Checkpoint3: Interactive Visualization with D3.js

Team members:
Hong Hong (NetID: hha7337)
Hui Ye (NetID:hyr1096)

Introduction

Our group attempts to conduct research on the diversity of police officers and discuss the relationship between policing diversity and misconduct rate, including race and gender representativeness. In this checkpoint, our group proposes two interactive visualizations to gain a deeper understanding of the relationship between the diversity of police officers and misconduct rate.

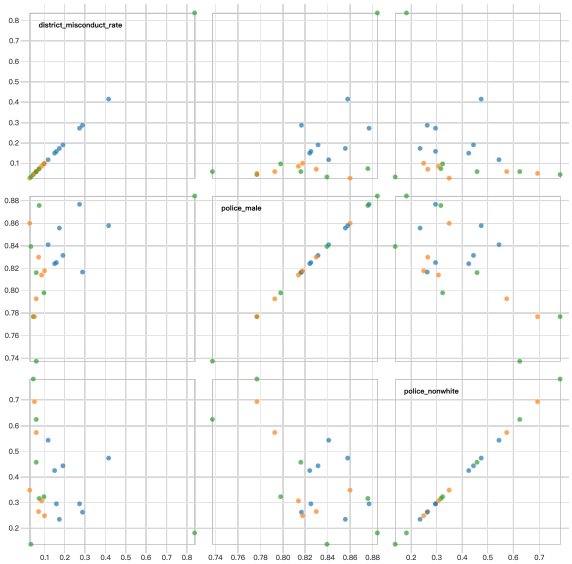
For the level definition in this analysis, you can refer to the data definition in our checkpoint1. In checkpoint1, we defined both the level of misconduct and the level of diversity by calculation and data segmentation.

Visualization 1: A scatter plot matrix shows the relationship between the proportion of male and nonwhite police officers, as the representation of race and gender, and police misconduct rate in each district.
Visualization 2: A stacked to group chart shows both trends and distribution relation among the police officers within a certain year period in the certain district that we have defined in checkpoint 1.

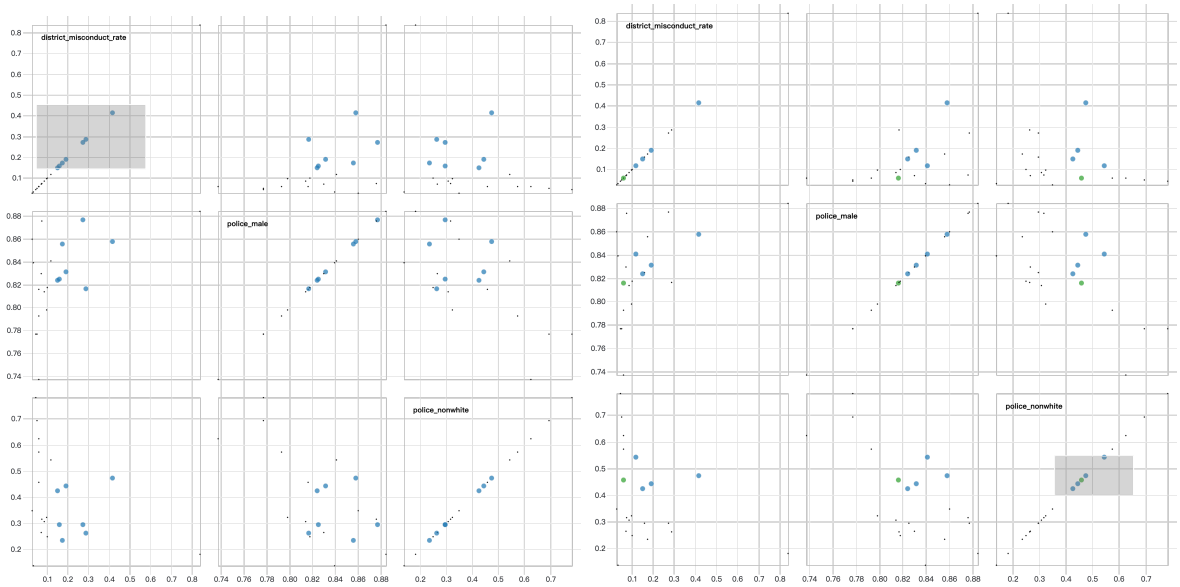
Our visualization and detailed analysis are provided below, then we provide a conclusion on our findings and raise some open questions.

