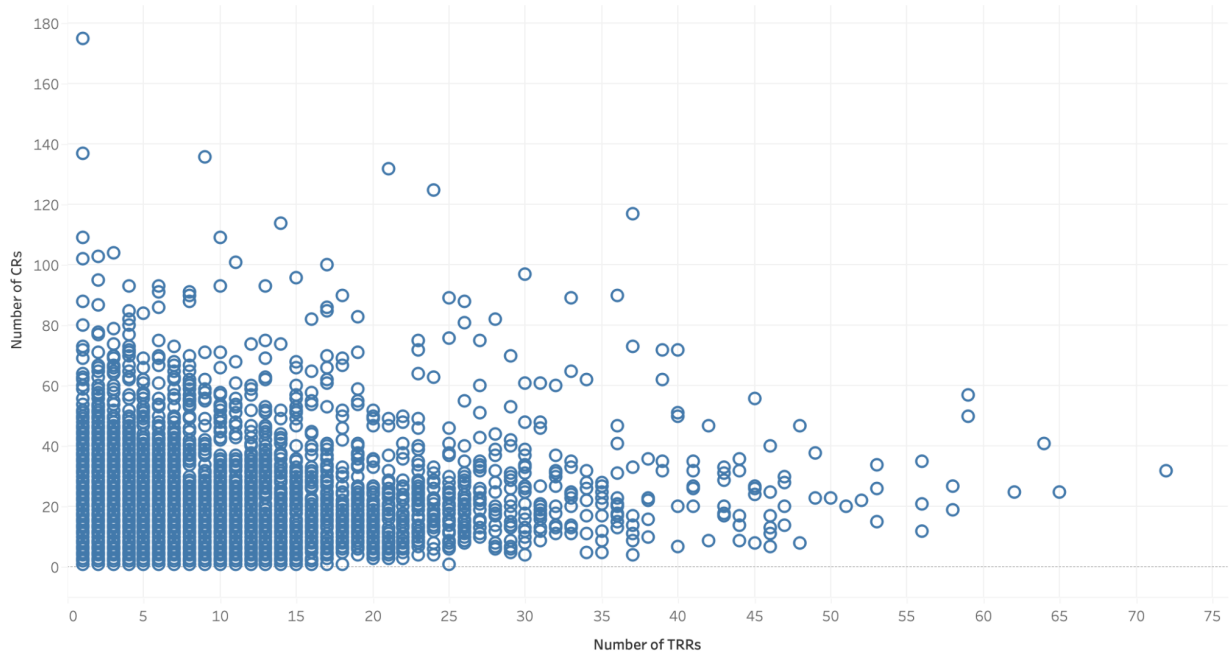
Checkpoint 2: Data Exploration

COMP_SCI 396/496: Data Science Seminar
Meenakshi Kommineni, Madison McClellan, Archana Ramasubramanian

Is there an overlap between officers with a high number of tactical response reports (TRRs) and officers with a high number of complaint reports (CRs) filed against them?

The goal of this visualization was to determine what type of correlation, if any, exists between the number of TRRs and the number of CRs filed against officers. Without looking at the data, we assumed that we would find a positive correlation between these variables. We decided on a scatterplot to visualize the relationship between these variables where each mark represents a single officer based on their unique ID. The x-axis represents the number of TRRs filed against the officer and the y-axis represents the number of CRs against the officer.

