Checkpoint 2: Data Visualization

The Silver Imps
Song Luo, Jing Jiang, Yuxin Chen

Introduction

Prior to analysis, we would like to restate our definition for repeaters: repeaters are the top 10% of officers with most complaints in a timespan.
For 2006-2010 we identify 887 repeaters among a total of 8874 officers.
For 2010-2014 we identify 744 repeaters among a total of 7445 officers.
For 2014-2018 we identify 557 repeaters among a total of 5567 officers.

Due to the definition we feel the original visualization questions in the Project Proposal are not adequate enough. Therefore we would like to present graphs that visualize all four questions in Checkpoint 1 - Relational Analytics with Tableau.

Q1: Using our definition for “repeaters” above, what percentage of total complaints are they responsible for?

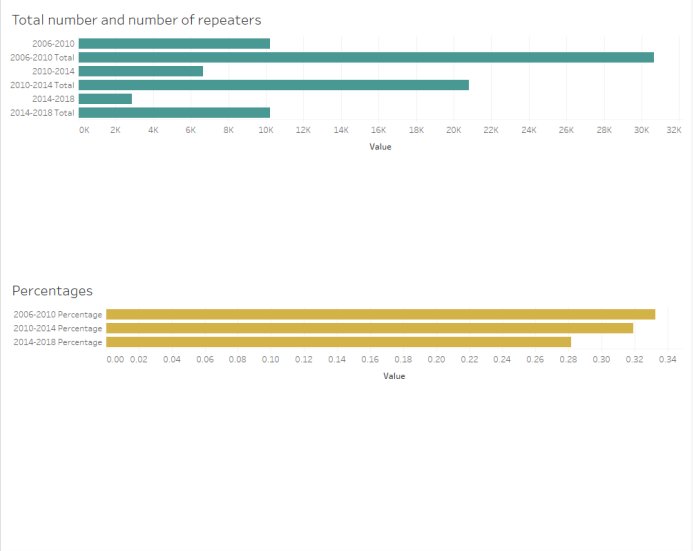

Figure 1 : percentage of “repeater” complaints

The graph on the top shows the number of “repeater” complaints and total number of complaints respectively in each timespan. The graph on the bottom shows the percentage of “repeater” complaints in each timespan. By using horizontal bar chart, we can see numerically there are significant drops both in “repeater” complaints and total complaints. However the “repeaters” steadily contribute around 30% complaints through all time, which is a significant amount considering that “repeaters” are only 1/10 of the total officers. It is therefore necessary to analyze deeper into the “repeater” complaint patterns.

Q2: What is the demographic Information (race, age and sex) of the “repeaters” using our definition above?

