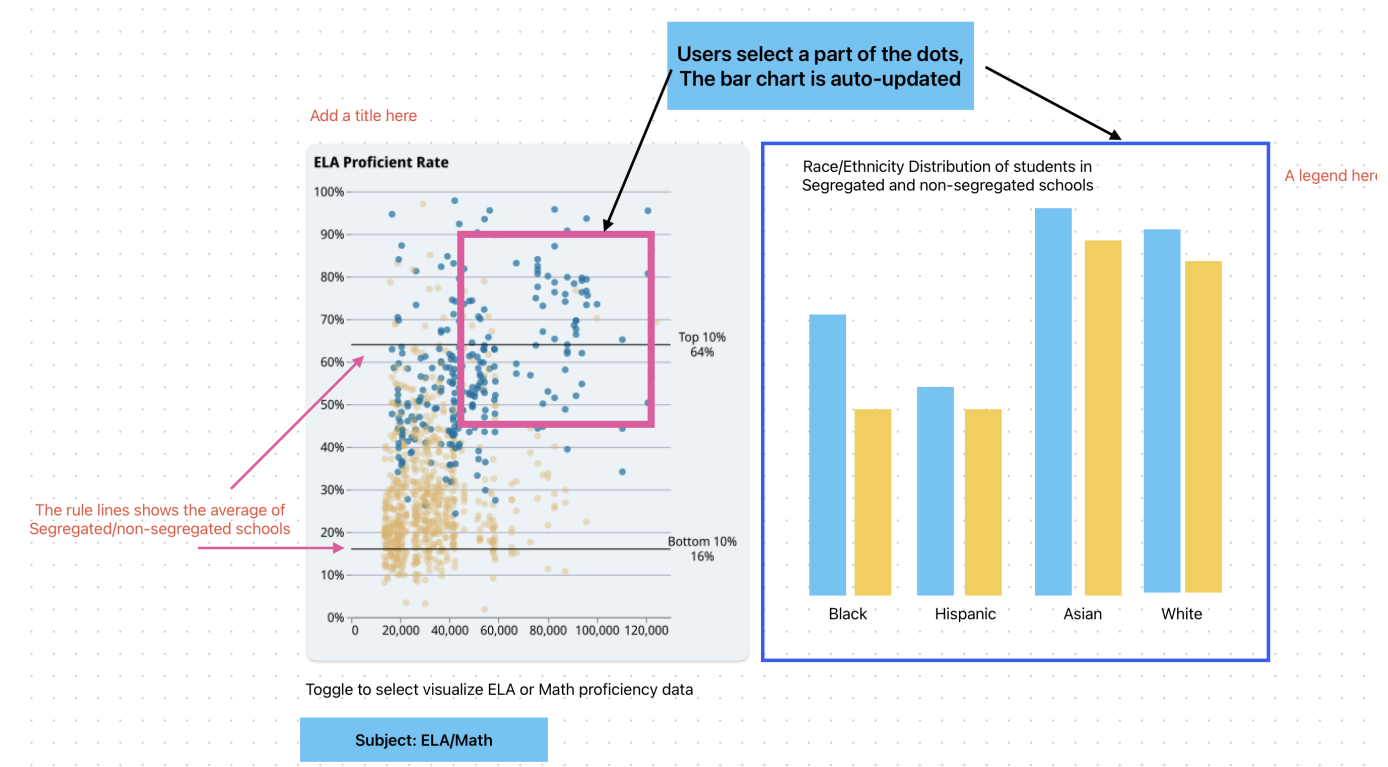
Another major update to the map is improving the rendering efficiency. I streamlined the datasets that the map layer and school layer use. It used to take more than 50 seconds to render the map on my machine. But now it only takes 6-7 seconds to render the whole dashboard. The map is now way more responsive.

