Tableau Report 4

Group 2: Tingting Ju, Aimee Li, Yizi Li, Leiwen Lin, Jiaying Shi

At group 2, we chose DC Metro Crime Data as the dataset to be visualized in our tableau project. This set includes records of “crimes in the DC metro police system ranging from Theft, Arson, Assault, Homicide, Sex Abuse, Robbery, and Burglary” (<https://www.kaggle.com/vinchinzu/dc-metro-crime-data>), between 2000 and 2016.

We chose this dataset for the following reasons. First, as newcomers to the DC metro area, knowing the crime rate and distribution around metro stations can greatly improve our ability to protect ourselves. As there are many in our class who are new to the area and are frequent metro riders, we believe this dataset will be of great interest to our classmates.

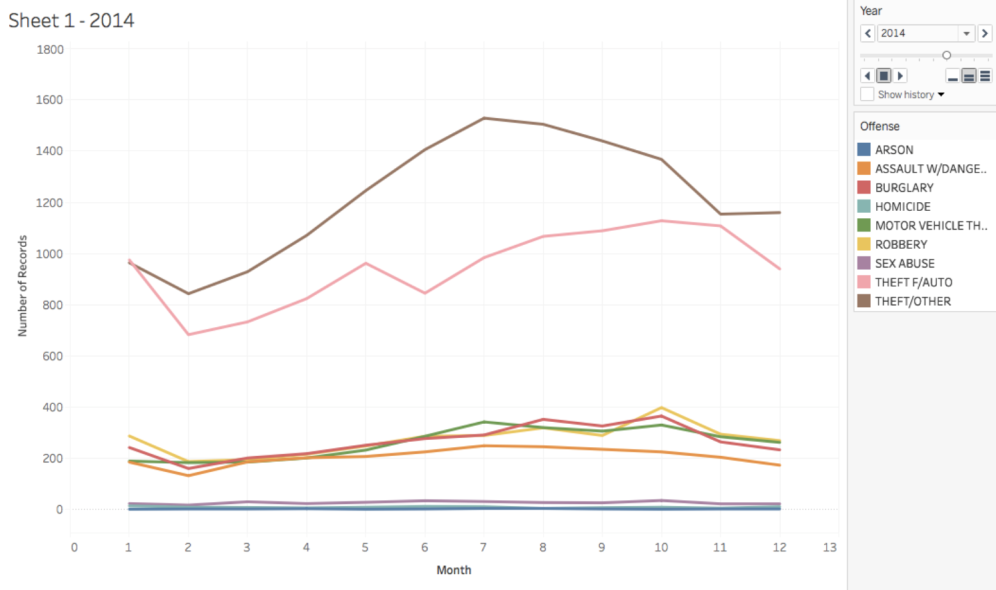
Secondly, this dataset contains time variables such as year, month and date. It also has geographic information, in addition to many other attributes. It can easily be visualized in multiple ways including mapping, time series, categorical analysis, cross-section comparison, etc. Such a rich dataset will enable us to explore the full potential of the tableau software.

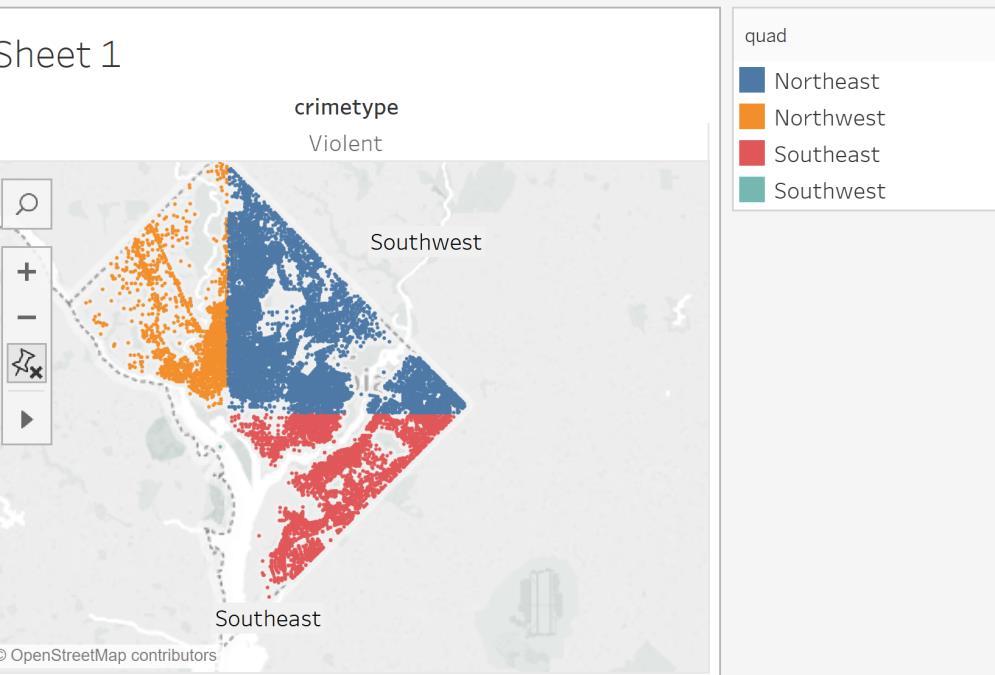
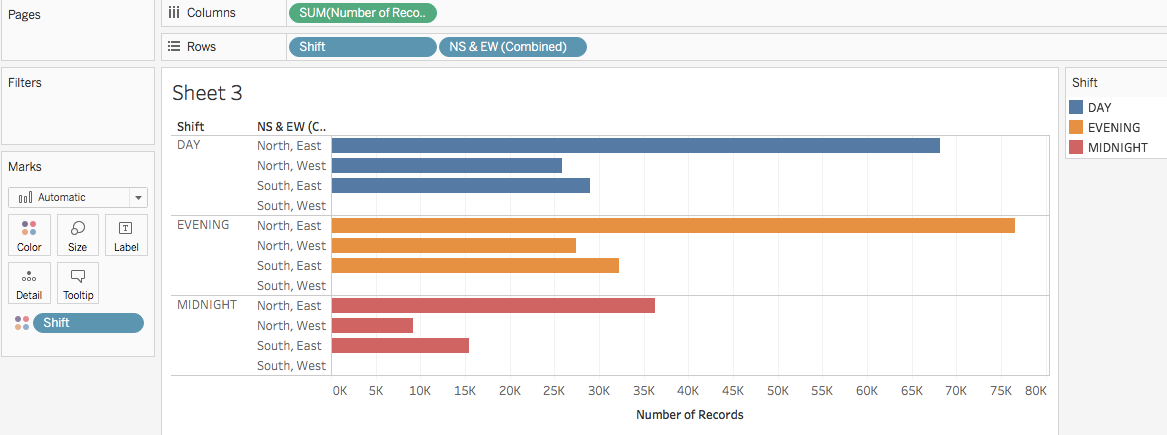
Thirdly, this dataset is rather complete. No null field is visible at a first glance. This simplifies the data preparation step. We could start looking for trends and correlations directly. In addition, there is the possibility of taking advantage of existing kernels for the analysis of this dataset.

After all our group members have completed the introductory training videos, we met up once after we gained basic understand of tableau to discuss how to apply it to our dataset. To create visualization of our dataset, we first imported it to tableau. Then unnecessary columns, such as Census Tract, CCN, were removed. We combined columns NS and EW to create a single dimension variable for easy visualization. We also learned how to use filters and pages panes to simplify the figure.

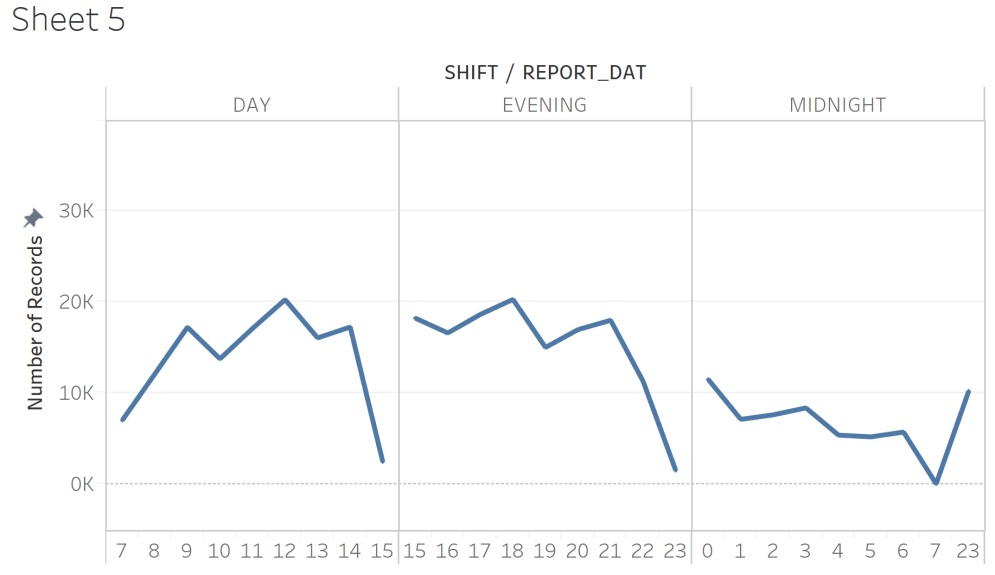
We created several graphs to analyze the dataset. First, we studied the frequency of different types of crimes and discovered that THEFT is the most common crime.