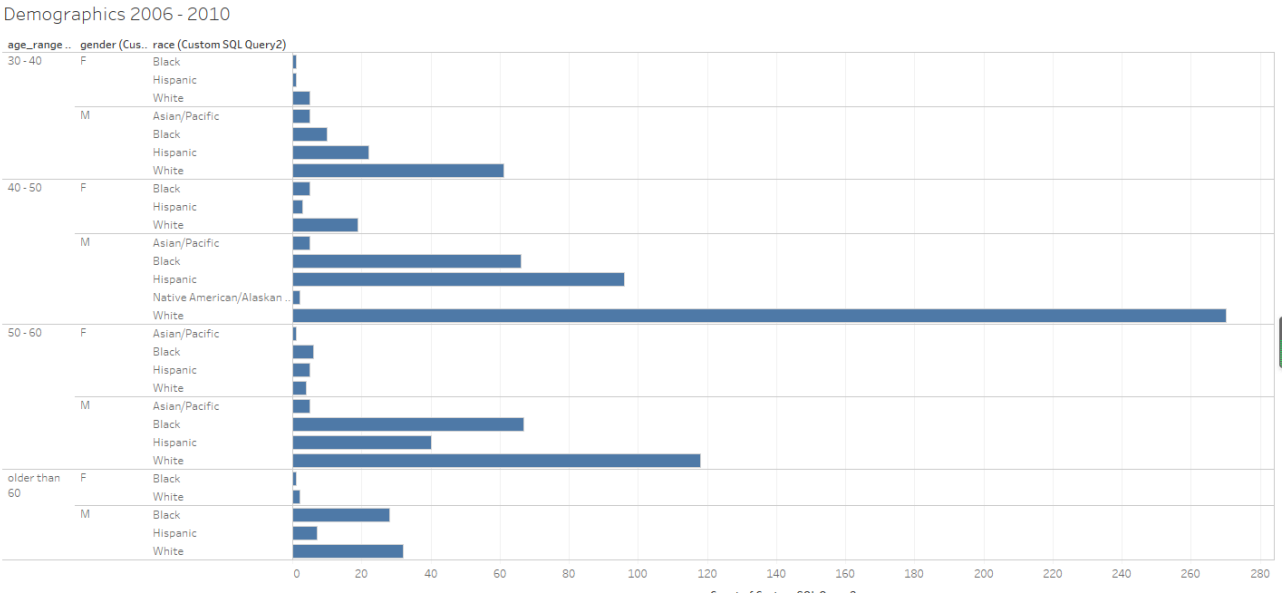

Figure 2: demographic information for 2006-2010(bar chart)

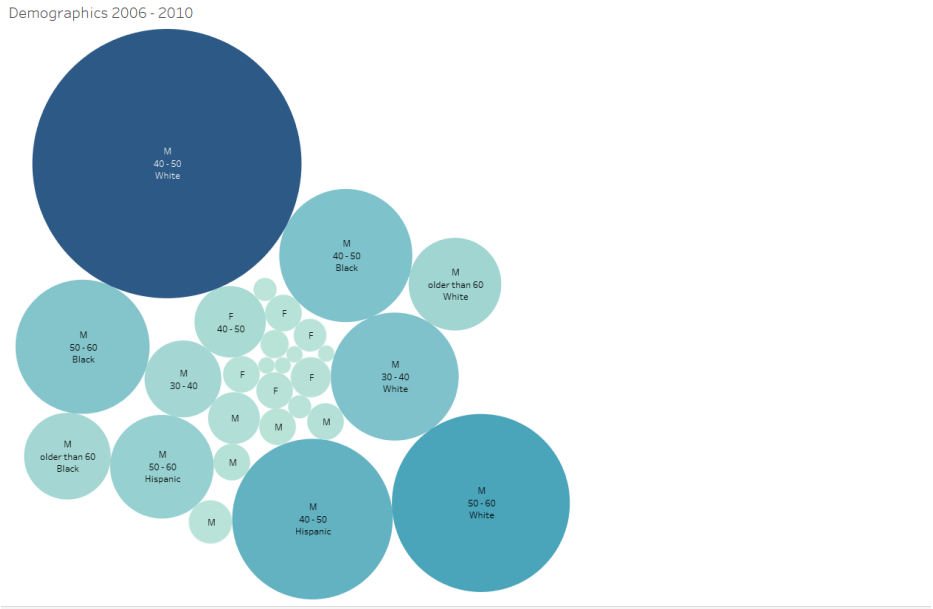

Figure 3: demographic information for 2006-2010(bubble graph)

First from the age aspect, it is clear that in 2006-2010 timespan the majority of the repeaters are clustered in age 40-50, as demonstrated in the bubble graph. There are also significantly more men than women among the repeaters: the fact that there are more male officers than female officers is part of the reason; still it is worthwhile to find out if male are easier to subject to misconduct and we can take training/other precautionary measures accordingly. In terms of the races, white, black and hispanic are three dominant races for repeaters.

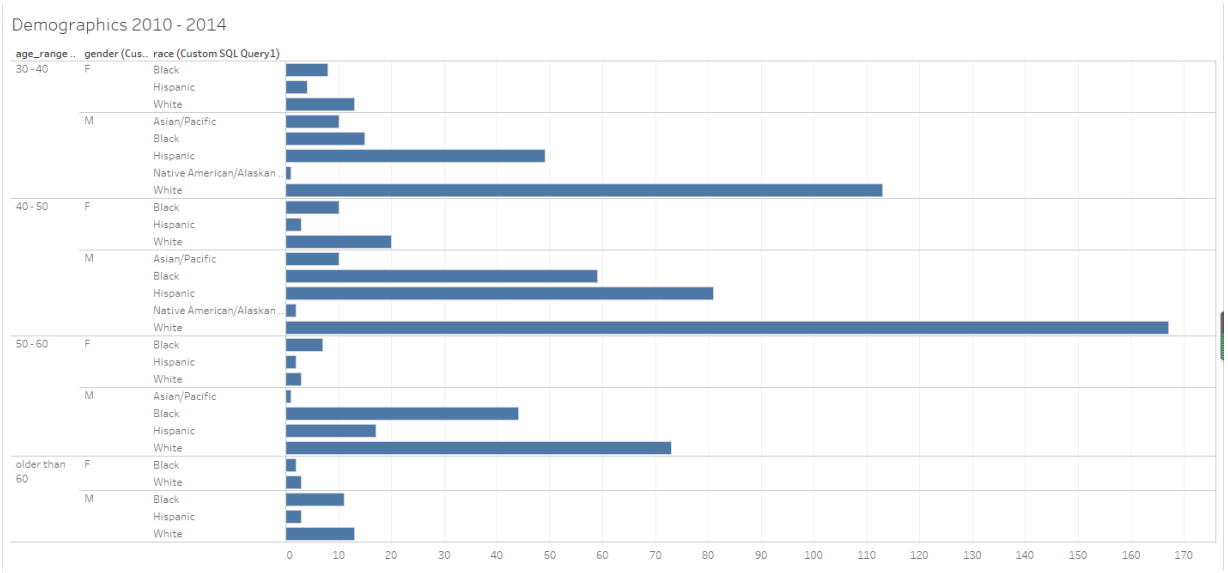

Figure 4: demographic information for 2010-2014(bar chart)

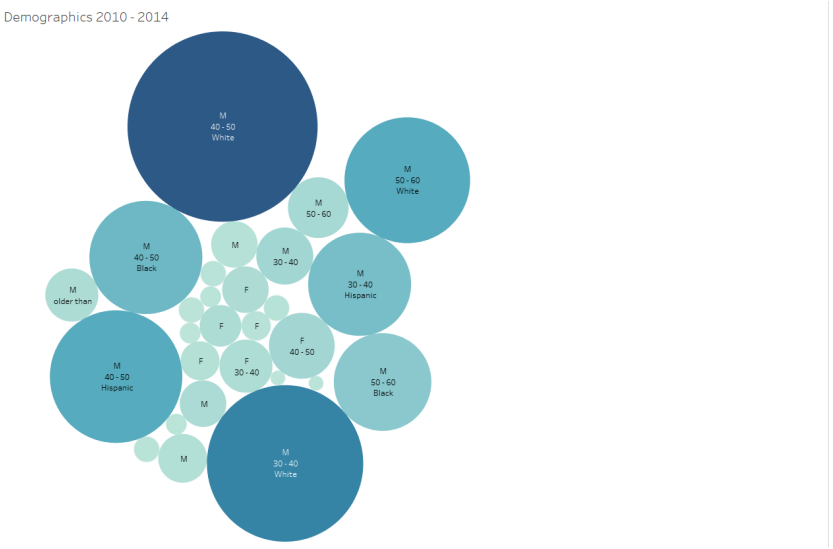

Figure 5: demographic information for 2010-2014(bubble graph)

We can see that most demographic patterns are consistent: the most dominant group is still “Male, White, Age 40-50”. However we should see that age 30-40 are also emerging as a main source of “repeaters”.

