Tableau Report 4

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

As 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.

The first step of our project is data cleaning. The unnecessary columns, such as Census Tract, CCN, were removed. Then we combined columns NS and EW to create a single dimension variable for easy visualization.

We made 6 graphs in total.

* Crime in different areas: This graph is the regional distribution. We 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.
* Bar chart types: We can see that over 80 percent of the DC metro crime are non-violent crime.
* Percentages of different crime: We discovered that THEFT is the most common crime.
* Numbers of crime in different period: This graph shows the hourly timing pattern of crimes. Generally speaking, 12:00 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.
* Trends of crime time: From the graph, we can figure out that THEFT not only is the most common crime but also keeps growing year by year. Comparatively, the frequency of BURGLARY, MOTOR VEHICLE THEFT, and ROBBERY are descending over time.
* Animation graph for each month: This is the animation of the regional distribution on a map. We can get our animation graph for each month of the selected year by clicking on the play button within the Pages' card.
* Bar chart of distribution of numbers of crime among different times and regions: From this graph, we can easily discover that northeast has the highest record of crime through whole day. The second one is southeast. Evening the time have most crimes in all area.

In conclusion: