**Student Name: Duy Ly, Trung Phan**

**Class: STAT 6610 – Section 8PM**

**Date: 03/13/2018**

**TRAFFIC VIOLATION ANALYSIS**

**1. Data description:**

This dataset contains the daily volume of speed violations which have been recorded by each camera installed in the City of Chicago as part of the Automated Speed Enforcement Program. This data is collected from July 1, 2014 until August 29, 2017. The reported violations were collected by the camera and radar system and reviewed by two separate city contractors. The speed camera data was collected and published by the Chicago Police Department on the City of Chicago data portal website.

We collected this dataset from the webpage [www.kaggle.com](http://www.kaggle.com). The dataset is initially contributed by Abigail Larion and updated by Jacob Boysen

The dataset comprise of six main variables which are ADDRESS, CAMERA ID, VIOLATION DATE, VIOLATIONS, LATITUDE AND LONGITUDE

* ADDRESS variable presents the location where traffic camera set up.
* CAMERA ID variable contains names of cameras set up
* VIOLATION DATE variable includes the date when speed violations happened
* VIOLATIONS contains the number of violation which occurred on specific day
* LATITUDE AND LONGITUDE are variables contains latitudes and longitudes which is used to navigate the location where cameras were placed.

**2. Data Visualization and report:**

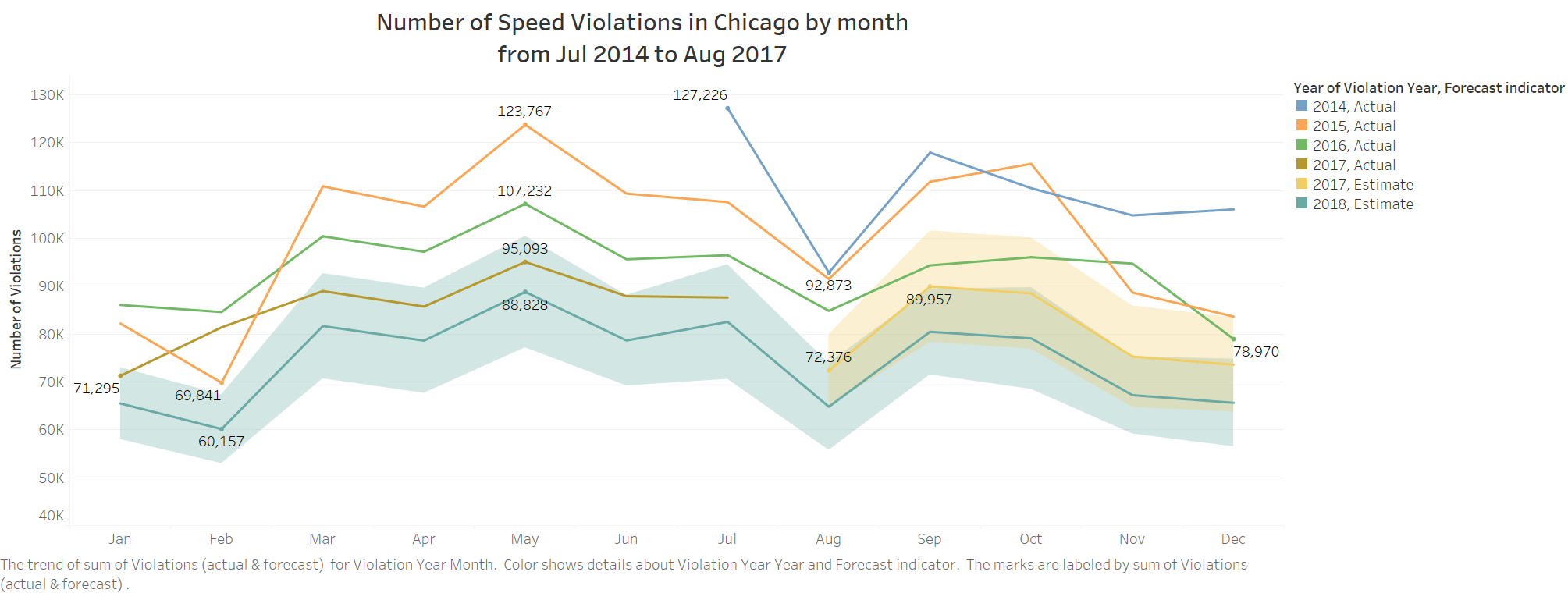


Figure 1

The figure 1 above showed the time series plot of number of Speed Violations in Chicago monthly from July 2014 to August 2017.