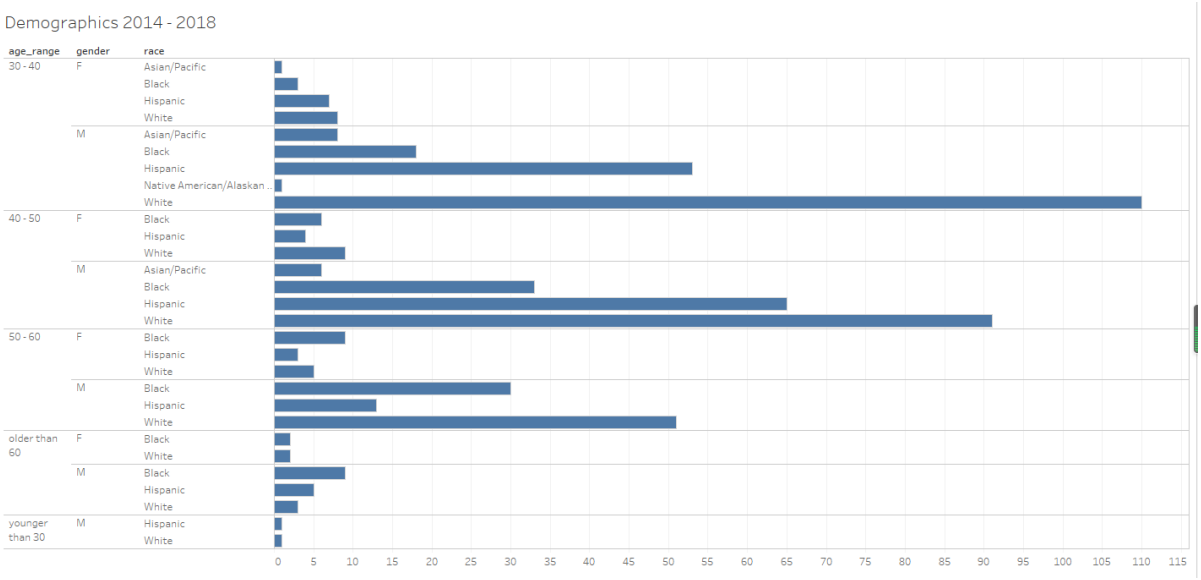

Figure 6: demographic information for 2014-2018(bar chart)

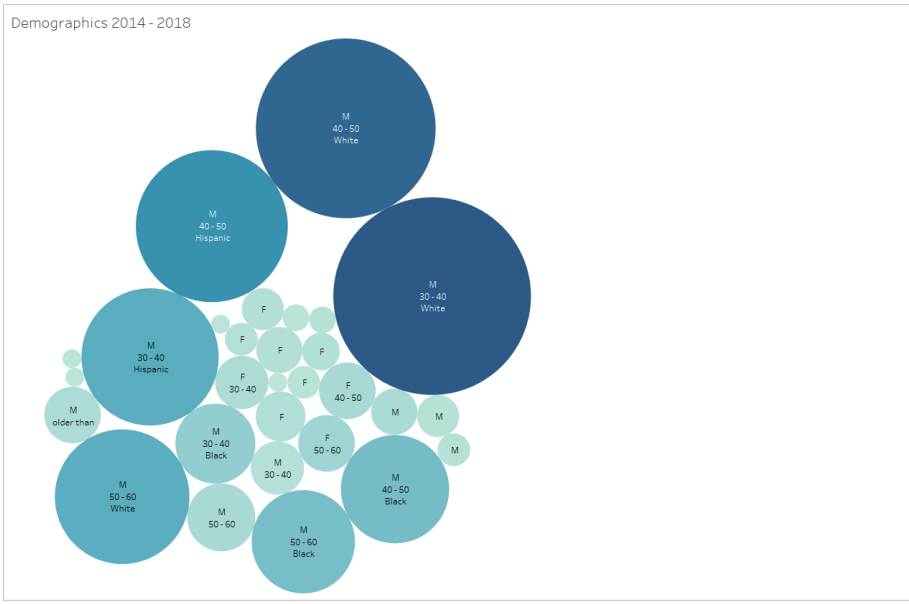

Figure 7: demographic information for 2014-2018(bubble graph)

We can see that “repeaters” keep getting younger: now the most dominant group is “Male, White, Age 30-40”. Also Hispanic is also becoming increasingly prevalent in the “repeaters”: there seems to be about the same number of Hispanic officers and that of Black officers, now the number of Hispanic officers is conspicuously larger than number of Black officers.

If we view these three timespans as a whole, we can deduce that the composition of “repeaters” are also changing. Otherwise we can expect to see Age 50-60 as the majority

group (when those Age 40-50 in 2006-2010 turn older). That “repeaters” are getting young is a concerning trend.

Q3: What is the distribution of the category of the misconducts for “repeaters”?

