**Checkpoint 2 Findings Report**

The Creative Wolves

*Sizheng Zhang, Masum Patel, Jiawei Zhang*

***Theme***: Understanding if there are commonalities between the top 4000 officers who received the highest number of complaints

**Introduction**

In Checkpoint No. 2 we decided to further investigate the commonalities between the top 4000 officers (named as “Repeaters”) who have received the highest number of complaints recorded in the CPDB records. We compared the traits and characteristics of the Repeaters to those of the police officers who have received at least one complaint (named as “Offenders”) and hoped to identify characteristics that are drastically different between the two groups.

In the first Checkpoint we explored several characteristics of the Repeaters such as salary, awards count, and we chose a few interesting traits from the list that we deemed worthy of investigating and performed visualization analysis. By visualizing the data, especially by comparing the results of Repeaters to that of Offenders, we managed to identify some significant commonalities among the officers with high number of complaints and spot key differences between the data of Repeaters to that of Offenders.

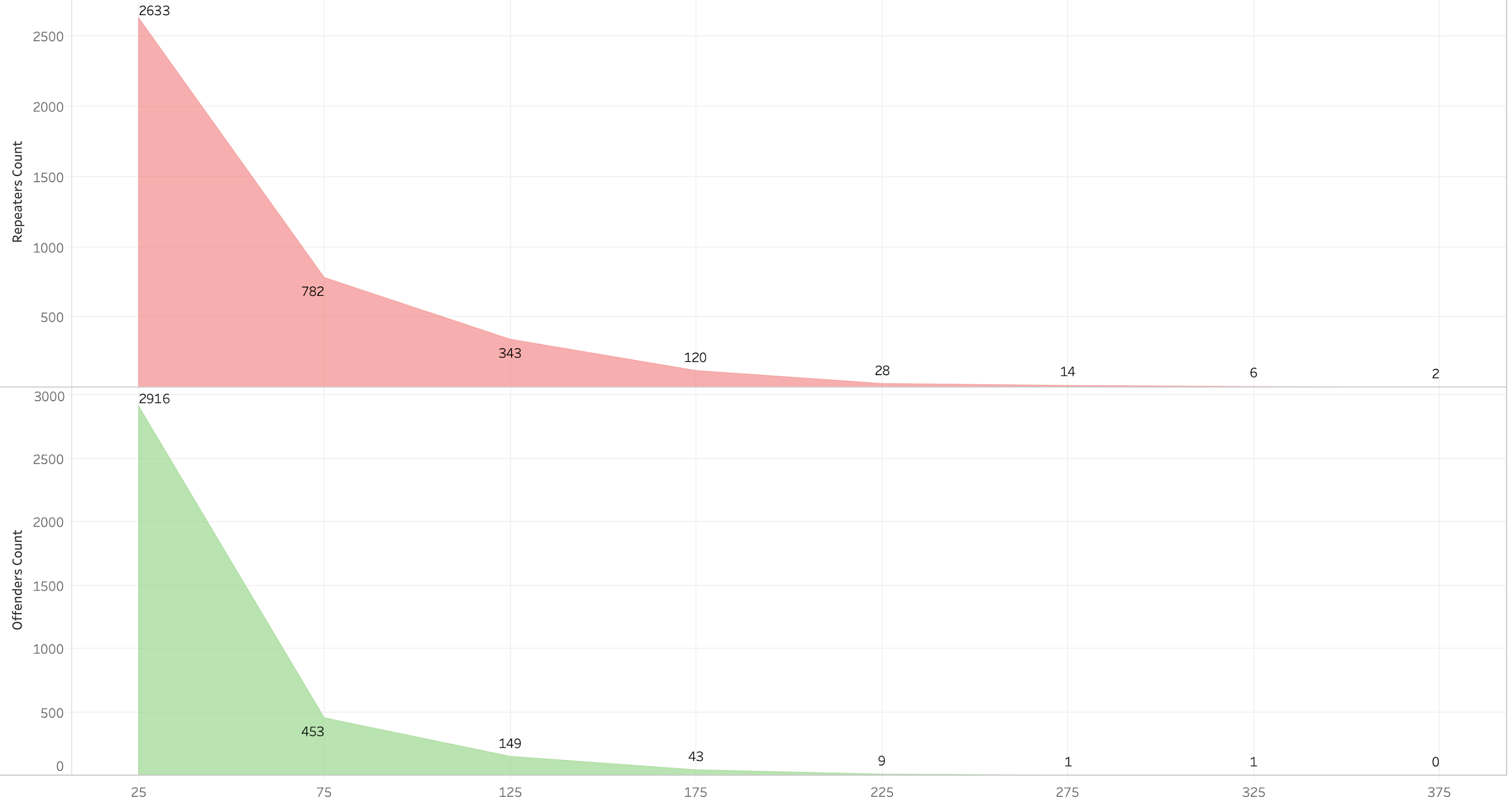
**Visualization Results**

We chose three particular traits (salary, award coun and complaints per beat from repeaters) that we think are interesting to investigate and compared the repeaters’ and offenders’ data on the three traits. Reason why we chose these three attributes is that we not only extracted significant data during our first checkpoint, but also realized that the pattern and distribution of these attributes might differ largely between the two groups. Please see next page for our results.

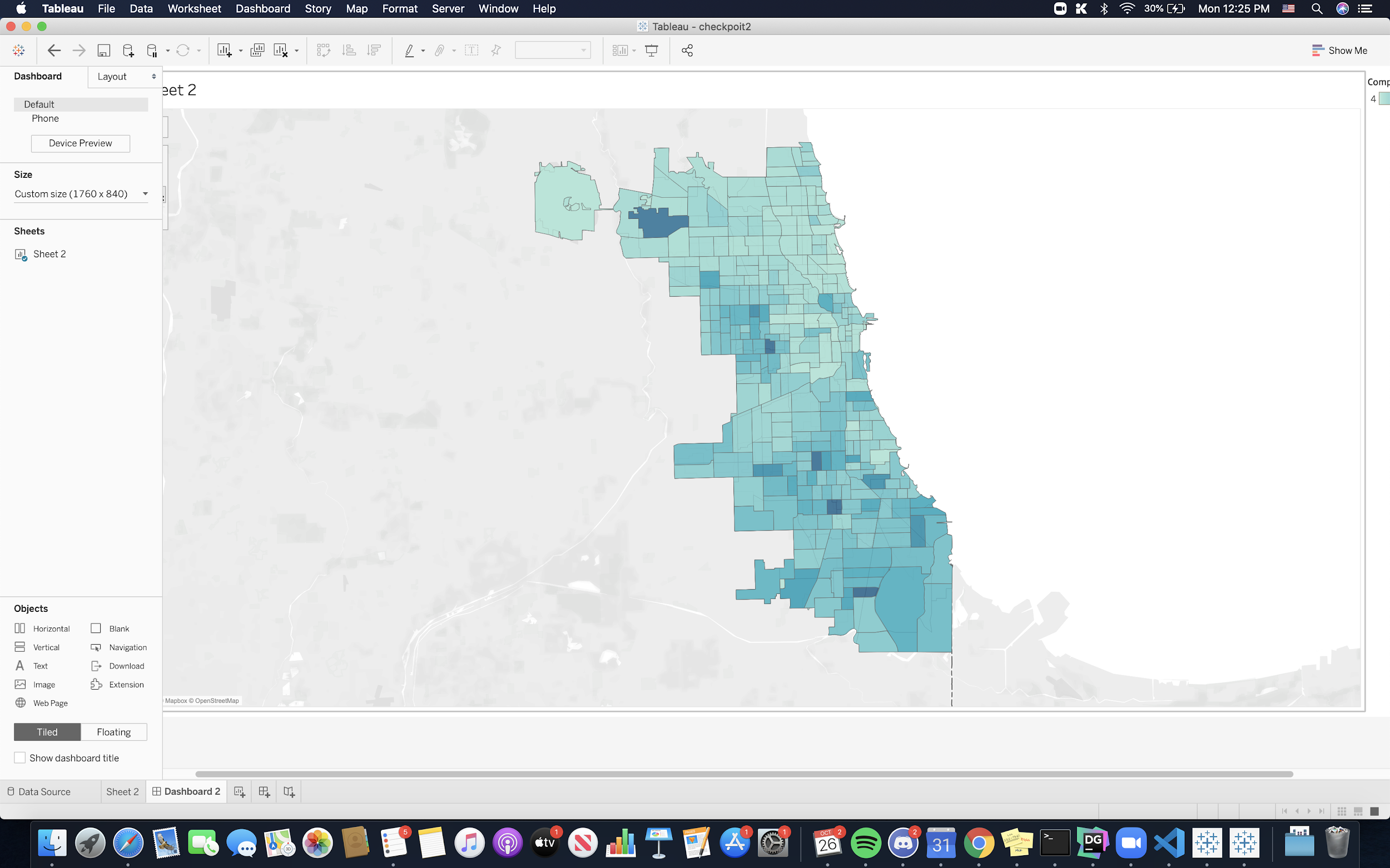
*Visualizing the Salary Range Distribution of Repeaters (in red) and a Random Sample of 4000 Offenders (in green)*

