Checkpoint 2 - Interactive Visualizations

Team Name: The Wild Blazers

Team Members: Anish Bhardwaj

Visualizations Chosen

1. Gender Distribution of Police Officers per District
2. Population Race Distribution vs Police Race Distribution per district
3. Population to Police Officer Ratio and Allegation Rate per police officer per district (Combined Interactive Visualisation)

SQL

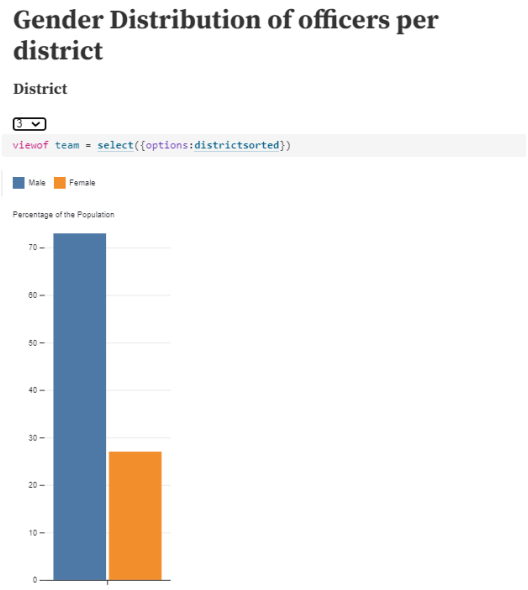
I modified the SQL queries and used the data I got from them to create the CSV file (d10re_1.csv in CSVs for Visualization) that I used for the visualizations seen below. The SQL code is available in the SQL folder on github. The ipynb present (Checkpoint 2 supplementary) is how I cleaned and modified the CSVs present in the RawCSVs_foripynb folder.

Latest Update: I fixed the data that I was sourcing the visualizations from after advice from Prof. Rogers.

Visualization 1: Gender Distribution of Police Officers per District

Link: https://observablehq.com/d/12675256968ae250

For this Visualization, I was attempting to first get comfortable with d3.js on observable and wanted to create a visualization where everything was ordered by district. Given the disparity I saw when performing the relational analytics as well, my plan was to graph out the Gender distribution of the police officers as well so that I could better showcase the disparity that I had observed. If the link is opened, we can see a graph like this one below.

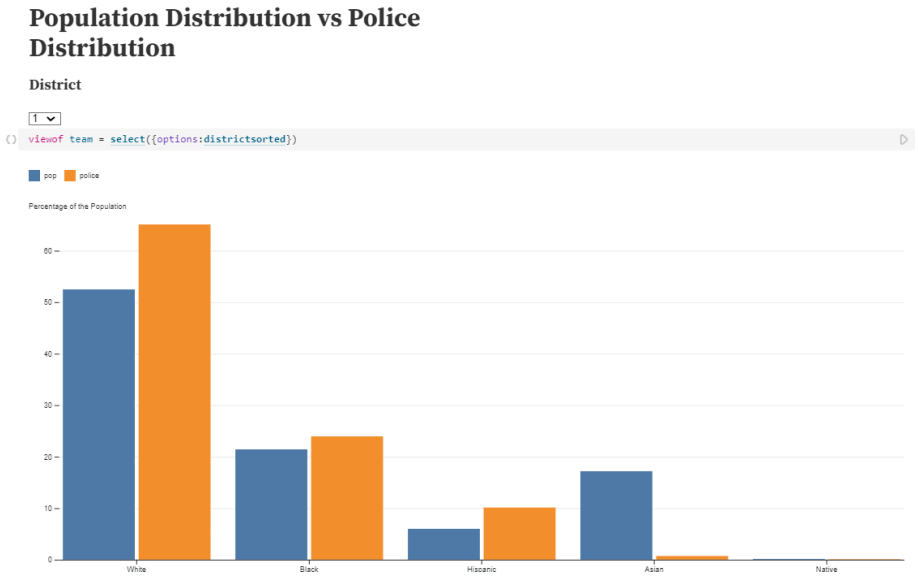


As can be seen, there is a dropdown menu wherein we can select the District (now numbered correctly for the corresponding Chicago Police district) and changing the district would show the values for each new district. While this was a practice graph, I think this was an important one as it allowed me to properly take a look at the sheer discrepancy and male dominated police statistics. The Male police outnumber female officers on a near 4 to 1 ratio in most districts. This graph was a great representation that also showed how important it is to visualize values as while the numbers were still high, it doesn't showcase the values as well as a graph.

Visualization 2: Population Race Distribution vs Police Race Distribution per district

Link: https://observablehq.com/d/03d15b41d6bb2884

The objective of this question was to see whether the officer demographics in a district are similar to that of the population they serve. To do this, I created a bar graph which plotted the different racial groups and the population percentage for these groups for both the police and residents of a district. An example of the graph can be seen below:



As can be seen in this graph, there is a dropdown to select the district which allows us to iterate over the districts and see whether the population ratios match across districts. In this example, we can see that with respect to the population, there is a large number of Asian/Pacific Islander population with respect to the police officers in this area. Even so, this district is relatively similar in terms of the distribution of Police and Population by race.