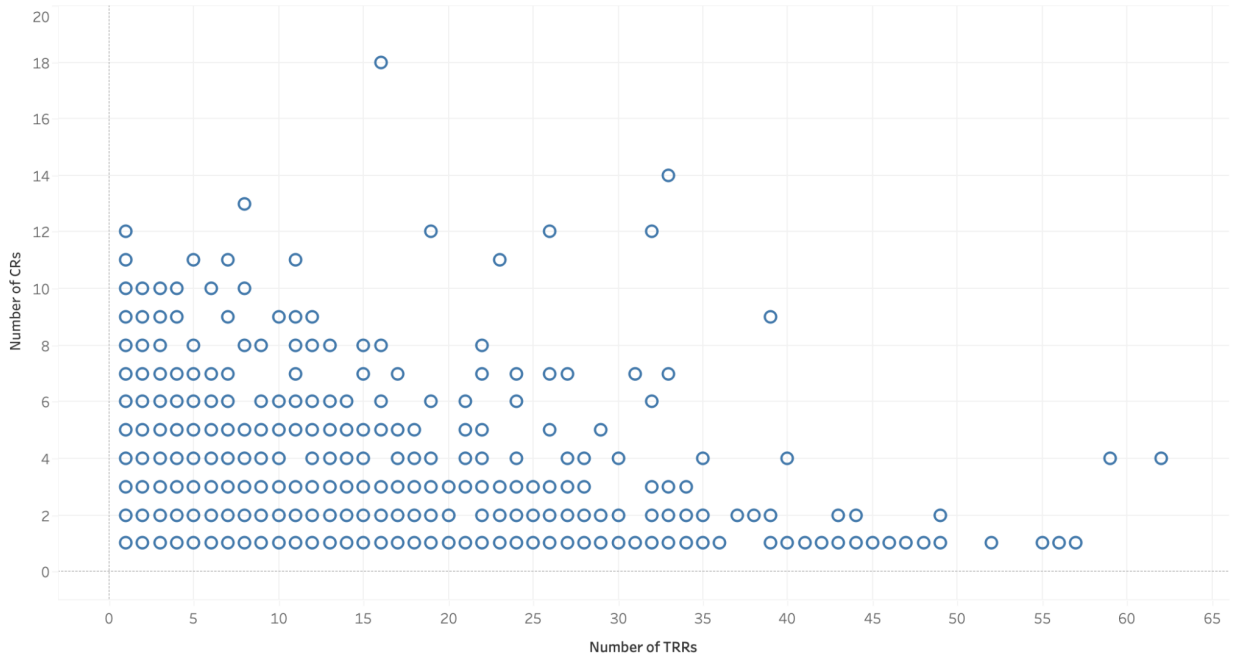
Figure 1.1: Number of CRs vs. Number of TRRs



From Figure 1.1 we can see that there is no correlation between the number of TRRs and the number of CRs filed against officers. Rather, we see that most officers fall within a lower range of both TRRs (< 30) and CRs (< 70), but that we cannot use the number of TRRs to predict the number of CRs or vice versa (i.e. there is no correlation between the variables). This disproves our initial hypothesis that we would find a positive correlation. Although we cannot make any definitive conclusions based on these results, we can infer that the behavioral tendencies or conditions that cause an officer to receive many CRs are different than those that cause an officer to use force excessively.

One potential reason that we did not find any correlation between the two variables is because CRs and TRRs encompass a wide range of officer actions. For example, an officer that receives a CR for engaging in bribery or corruption is not necessarily going to be an officer that also engages in excessive use of force because these behaviors have no overlap. Thus, we thought a more interesting question might be whether we can find stronger correlations between specific types of CRs and specific types of force usage. For example, we assumed that an officer with a high number of CRs in the verbal abuse category would also have a high number of TRRs for verbal commands. However, we actually found that this was not the case. Similar to Figure 1.1, we can see in Figure 1.2 that there is no correlation between officers with a high number of verbal abuse CRs and officers with a high number of TRRs for verbal commands.

Figure 1.2: Number of CRs for Verbal Abuse vs. Number of TRRs for Verbal Commands



