

Checkpoint 2: Data Visualization

Team: The Powerful Turtles

Introduction

In this project, we would like to study how the demographics comparison between the complainants and the respective police officers reported correlating to the number of complaints. In this checkpoint, we perform data visualization of race distributions among the officers and the complainants using Tableau. The results from this checkpoint provide the foundation for the interactive data visualization tasks in the next checkpoint, where the cross-correlation between the officer race and complainant race with respect to the community race composition. Together, these visualizations will provide guidance for future checkpoints.

Relational Analytics Questions

In this checkpoint, we would like to address the following questions:

1. *A pie chart showing the race distribution among the officers and the allegation counts within each race distribution.*
2. *A bar chart showing the race distribution among the complainants*

Results

1. ***A pie chart showing the race distribution among the officers and the allegation counts within each race distribution.***

- ❖ A pie chart showing the race distribution among the officers
 - We first drag the **data_officer** table to the data source and check if there are any strange values using filters.

Enter search text

☐

Amer Ind/Alaskan Native

☐

Asian/Pacific

☐

Asian/Pacific Islander

☐

Black

☐

Hispanic

☐

Native American/Alaskan Native

☐

Unknown

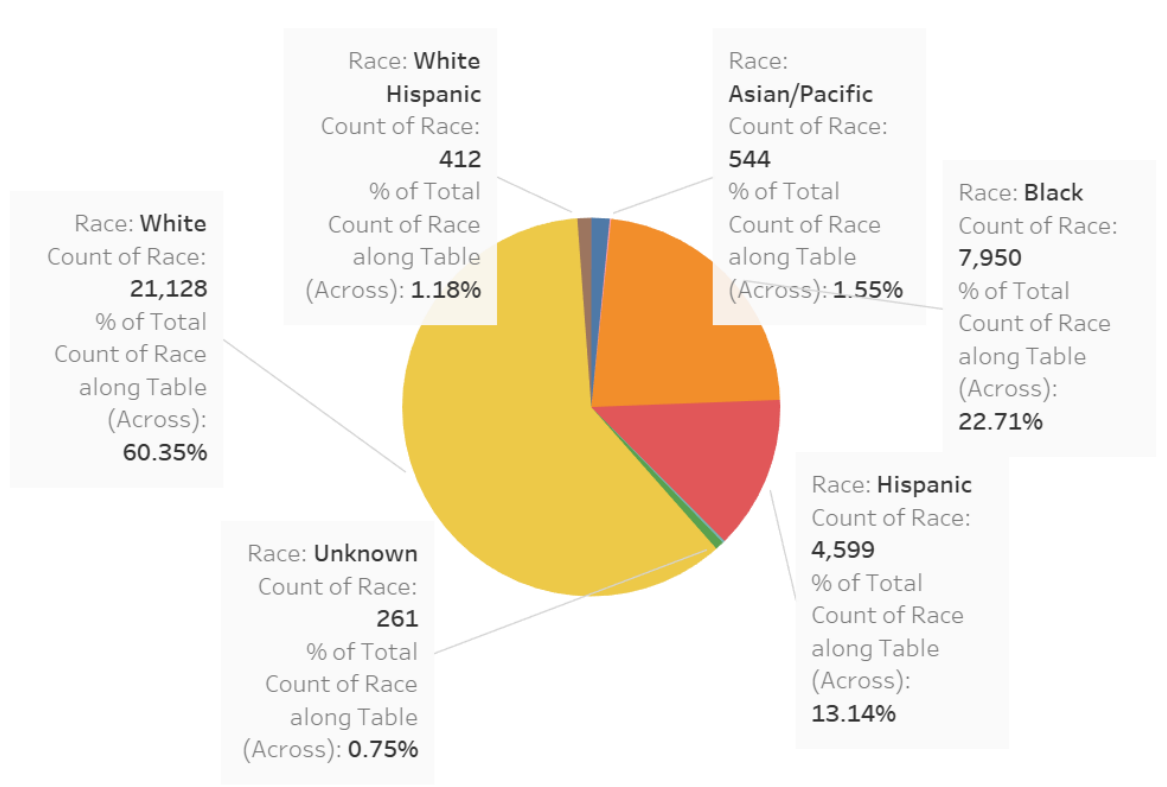
☐

White

☐

White Hispanic

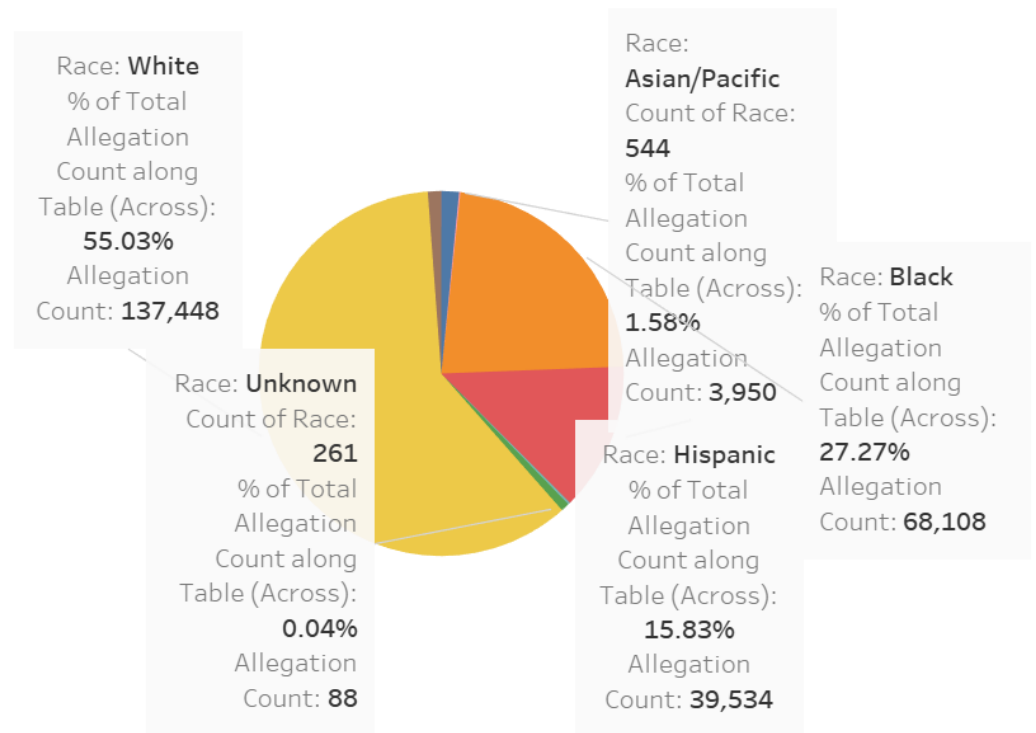
- Then, we drag **Race** as the rows and **Count (Race)** as the columns. Choose Pie chart to visualize. Add labels and marks to make it clear to the audience.