Visualization 1

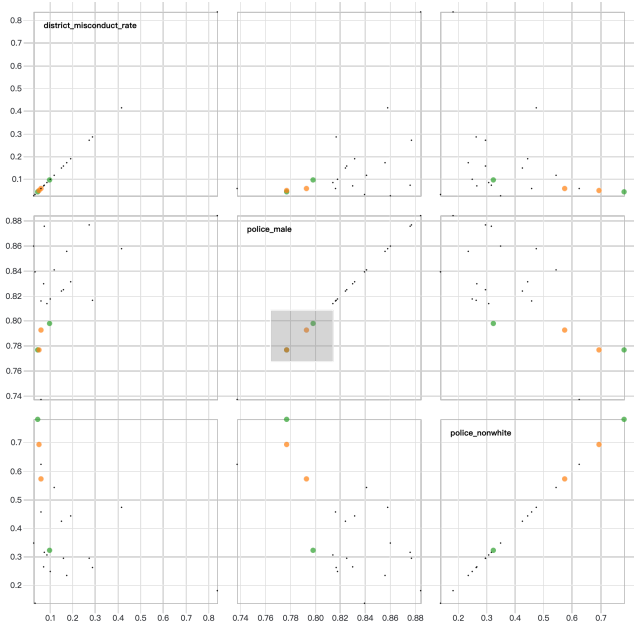
External link to interact with our visualization:
https://observablehq.com/@honghong1012/checkpoint-3-interactive-visualization-with-d3-js-1



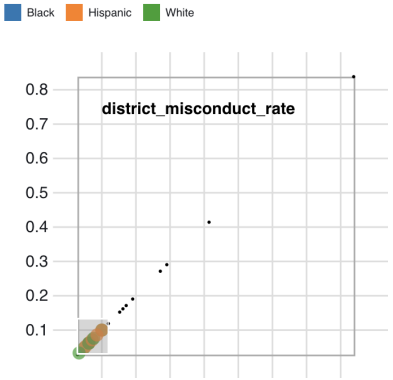
The above is the overview of our visualization 1, where we combine the three column vectors 'district_misconduct_rate', 'police_male', and 'police_nonwhite' and draw the brushable scatterplot matrix.

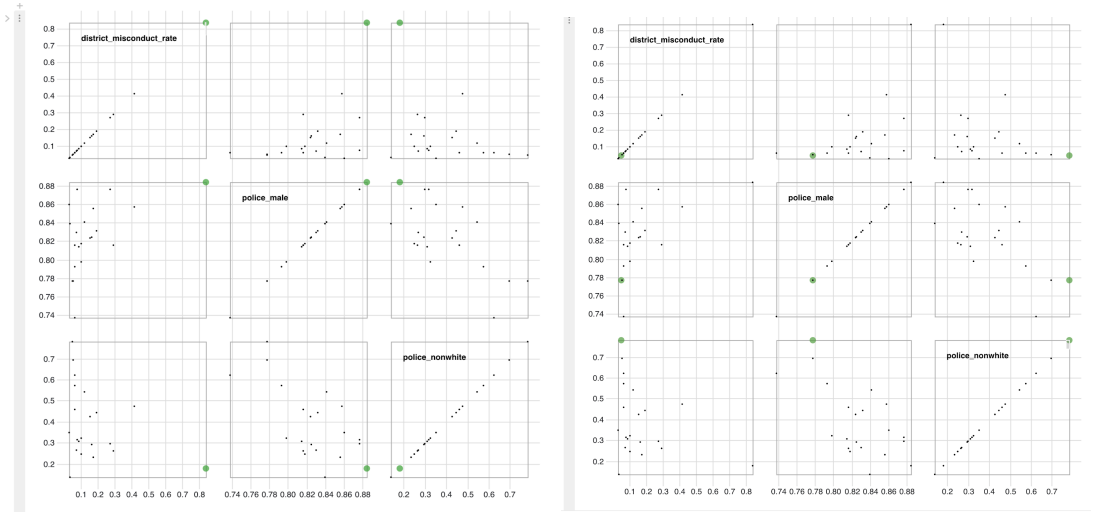


When we brush on the matrix, we are able to see that more of the districts that are misconducted are black-dominated districts, and in these districts, most of the police officers serving are male, with the proportion of male police officers above 82%, while the proportion of non-white police officers is distributed in the intervals of 40%-50% and 20%-30%. When we select the districts that belong to medium diversity policing, we are able to correspondingly find that the misconduct rate of these districts is relatively low compared to that of the low diversity policing districts, which is basically low-level of misconduct rate or medium-level of misconduct rate.