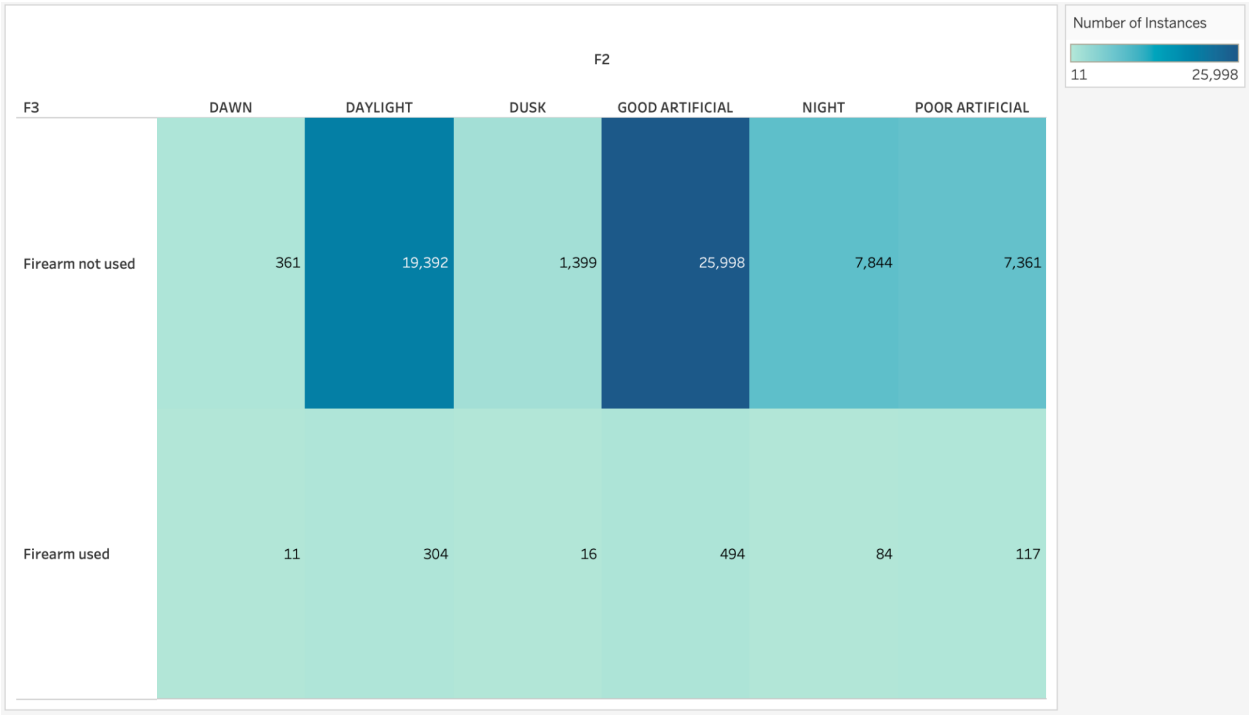
We also explored relationships between other types of CRs and TRRs that we assumed would have a positive correlation, such as CRs for excessive use of force and TRRs. However, we found a lack of correlation for these variables as well. These findings were surprising for us and raised questions not only about officer tendencies and environmental conditions that lead to CRs and TRRs, but also about how the data is recorded. We would like to explore these topics further, but believe it is outside of the scope of the available data and project timeline.

\Is there a correlation between an officer's use of a firearm and the lighting conditions of the incident?

In this question we wanted to explore the conditions under which officers use a firearm. We specifically wanted to examine whether using a firearm was more common under certain lighting conditions. We extracted data from the TRR table, grouping by lighting condition and use of

firearm. Our initial assumption was that we might find more frequent use of firearms under poor lighting conditions or nighttime because it might lead to more ambiguity in the situation. We decided to use a variation on a correlation matrix to investigate this. The rows indicate whether or not a firearm was used and the columns represent the six different types of lighting conditions: dawn, daylight, dusk, good artificial, night, and poor artificial. For each combination of firearm use and lighting condition, we counted the number of instances and labeled the corresponding box in the matrix. We also used a color gradient to visually represent the counts. The resulting matrix is pictured in Figure 2.1. The first conclusion that we can draw from the matrix is that most TRRs do not involve a firearm. Secondly, we can conclude that most TRRs occur under good artificial lighting and daylight, regardless of whether a firearm was involved. This disproves our initial hypothesis that we would find many instances of firearm use under poor lighting or at night. Although we think this type of visualization is appropriate for the data, we realize that the magnitude of the difference between TRRs where a firearm was used and where a firearm was not used makes minor differences in counts difficult to distinguish by looking at the color alone.

Figure 2.1: Correlation Matrix for Firearm Use vs. Lighting Condition



The next question we wanted to explore was whether we would find this to be true for all districts. We extracted the same data, but this time filtered by district. We found approximately the same distribution of TRR instances for each district. We then attempted to identify which districts would have the greatest difference in distribution. Figure 2.2 and Figure 2.3 represent the distribution for the 4th and 20th Districts, respectively. Although these matrices resemble Figure 2.1 for the most part, we can see some notable differences. First, we can see that the 4th District has far more TRR instances in general. We can also see that in the 4th District, a large

proportion of TRR instances without firearm use occurred at night, whereas very few occured at night in the 20th District. Additionally, we can see that while there were 7 instances of use of a firearm at night in the 4th District, there were none in the 20th District. These findings suggest that there is generally a stronger police presence at night in the 4th District. To understand why this phenomenon might be true, we would need to understand more about the nature of these districts, which is outside the scope of this data set.

Figure 2.2: Correlation Matrix for Firearm Use vs. Lighting Condition in the 4th District

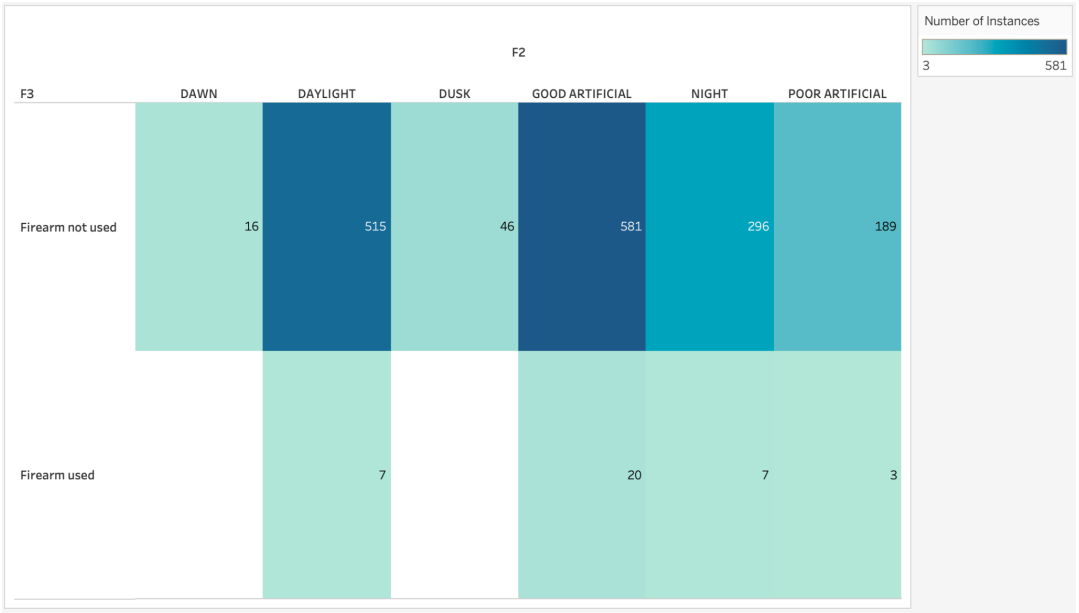
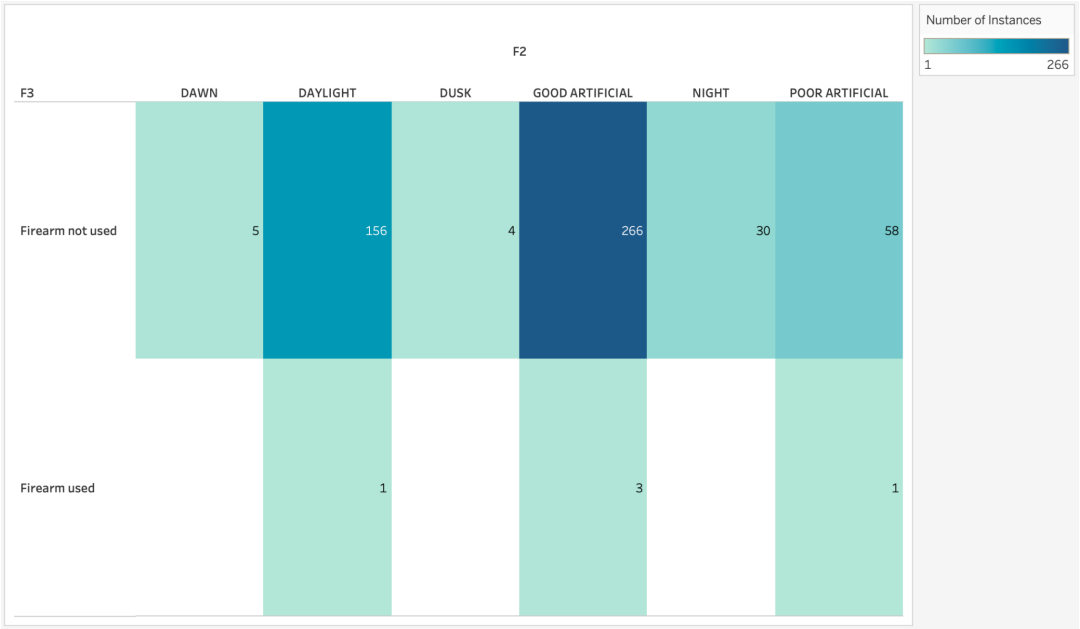


Figure 2.3: Correlation Matrix for Firearm Use vs. Lighting Condition in the 20th District



Our Experience with Tableau

To create these visualizations, we started by writing SQL queries manually in DataGrip to extract the data we needed. We chose this method because we were experiencing difficulty extracting the data in the same way on Tableau via the drag-and-drop functionality. Additionally, to extract the data we needed in an accessible format, we needed to create views then query over those views. Overall, this was much easier to do manually. We then exported the results of the SQL queries as a CSV and imported the file into Tableau. From there, dragging the attributes to the row and column boxes in Tableau then choosing the desired visualization was fairly simple.