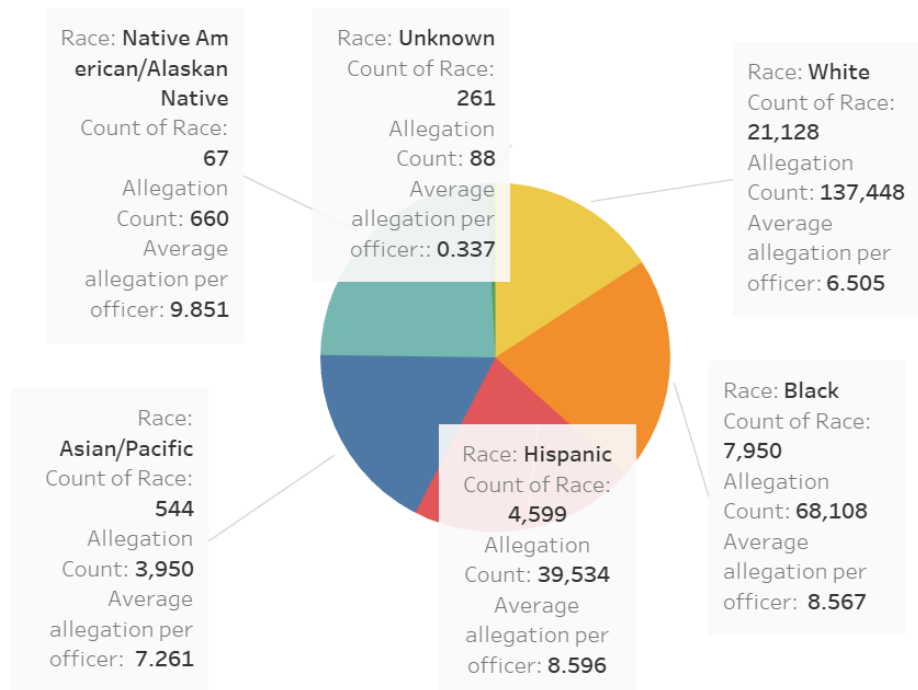
Analysis: From the pie chart, we can intuitively see the race distribution among the officers. Among the officers, more than half are White, followed by Black, Hispanic, Asian/Pacific, White Hispanic, and Unknown. The number of White officers is more than 2.5 times the number of Black officers and more than 4.5 times the number of Hispanic officers. White, Black, and Hispanic officers make up more than 95% of the officers. The results show a very unbalanced distribution among officer races, which implies an interesting potential problem that we need to pay attention to.

- Then we want to see how many allegations in each group, so we drag **Allegation Count** into the details, and also the pie chart of allegation count divided by race count is calculated.

The race distribution and allegation counts among officers - pie charts



Average allegation per officer

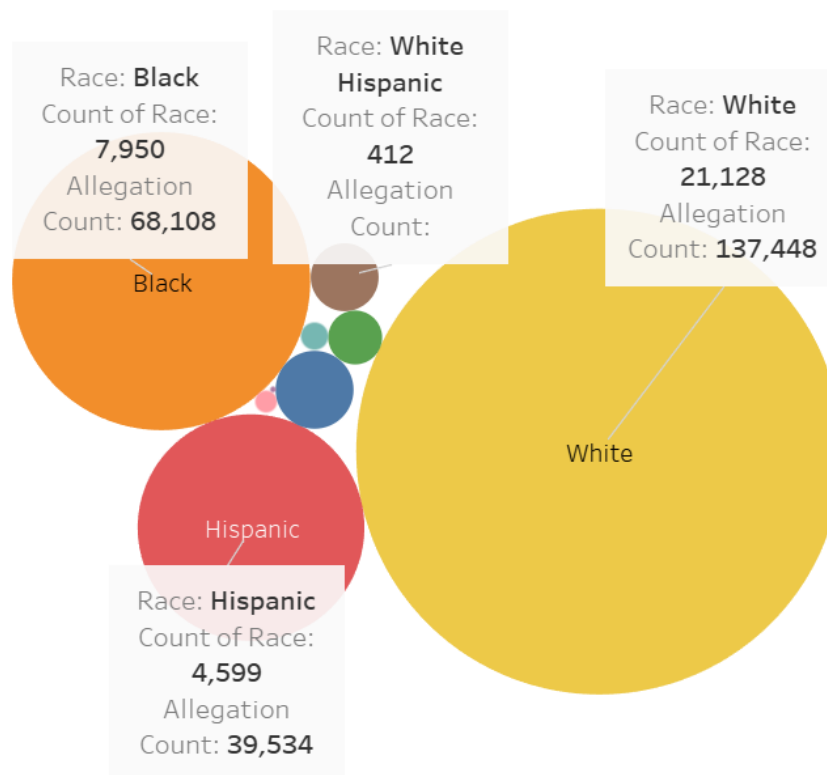


Analysis: From the allegation counts of each group, we can see that white officers have the most number of allegations, followed by Black, Hispanic, Asian/Pacific, and Unknowns. Note that White Hispanic officers have no allegations.

Although white officers have most allegations, due to their high officer number, their per capita allegation number is actually less than all other races (omitting the allegations from the category in unknown races). On the other hand, Hispanic officers have higher per capita allegation than other races, and Black officers follow closely after. This leads us to hypothesize that officer race is a crucial factor in predicting officer misconduct using machine learning models.

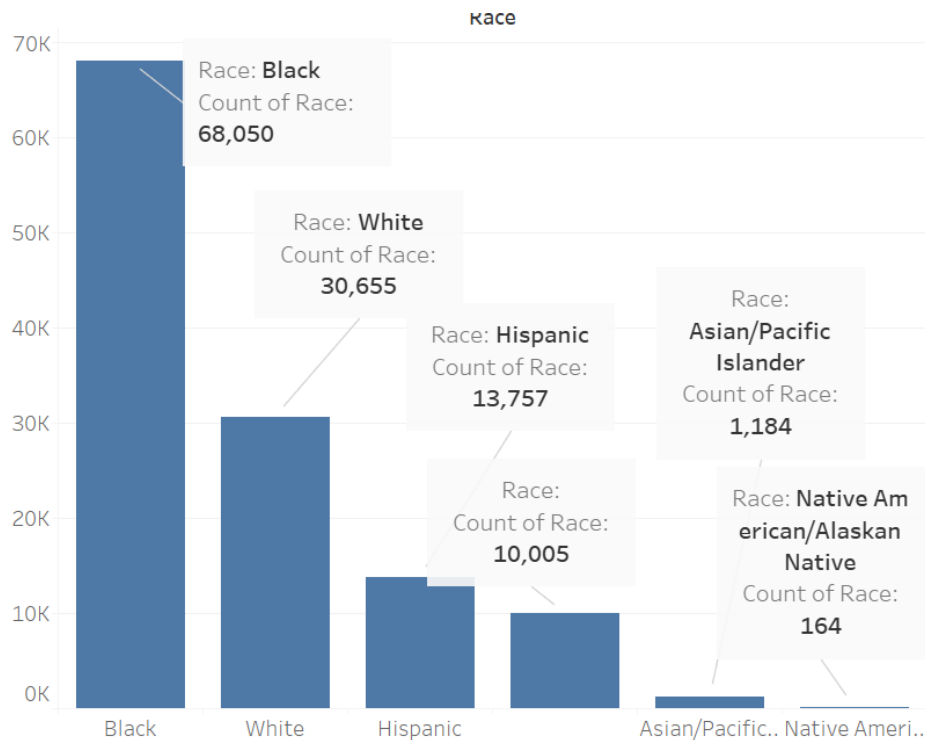
About Tableau:

It is really easy to have good visualizations in Tableau. You can choose all kinds of charts to show your results. Any settings and details are easy to add on to the charts. For example, we can change quickly from the pie chart to the bubbles shown below.



2. A bar chart showing the race distribution among the complainant

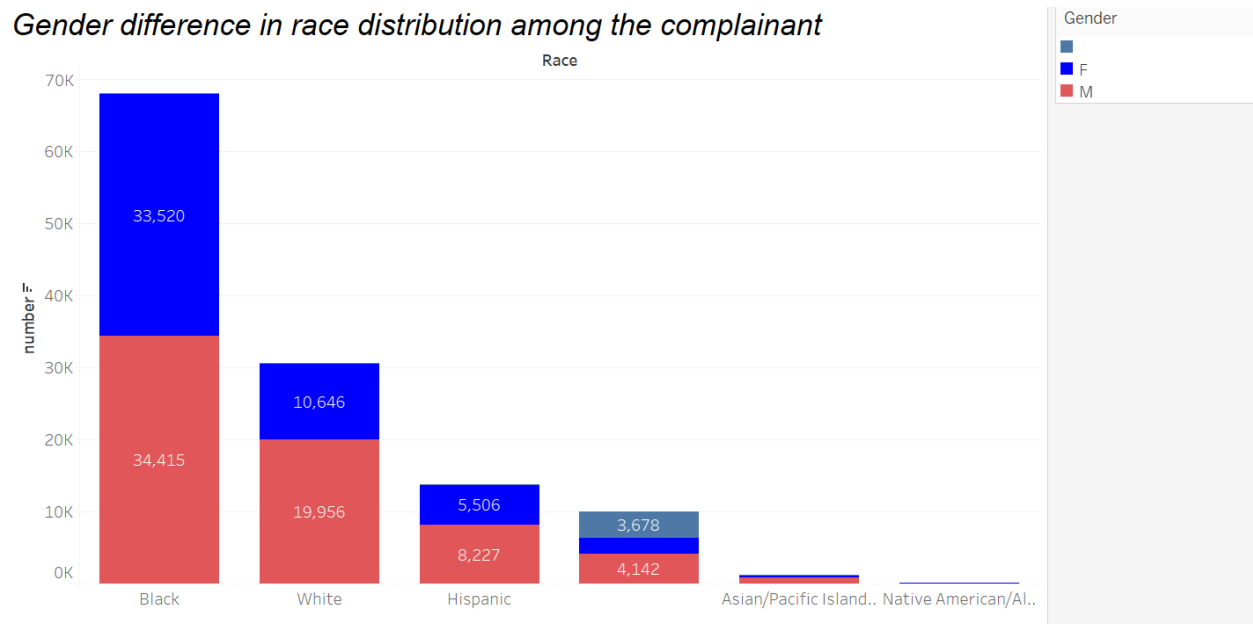
- We first drag the **data_complainant** table to the data source and check if there are any strange values using filters.
- Then, we drag **Race** as the rows and **Count (Race)** as the columns. Choose a horizontal bar chart to visualize. Add labels and marks to make it clear to the audience.



Analysis: From the complainant distribution of the table, we can see that about two-thirds of the complainants are black, followed by White, Hispanic, unknown, Asian/Pacific, and Native American/Alaskan Native. The number of black complainants is more than twice the number of white complainants, four times that of Hispanic complainants, and sixty times that of Asian and Pacific Islander complainants. Based on the results from the previous section that the vast majority of the officers are white or black, and here most allegations are onto black complainants, we can infer that black people are more likely to experience misconduct from the officers, and victim race is a significant factor in predicting officer misconduct.

- Next, we want to see if any gender differences exist in race distribution, so we add the gender into the details and use stacked bars to show the comparisons.

Gender difference in race distribution among the complainant



Analysis: From the table, we can see that there is not much gender difference in the black complainant, but the males are more likely to complain than females among white and Hispanic people. Overall, there are more male complainants than female complainants, indicating that either males are more likely to experience police misconduct, or males are more likely to file complaints against the officers than females. Since the difference in gender is not as significant as race, we hypothesize that victim gender plays a less significant factor in predicting police misconduct.

About Tableau:

Challenges:

- We spent almost 4 hours just trying to connect to PostgreSQL, without even being able to do ANY analysis. The tutorials online are not that helpful. The compatibility of Tableau and PostgreSQL needs to be addressed here.
- Data loading is relatively slow and time-consuming if used without excellent internet speed.

Strengths:

- It is easy to have all kinds of visualizations just by dragging the tables without any programming. It will be a powerful tool to use in future analysis and research.
- The visualization formats (bar, pie, plots, etc.) and designs (color, etc.) are relatively simple to use and modify, and the default settings are already very visually pleasing.