Secondly, we turned to the regional distribution and plotted the crime location on a map to find out which area in DC has the highest criminal rate. The graph clearly shows that the northeast of DC has the highest number of crimes in all time periods.



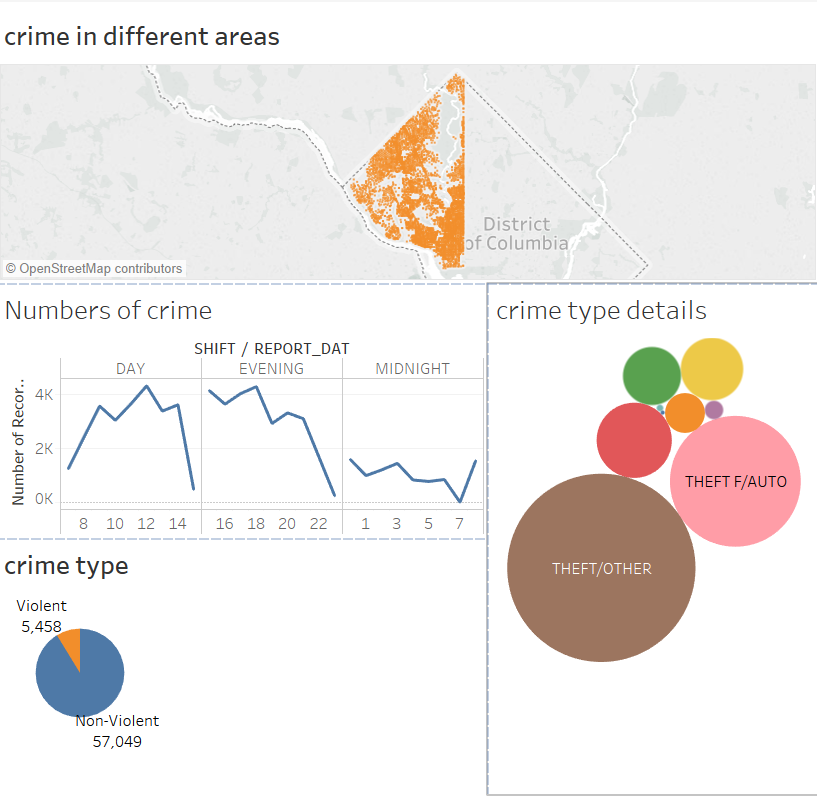
 

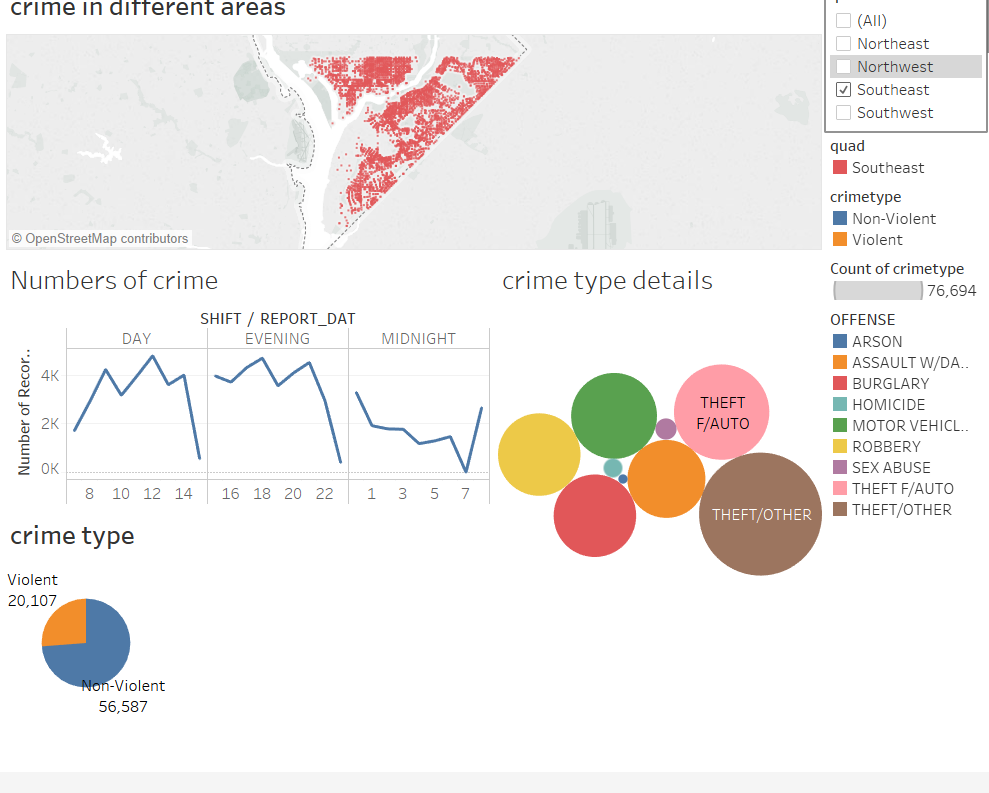
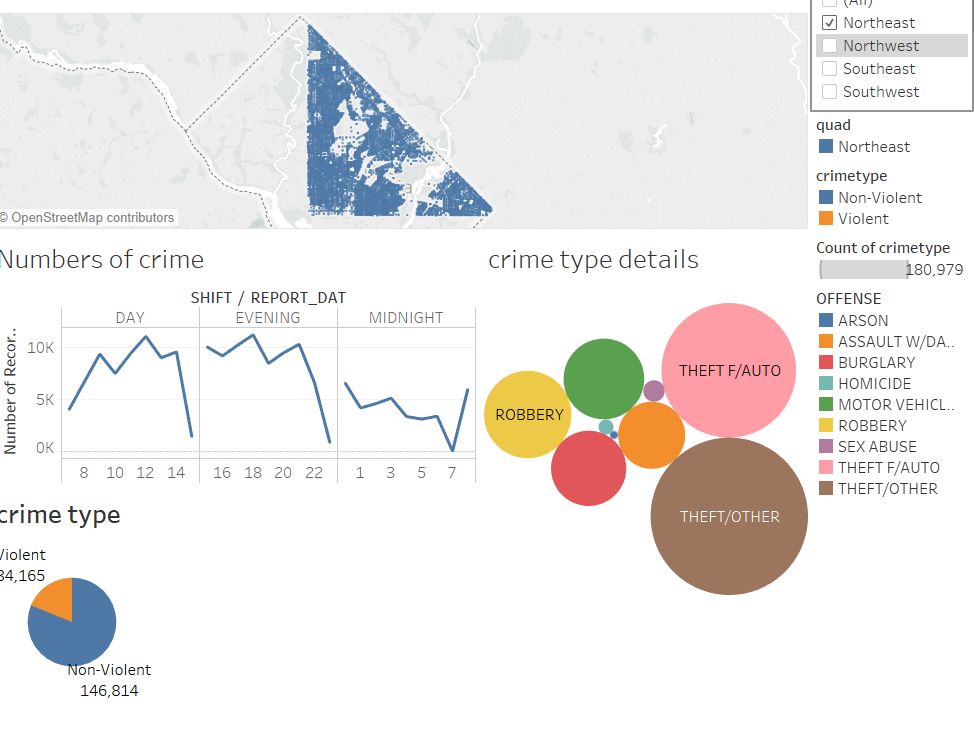
Then, we tried to figure out the hourly timing pattern of crimes. Generally speaking, 12:00pm and 18:00 are the most dangerous time around the metro stations. We recommend people to plan their activities accordingly or be extra vigilant if obligated to go out around those hours.