When we attempted to examine the relationship between gender and level of misconduct rate, our selection was shown in the figure above. What is surprising is that when the gender diversity in the police composition is greater, the non-white composition is also relatively greater. At the same time, we were able to find that the selected districts are predominantly White or Hispanic neighborhoods. All of these districts are in the low misconduct level range (less than 11%). This supports our hypothesis that when the gender diversity of police officers is greater, the misconduct rate of their districts is relatively smaller. Of course, here we also need to consider the residential population of the districts, in which the main residential population may also affect the misconduct rate of police officers.



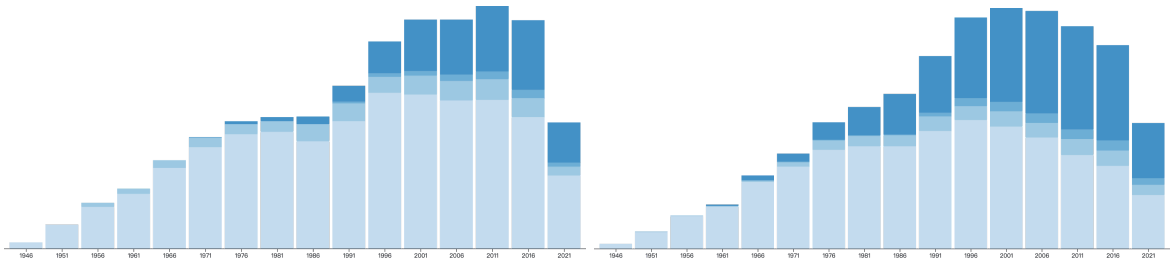


We can read from the top graph that the white dominated districts usually have a low police misconduct rate. However, we find a special case that a white-dominated district has a highest police misconduct rate (as the green circle selected in the left graph shown above). This district also shows the highest percentage of male police officers with 88% and a low non-white police officers rate with no more than 20%. At the same time, we also can see from the right graph that there is a white district with a pretty low police misconduct rate. Looking into the police diversity, the police in this district has relatively low male proportion and has the highest nonwhite percentage. The comparison of these two cases shows that there is a possibility that the male-dominant and while-dominant police group has a high possibility to have a misconduct, while the police with diversity on race and gender tend to have a lower misconduct rate.

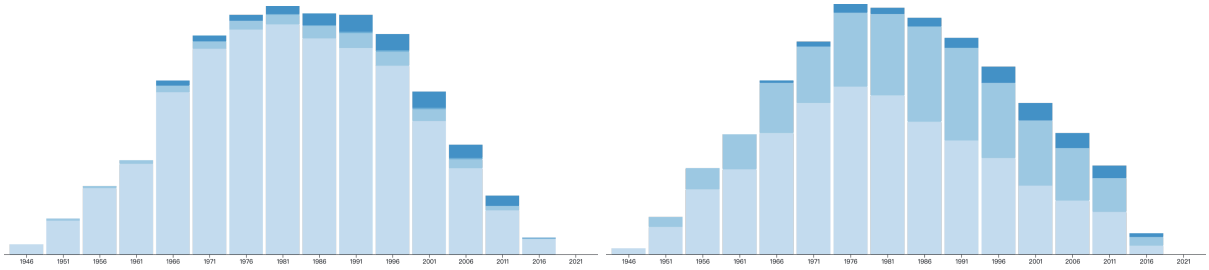
Visualization 2

External link to interact with our visualization:
https://observablehq.com/@honghong1012/checkpoint-3-interactive-visualization-with-d3-js-2