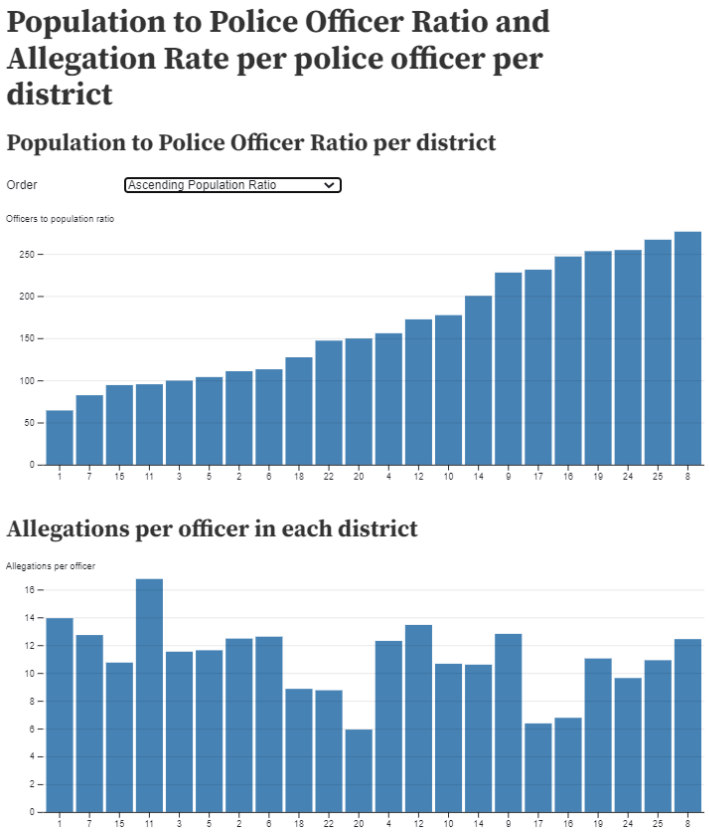
Some examples where we see large discrepancies are District 11 and District 15. Both these Districts have a large black population but the police officers distribution is skewed White with an extremely large number of white officers as compared to their ratio in the population. Correspondingly, if we take a look at the last visualization or the data itself, we see that District 11 has the largest number of allegations per police officer as well which may showcase a potential correlation for racial bias in a district like this.

By visualizing the distribution of race across the police and the districts they serve, we can not only draw conclusions regarding whether the officers are representative of the community but also draw inferences regarding whether there is a potential for racial bias after comparing the distribution with allegation rates in that district.

Visualization 3: Population to Police Officer Ratio and Allegation Rate per police officer per district (Combined Interactive Visualisation)

Link: https://observablehq.com/d/9639fa4f48c4de3d

This was my favorite visualization of the 3 as not only was the interaction level higher but it brought up a very interesting point. The purpose of this visualization was to compare the Ratio of the Population to Police Officers to the Allegation Rate of Officers in a District. The reason this was an important visualization is because when trying to understand whether police officers may be biased or require sensitivity training based on their allegation rate or potential risk, it is important to first check whether the Allegation Rate corresponds to an increase in the population being served. This is because there is a chance that the police officers in a district with a higher population may be overwhelmed or may just potentially be subject to more allegations as a result of having to serve a much larger population. A still of the graph can be seen below.



This is definitely a big graph but as can be seen there is a dropdown menu which allows a viewer to select the order of how they want to see this. Here, I have set it to an ascending population ratio which means that as we go from left to right in this graph, the number of officers relative to the population decrease. As can be seen, there is no linear correlation between these 2 graphs and adjusting it in other ways shows the same thing. This showcases that the police officers serving a district are not getting more allegations as a result of an increased population and points to the presence of an underlying bias. This visualization was extremely helpful and coupled with the racial distribution across the districts showcases clear progress towards the goal.

Learnings from Visualizations

Other than learning how to create visualizations in d3 and what has been summed up above, the visualizations I created up above do have some interesting insights.

Key Questions

- *What are the key takeaways from this analysis?*
- *How can we relate the demographic makeup of these police units with the complaints registered against their officers?*

Visualization 1 - Gender distribution

The gender distribution graph for each district shows us that there is a clear disparity in the gender distribution of officers in a district. If we comb through each district's values, we can see that most times the male officers outnumber female officers in a nearly 4 to 1 disparity. This opens the floor to further questions in researching this data that could be conducted which could relate to whether based on population demographics of police officers in a district, do districts with better gender and race distribution amongst the officers lead to a lower allegation rate per police officer? This first visualization was very good for me to also get familiar with the d3.js environment and as a result of it, I was able to easily navigate the code and mechanisms for my other graphs.

Visualization 2 - Population vs Police Racial Distribution

The racial distribution of police officers vs the districts they serve was a very interesting graph. In this, we can see the clear distribution of officers and district demographics as well as where disparities exist. If we take a look at a district like district 11, (which has the highest number of allegations per officer as seen in the 3rd visualization), we see that there is an overwhelmingly large number of white officers in this district. The district is also majority black in terms of population and the ratios of the police officers to the district does not match. This showcases a potential of racial bias in the district. The converse can be seen in district 20 which has the lowest allegations per police officer. The distribution of police officers in this district is closer to the distribution of the population. The district also has a white majority in the civilians. It is data like this that shows that while principles are in place to reduce bias, there is still a potential bias evident in the districts across Chicago

Visualization 3 - Population to Police Officer Ratio and Allegation Rate per police officer per district

This visualization is a very important one that was used to work against the hypothesis that the allegation rate is greater when there are more police officers in the district. In this interactive visualization, users can take from multiple different sorting options (Ascending and Descending Police per capita, Ascending and Descending Allegations per police officer in a district, and District wise) to see whether this hypothesis is true. Based on the visualization, it is evident that there is no clear correlation between the ratio of police officers to the citizens in the district and the Allegation rate per police officer in the district. This visualization, in conjunction with the 2nd visualization, also allows us to see how officer and district racial demographics seem to potentially play a role in the allegations made.

Final Thoughts

These visualizations were extremely fun to make and helpful in showcasing that there is a potential for bias amongst police officers and the districts they serve. By going through these visualizations, I feel more confident in creating ML based models that can help determine if based on an officer's information and allegations that are coming in we should recommend officers for sensitivity training.