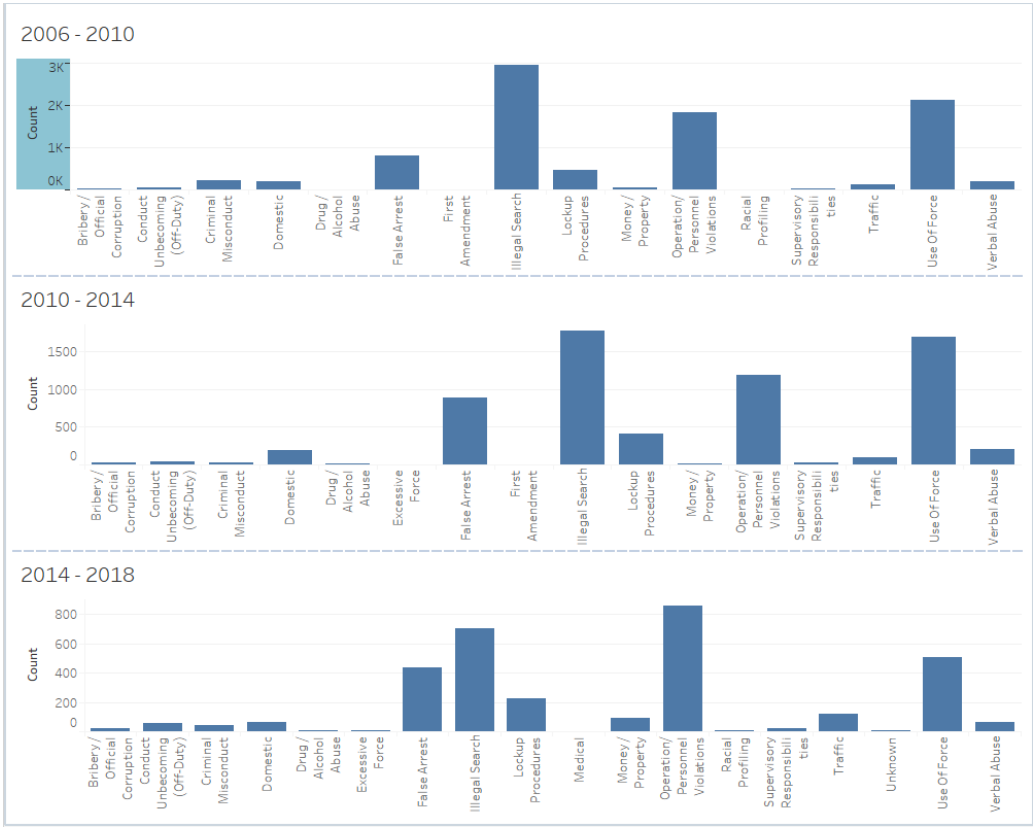

Figure 8: distribution of misconduct categories

According to the histogram, we can see that Illegal Search, Operational/Personnel Violations and Use of force, False arrests remain the four top categories of misconduct. It should be noted that these three similar graphs actually adopt different scales, so overall we can see a sharp drop of misconducts in all categories over time. The percentage of Operational/Personnel Violations keeps rising and in fact, it has become the leading misconduct as of 14-18. However, this is the more trivial side of misconduct as the description includes “smoking violations”, “invalid license plate”, etc. What we should focus on are illegal search use of force and false arrests, all of which against individual will and harmful to the innocent. I believe more stringent protocols should be established as to when and where the searches should be performed in and the force used, in an appropriate manner.

Q4: Among the allegations against “repeaters”, what percentage of these cases lead to the officers being disciplined?