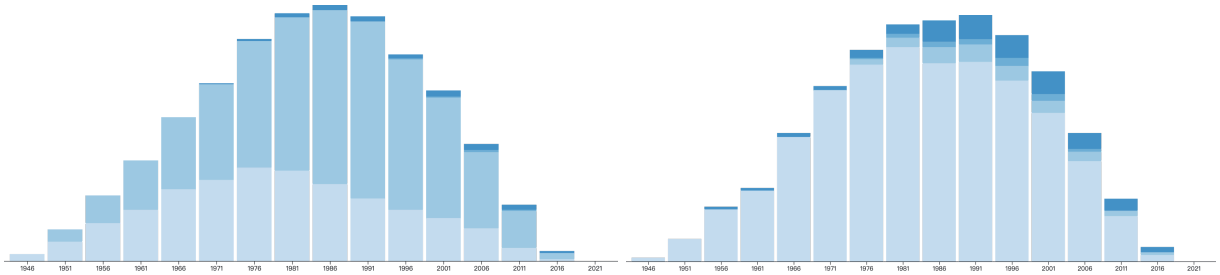
Since there is no way to map the visualization functions and the legend correspondence one by one in code, here we illustrate the meaning of the different color partitions of the bars. From bottom to top, the four colors represent White, Black, Asian and Hispanic police from light to dark.



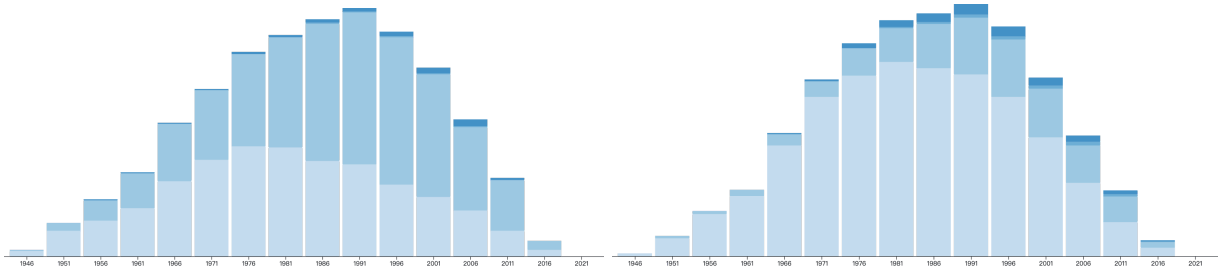
Black-dominant district 9 and 14



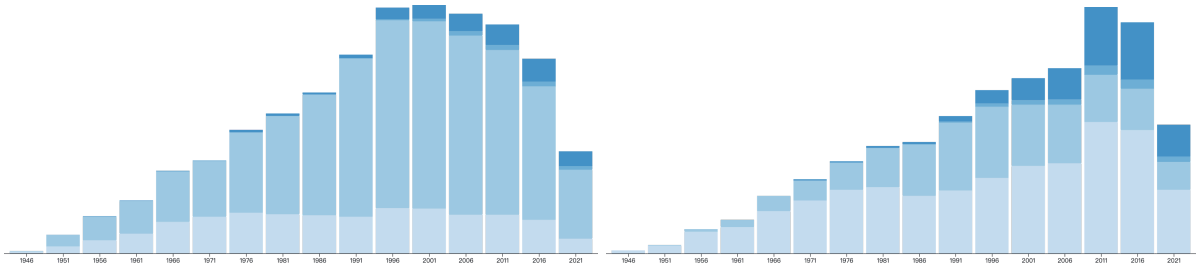
Black-dominant district 8 and 11



Hispanic-dominant district 6 and 24



White-dominant district 5 and 18

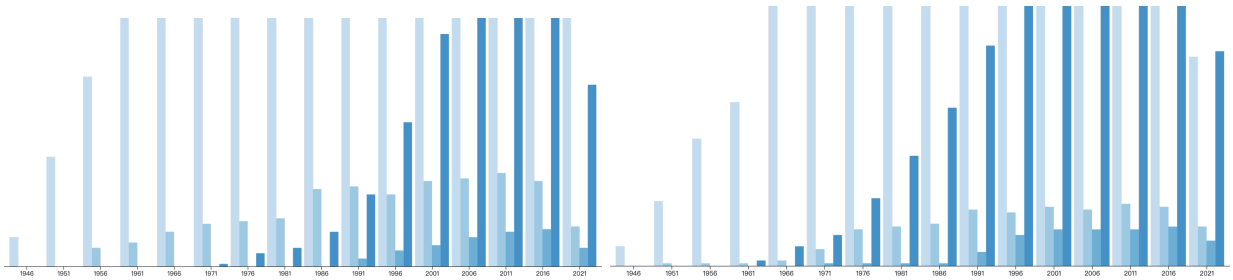


District 2 and 15 (Asian population relatively higher)

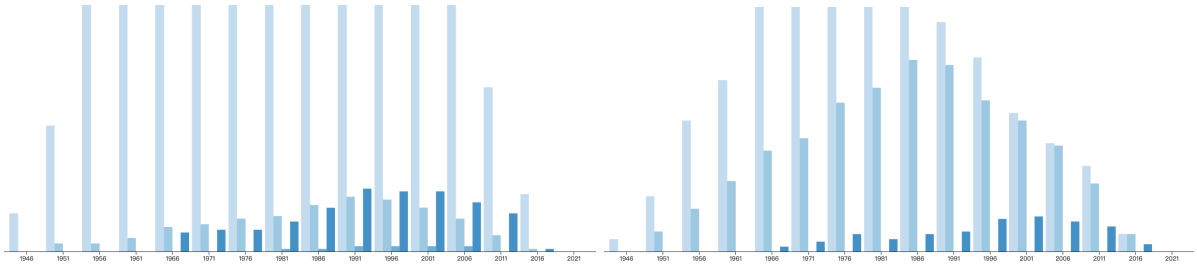
We can first observe the bar chart in a stacked method , and according to the classification of checkpoint1, we have visualized the data of different pairs of districts. We divided the years of service of the police officers according to the collected data, in which we set an interval of 5 years and calculated the number of police officers in each year by calculating their assigned and resigned times. (If the resigned time field is empty, we consider them to be still in service) Looking at the overall ratio, we can see that for districts with similar dominant race composition, the ratio of police officers actually varies, and we cannot see a specific allocation ratio from this. However, we can see that in our visualization of these 10 districts, white and black police officers are in the majority. And compared to white police officers and black police officers, the number of hispanic police officers is very small, and the number of asian police officers is close to zero.

Secondly, we can observe the overall trend. According to the data collected, we can see that the trend of the number of police officers rising and falling is roughly similar. Especially for the similar districts, we can observe that the trend of the bar chart is very similar. For both district 9 with 14 and district 2 with 15, the total number of police officers is showing an increasing trend. For the other districts, the total number of police officers starts to rise in 1946 and peaks in 1980-1990, and then declines. We attribute this late decline to two possible reasons, one is from database problems, such as recent police data not being updated. The second one could be the reason of data cleaning.

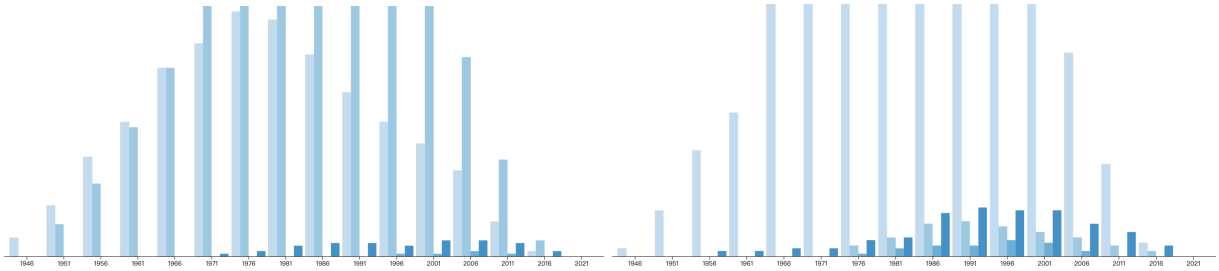
According to different districts, districts with predominantly black populations are generally dominated by white police officers. For example, in districts 2 and 15, where asian residents have a relatively high proportion of police officers, black police officers make up the majority of police officers, which can be said to be a high diversity district.