School Performance-Neighborhood Income Chart



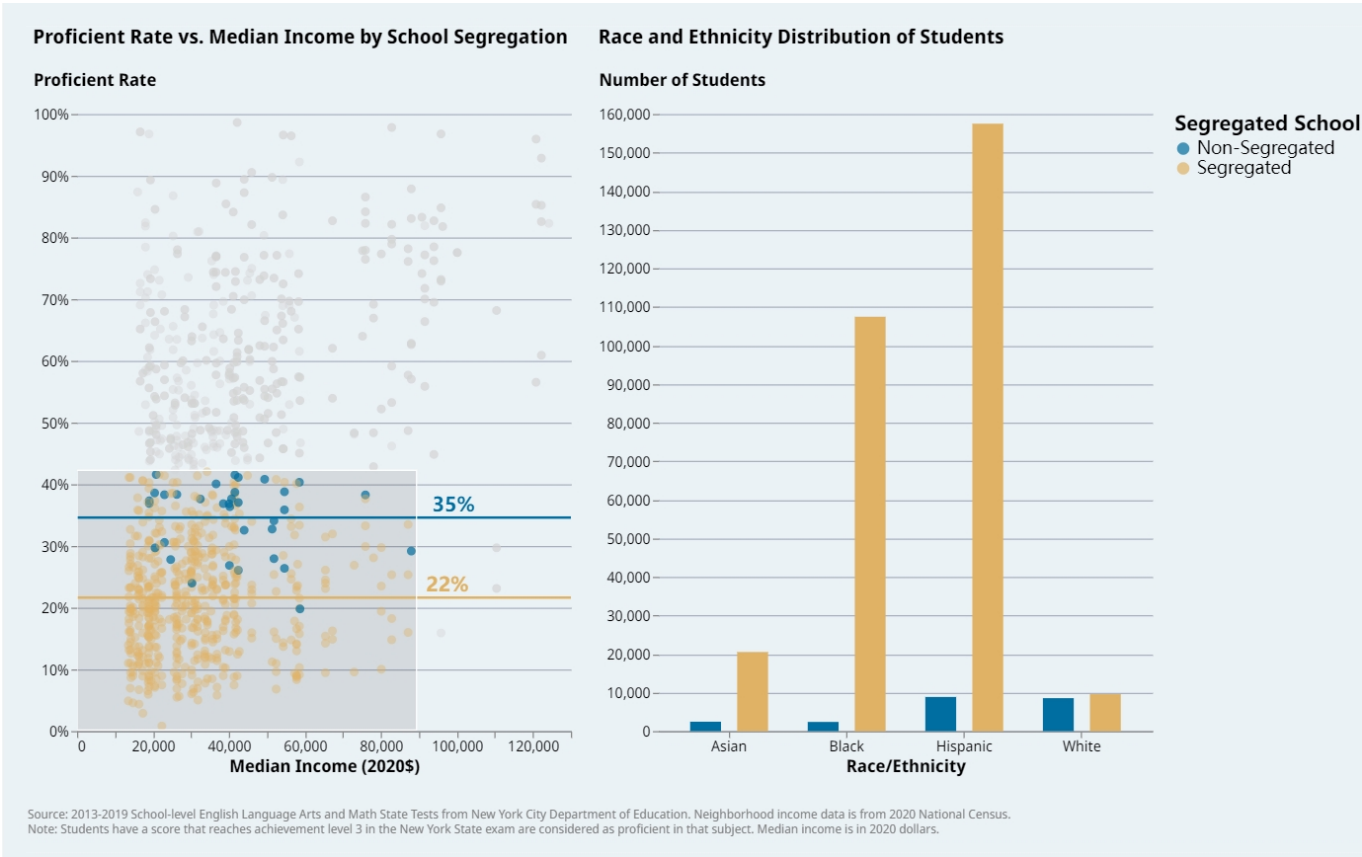
The static version of the scatter plot shows the strong correlation between school performance and the household income of the neighborhood where the school is located. It also shows the categorical relationship between racial segregation and income inequality: non-segregated schools (blue dots) are more likely to be in higher-income communities and higher up in proficiency scale.

The main goal of the map, again, is to visualize the intertwined relationship between school segregation and racial segregation. What it does not fully achieve is that there is a lot of variance embedded in the dichotomy of segregated and non-segregated schools. Based on my experience, I found that though Asian students are more likely to be in segregated schools, their test performance is still very strong. The NYT article mainly tells the story from a Black American perspective. Though it is representative, it does not fully show the story of underrepresented students in NYC.