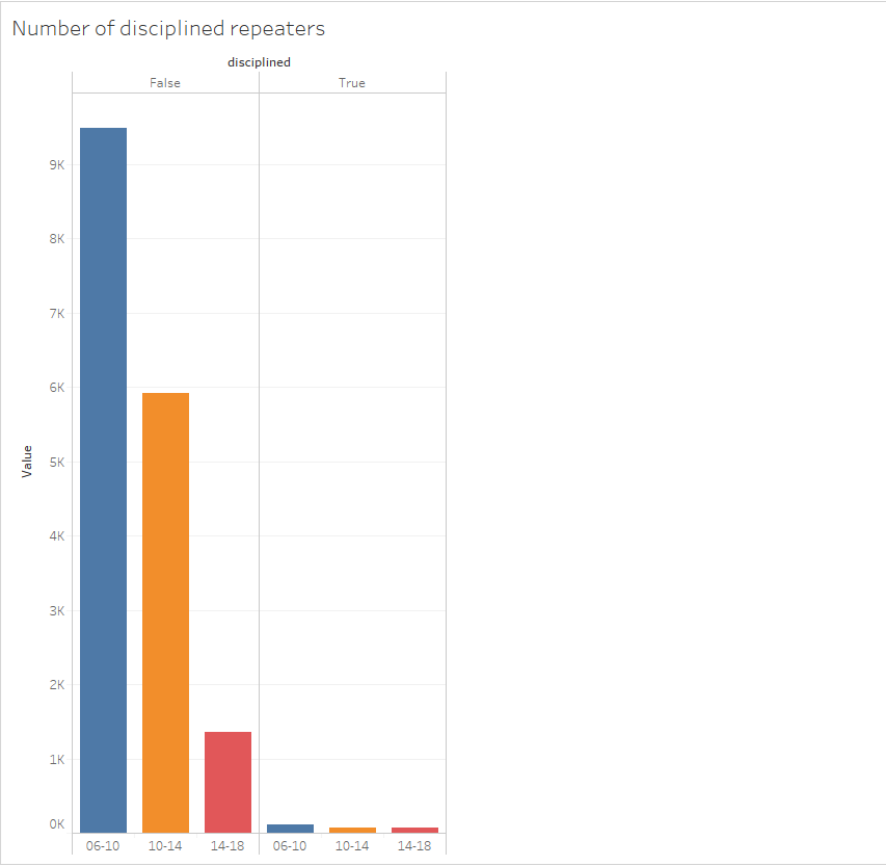

Figure 9: Disciplined v.s. Not Disciplined

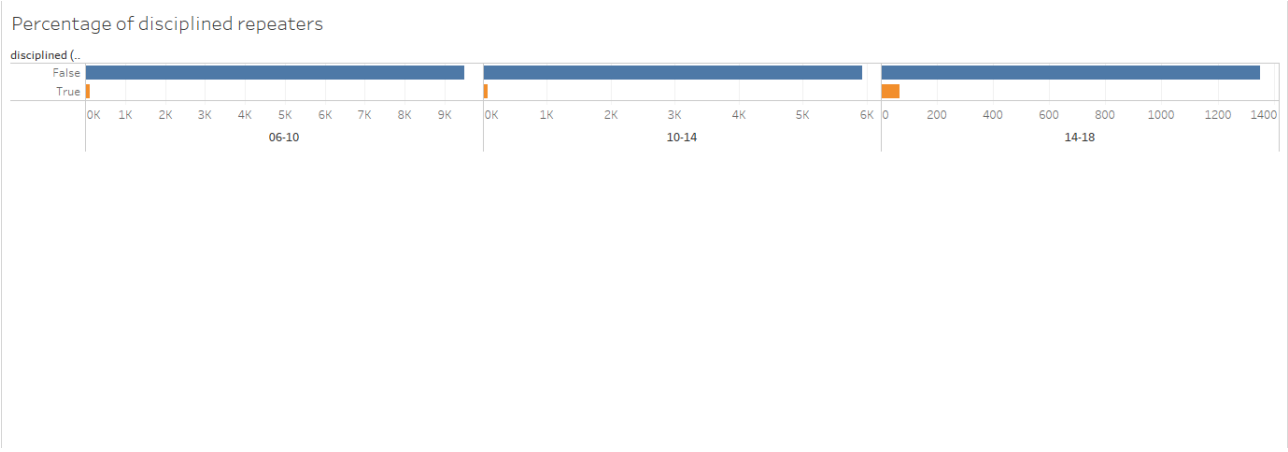

Figure 10: Disciplined v.s. Not Disciplined(cont.)

When we are querying we did not include data entries that has null value in “disciplined” field: indeed there are quite some amount of allegations that we have no information about disciplinary decisions. I suppose more efforts could be dedicated to clear these null values. From Figure 9, we can see a sharp drop of allegations that lead to disciplinary decisions over time. As we have indicated in Checkpoint1, we need to see why there is such a considerable drop in total allegations: is the count going down because repeaters are behaving better or are people filing fewer allegations for some reason?

In Figure 10, from the comparison we can see how few allegations that actually lead to officers disciplined. This is concerning because according to the data in Q3, illegal search and use of force consists a significant amount of police misconducts; yet so few of them are facing consequences. Maybe the disciplinary system also needs to be more transparent about how the decisions are made.

Experiences with Tableau

It is with no doubt that Tableau is getting popular as a data visualization tool. Yet we experience serious connectivity issues with Tableau. We have both Windows and Mac users in the group. Neither of us can connect to our local Postgresql Database. The documentation is poor and the error code, upon searched, returns no meaningful tips or fixes at all. The Windows side can connect via AWS when drivers are installed; however it takes a long time for table to show up and query to be executed, and still faces intermitting connectivity. Also Tableau does not support creating temporary view and only supports SELECT statements. We have some complex logic in our SQL statements yet we have to convert them to a gigantic nested loop of SELECT statements, which is quite counter-intuitive.

Tableau is not without advantages. There is a variety of visualizations template to choose from, and also it provides built-in visual best practices. Overall I would say Tableau is a good attempt for creating data visualizations in a fast and intuitive way, but it is ill-suited for business purposes just yet.