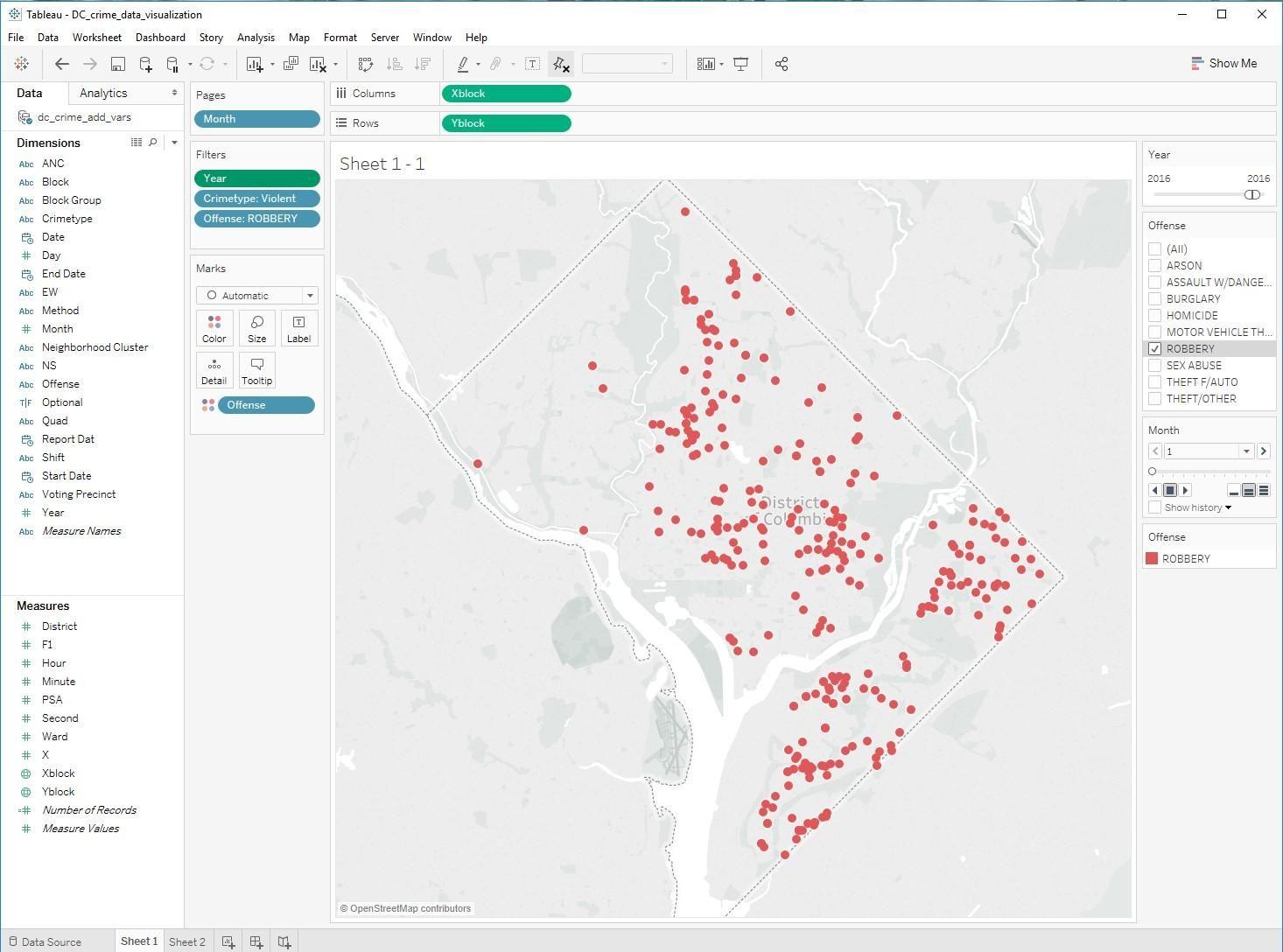
Finally, we considered and combined multiple graphs together, including the crime types, the regional distribution, and the time distribution, in order to get a multidimensional understanding. From the following figures, we concluded the following:

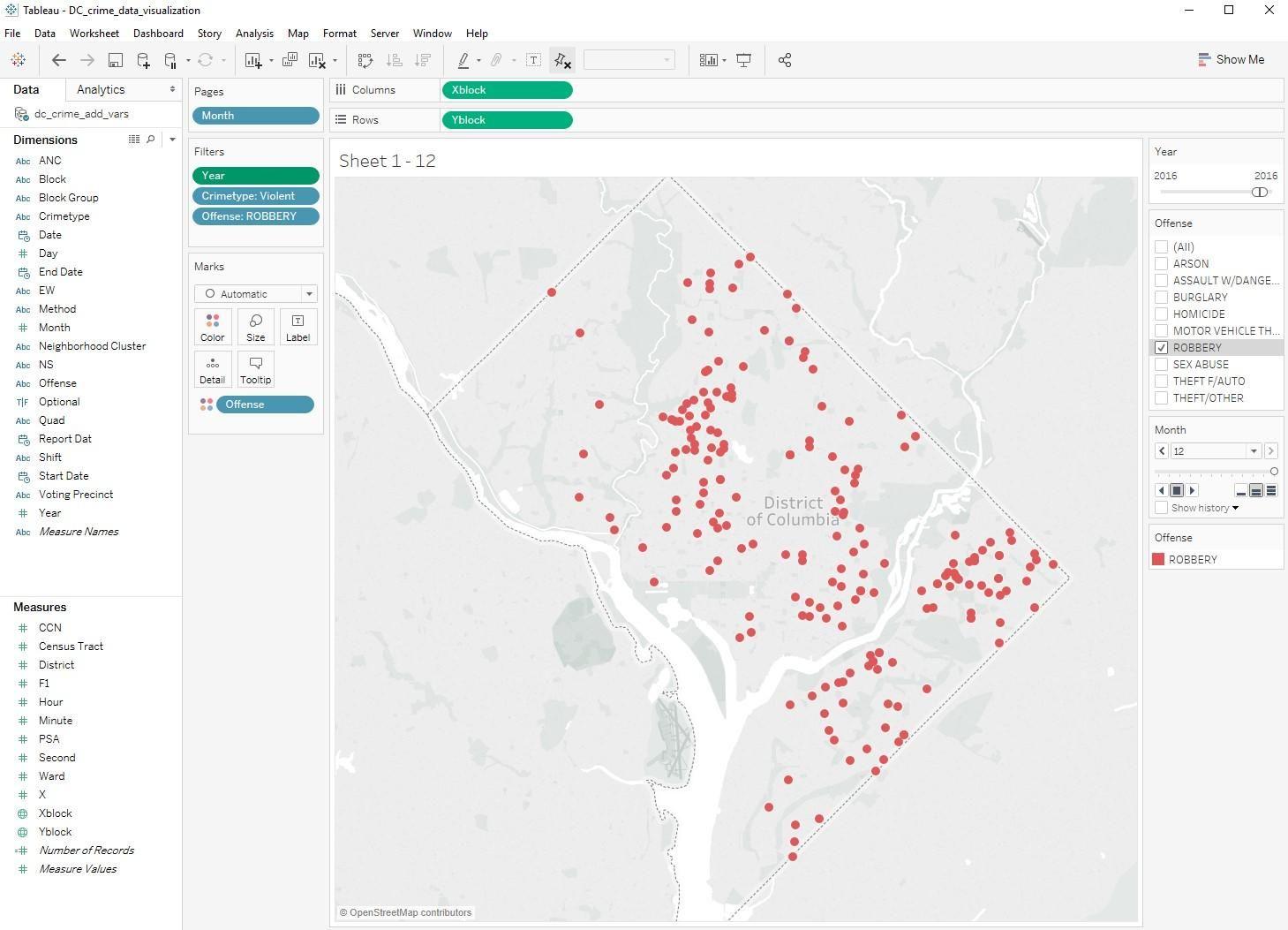
1) Northwest has higher crime rate than other quarters. 2) In all areas, 12:00pm and 18:00 are the most dangerous time of the day; 3) THEFT type of crime are the most likely to happen; and 4) non-violent cases are far more than violent cases in those three areas.