*Figure 1: The Comparison of salary range distribution of “Repeaters” and “Offenders”*

*Visualizing the Avg. Award Distribution of Repeaters (in red) and a Random Sample of 4000 Offenders (in green)*



*Figure 2: The Comparison of average awards distribution of “Repeaters” and “Offenders”*

**

*Figure 3: Number of complaints from top 4000 repeaters per beat mapped on city of Chicago*

**Analysis of Results**

1. In our first visualization, we presented the distribution of salary ranges of Repeaters and Offenders. We first grouped the individual salary values into ranges represented by the average these values. For example, officers with salaries from 70,000 to 90,000 will be grouped into the 80,000 salary group. We then calculated the percentage of each group relative to the total population. Since Repeaters have a population of 4000, we selected a random sample of 4000 officers from the Offenders to provide a standardized comparison.

Our results showed that most of the Repeaters (62%) have an average salary of 100,000 dollars per year while only 47% of Offenders have average salary in this range. We chose to use area charts instead of bar charts because it is efficient at showing both the volume and the distribution across different salary ranges. From the shape of the chart we realized that Offenders tend to have lower salaries on average than Repeaters and have higher percentages in the low salary ranges (80k and 60k). Repeaters also have more officers with very high salaries over 160k. The distribution and comparison between the salaries of Repeaters and Offenders are very perplexing because we expected lower salaries for Repeaters due to their high number of complaints received and potential penalties they received in forms of salary cut or fines. However, we observed that Repeaters have higher salaries overall than Offenders and have more individuals with very high salaries, making the salary policy questionable in the Chicago Police Department.

1. In our second visualization, we followed a very similar process to that of the salary analysis. We collected 4000 “Repeaters” and 4000 randomly selected “Offenders” and calculated their total number of awards received. Then we categorized their award count into several levels. For example, officers with awards from 0 to 50 will be grouped into the 25 award count group. This categorization process allows us to clean up the noise in the data and streamline the analysis process.

Our results showed that, surprisingly, “Repeaters” tend to receive more awards than the “Offenders”. While both groups have most of their officers fall into the first category of 25 awards on average, more officers receive higher awards count in the “Repeaters” group. For example, over two times more officers in the “Repeaters” group received on average 125 awards than the “Offenders” group, and almost three times more “Repeaters” received on average 175 awards than the “Offenders”. More “Repeaters” also received extraordinarily high awards counts over 300. This is also something unexpected because by logic one would assume that officers who receive complaints more often tend not to receive awards from the department, yet the data shows the direct opposite.

1. In our third visualization, we have taken the queries we created from the 1st checkpoint, wherein we collected the top 4000 officers with most complaints and mapped their total number of complaints per beat. And we used the region boundaries to map all the beats with those number of complaints on the city of Chicago.

Our results show that only taking a quick glance at the picture, its apparent how south of chicago has more complaints from repeaters than north of chicago. The beats with most complaints are spread throughout Chicago however. This paints a picture of the officers behaviour towards misconduct correlated to the regionality. There are a few outliers which need to be further investigated like Norridge and Harwood heights have high number of complaints from repeaters in the northern part.

**Conclusion**

As a result, we think that our data visualizations showed significant and surprising results on the potential commonalities among the “Repeaters”. We now understand that “Repeaters” tends have higher than average salaries and awards count and their complaint patterns are correlated with the regions they work most in.

**Experience with Tableau**

As first time users of Tableau, we found the learning curving was very steep for a professional data visualization tool like Tableau. The design principle and language is very different from what we are used to. As a data visualization tool, its data query interface is vastly different from tools like MySQL Workbench or DataGrip, and its data visualization interface is new and confusing to use with numerous setup or data transformation options. For example, in the data visualization panel, the notion of rows and columns is very hard to understand as opposed to x-axis and y-axis that one would expect in a coordinate-based visualization window. We all believed that was one of the most confusing and dis-orientating tools to use for beginners. As a result, neither the data extraction process nor the data visualization process was easy to pick up.

To make things worse, most of the Tableau tutorials and instructions are based on the sample data provided by Tableau while the data we analyzed in reality is usually vastly different from that in the sample. Therefore, the tutorials provided very little value.

However, we firmly believe that Tableau has great potential and flexibility; it can be extremely powerful under the right hands. Yet we believe masting Tableau will require formal training over long periods of time to fully understand the principles behind Tableau design and the functionalities the Tableau provides.