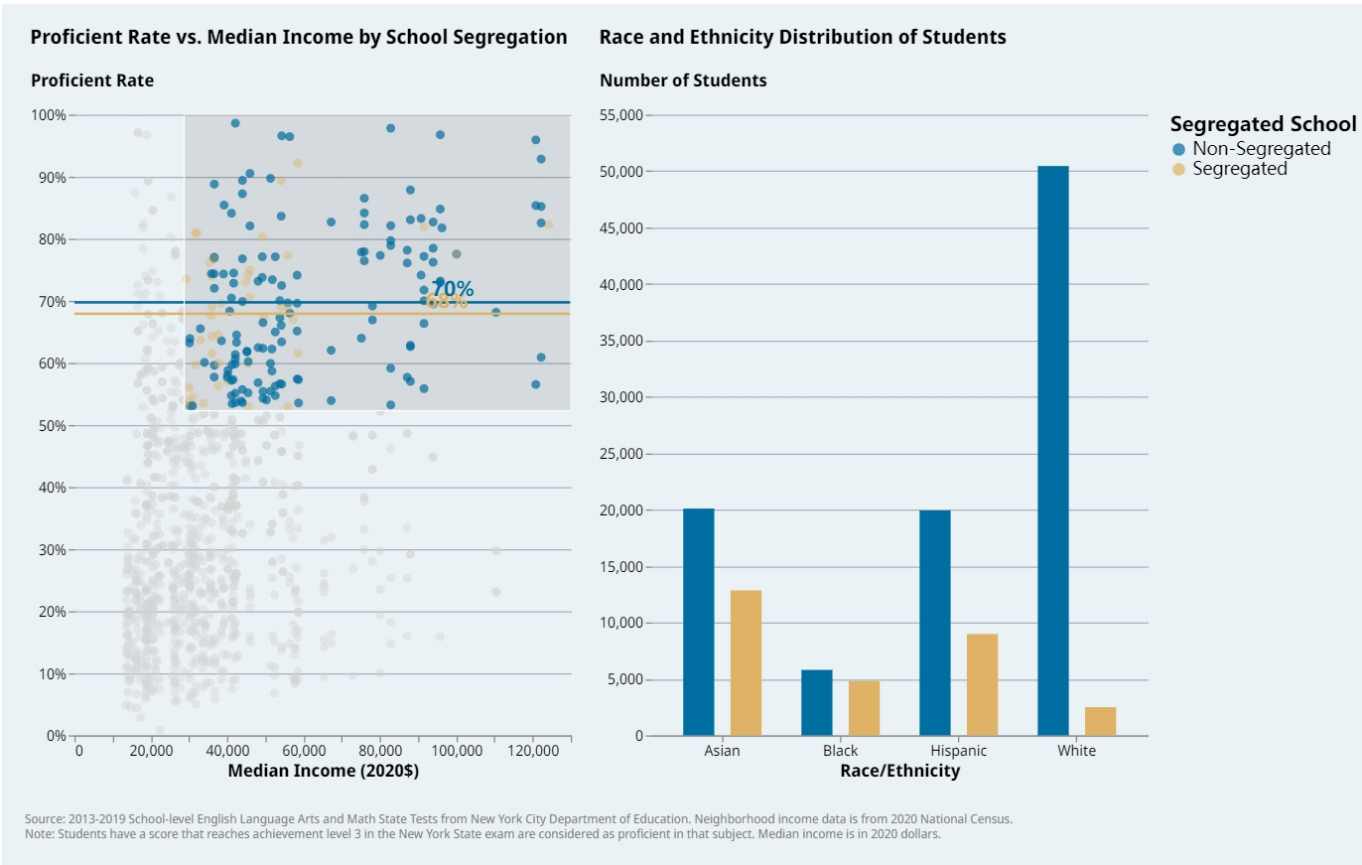


The major revision is to add a race distribution to disentangle the dichotomous label of segregated and non-segregated school.

The left part mainly reused the static version of scatter plot. But I added a toggle to let user choose between ELA or Math proficiency data. And the threshold lines are changed to showing the average proficiency rate for segregated and non-segregated schools respectively. The right part shows the racial distribution in segregated and non-segregated schools which are selected. Users can brush to select the schools they are interested in and then the bar chart will be automatically updated.



More Hispanic and Black students in low-performing schools.



White students are over-represented in high-performing schools. But also more Asian and Hispanic enrollment in non-segregated schools.

Here I used two examples to show how the interactive version works. If we select the schools in the lower-left region: lower-income and lower-performing, we found much larger population for Hispanic and Black students. In the upper graph, I brushed the schools in the upper-right region: high-income and high-

performing. We found that White students are clearly over-represented. But we also found there are about relatively large Asian and Hispanic population in those schools. And for those high performing schools, there is a larger Asian and Hispanic population in non-segregated schools.

It resonates with the argument that underrepresented students may perform better in non-segregated schools than in segregated ones. But it also expand the story: there are also a relatively large Asian and Hispanic enrollment in high-performing and non-segregated schools compared to Black students. It is kind of aligned with the previous example in the NYC map: Most public schools in Washington Heights, a neighborhood with large Hispanic population, are segregated though, the schools perform not in a bad way.

Conclusion

The designs are embedded in the jupyter notebook. I enjoyed the process of thinking and designing these graphs. It is very related to my study in education policy analysis. I am looking forward to learning more your insights.