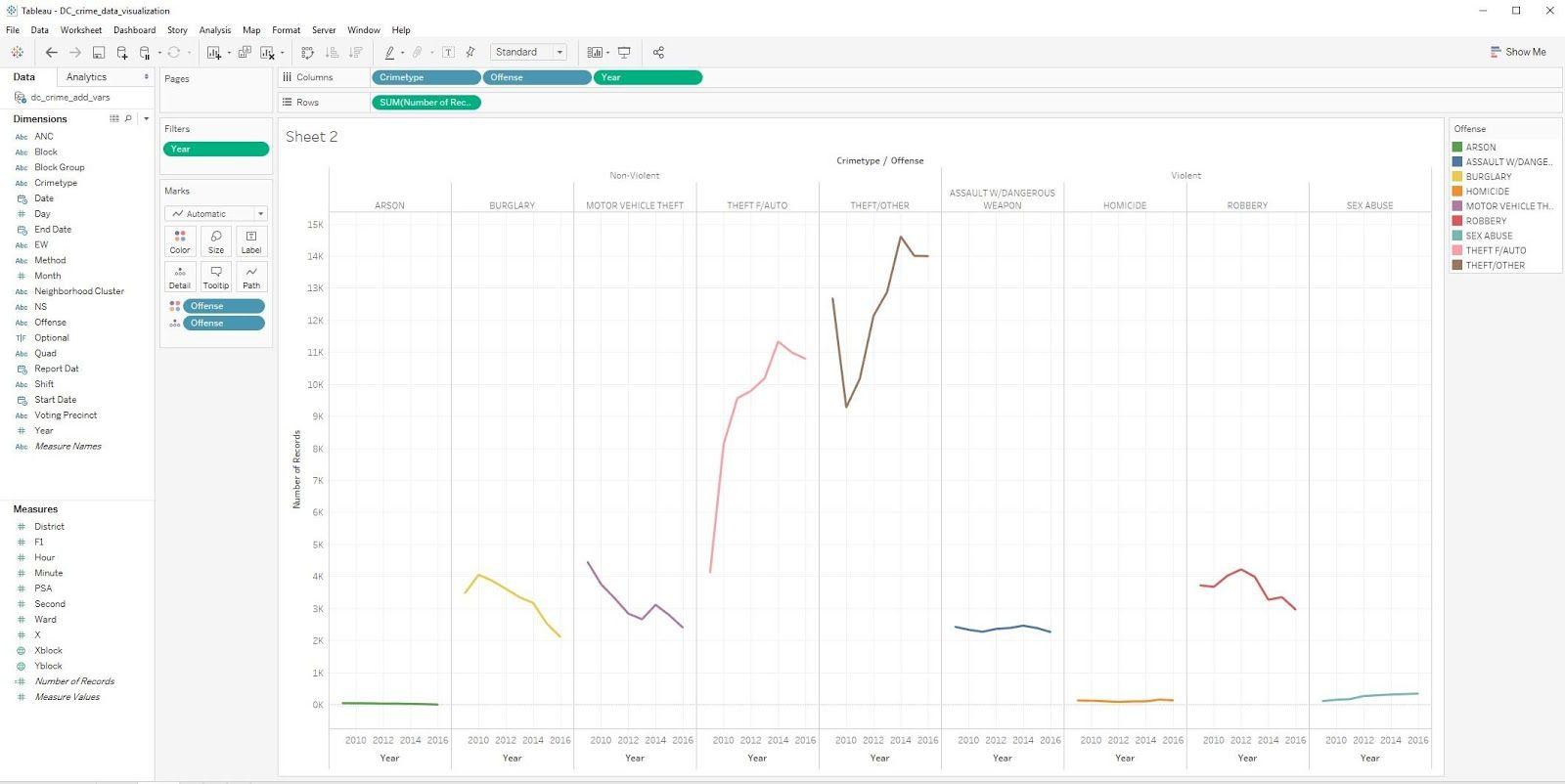
Next, we pay more attention on how to show an animation of monthly crime distributions on one specific crime type and how to visualize the time series trend of each different crime type.

Firstly, we work on the animation of the regional distribution on a map. Take the crime type as ROBBERY for example. 



The first graph is showing where Robbery happened during the month of January in 2016, and the second graph is showing where Robbery happened during the month of December in 2016. And we can get our animation graph for each month of the selected year by clicking on the play button within the Pages' card.

Secondly, we try to display the trend of each crime type over time.



From the above graph, we can figure out that THEFT not only is the most common crime (the conclusion from our last report) but also keeps growing year by year. Comparatively, the frequency of BURGLARY, MOTOR VEHICLE THEFT, and ROBBERY are descending over time.

The primary challenge we faced is how to find meaningful and informative ways of presenting the data. Tableau offers a myriad of tools, therefore a rich dataset such as ours can be visualized a million ways. Yet identify which ways are visually appealing and can tell a nice story is a challenge. We managed to use filters and pages to simplify the visualizations and made a few interesting graphs.

In summary, visualizing DC Metro Crime Data not only enable us to learn tableau well, but also can benefit many in the current class.