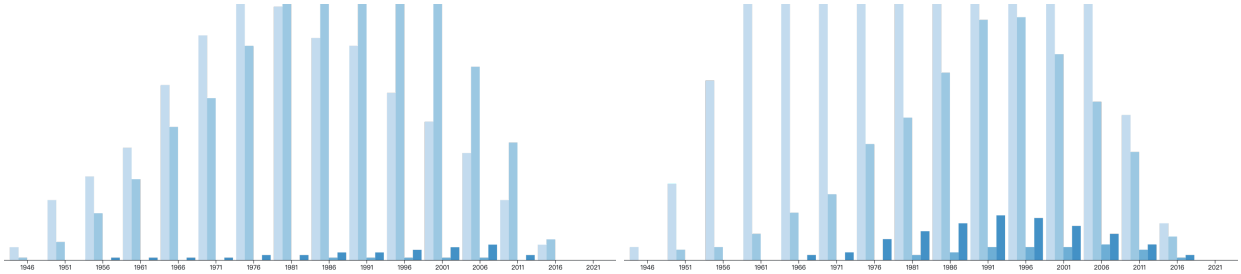
Black-dominant district 9 and 14



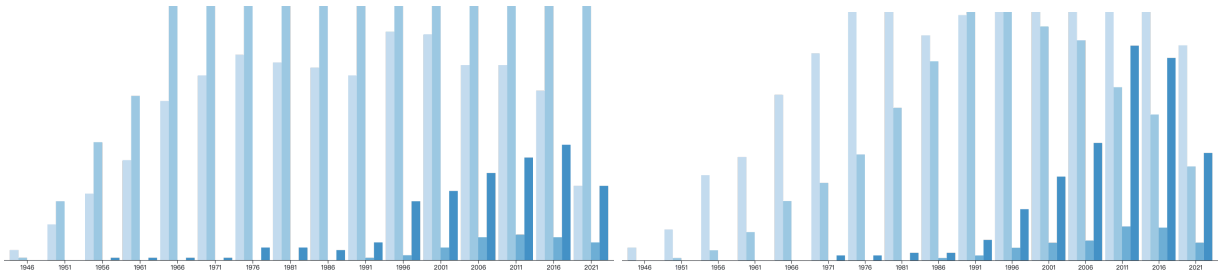
Black-dominant district 8 and 11



Hispanic-dominant district 6 and 24



White-dominant district 5 and 18



District 2 and 15 (Asian population relatively higher)

Finally, we can observe the bar graph in a grouped method, (four groups of graphs from left to right are white, black, asian, and hispanic). It is reassuring to see that the diversity of police ratios has mostly increased over time. In the black-dominant district, the ratio of black police officers is also increasing year by year. Accordingly, the number of white police officers remains a stable number in most of the years, which may be a side indication that the police department will consciously increase the diversity of police officers when recruiting new officers or assigning officers to the corresponding district on duty, which is a positive phenomenon.

Conclusion

Overall, the percentage and metric is defined according to the data, the result may be a little less reliable, but we would say that the findings by interactive visualization somewhat support the question or give us interest to explore in this direction.

Specifically, from visualization 1, we could tell from the interaction with the graph that the predominantly black districts had the highest level of misconduct rate on average. And looking at the racial and gender diversity in them, the diversity level of these districts is at low level or at medium level, reflecting to some extent the relationship between the misconduct rate and the diversities of the police officers on duty.
From visualization 2, we could tell from the graph that the composition of white police officers is still the largest and has remained stable. However, it also reveals a positive phenomenon that the racial diversity of police ratios is gradually increasing, which may reflect a shift in the police department's strategy.

The bar chart gives us a more visual representation of diversity, unlike the checkpoint1 where only the data is calculated and partitioned, in the visualization we can more directly feel the difference in diversity and the level of diversity in different district police.

In fact, there are many open questions to explore from this visualization. That is, for those districts where police ration diversity is increasing, how does the misconduct rate or level change in each year interval? And what are the trends? Is it possible that as the diversity of police officers in a community increases, its misconduct rate can decrease?

To dive deeper into the theme we want to conduct experiments on, more work needs to be done.