The number of Speed Violations recorded had decreased over the time. We can clearly see the seasonal pattern of violations volume in the time series plot. The number of Speed Violations increase over the first quarter of every year, hit the peak volumes in May then gradually declined until reach the lowest points in August before recovering again during last quarter.

As we can see, the time series plot showed a strong seasonal pattern where speed violation occurred more frequently in mid-spring and less frequently in mid-winter and at the beginning of summer.

Based on this seasonal trend, the total number Speed Violations is predicted to be continued dropped down in 2018 with the highest point (88,828 violations) in May and the lowest point (60,157 violations) in February.

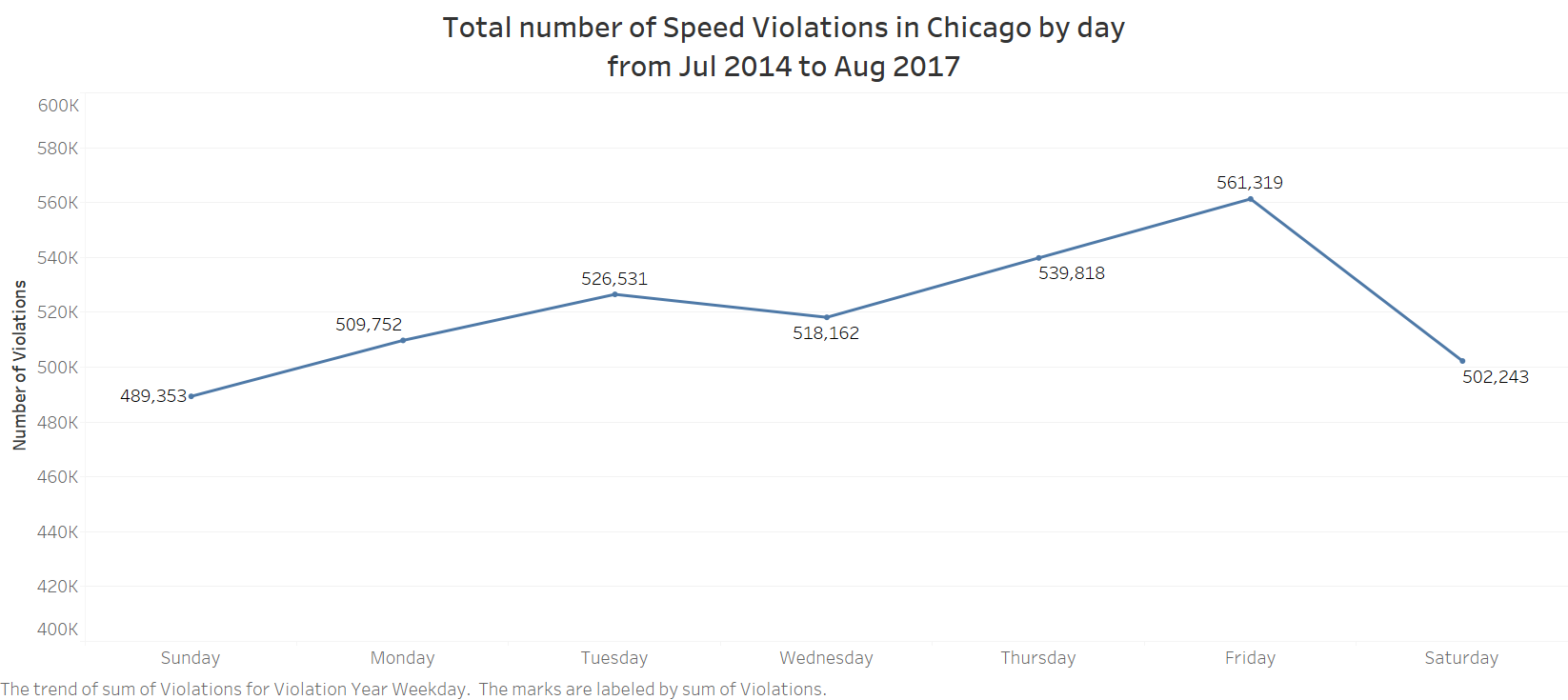


Figure 2

Figure 2 above illustrates the breakdown for the total number of Speed Violation into days of week from July 2014 to August 2017. As we can see, violation volume was mostly found to be increasing during weekdays, especially in Friday and then fall down in weekends

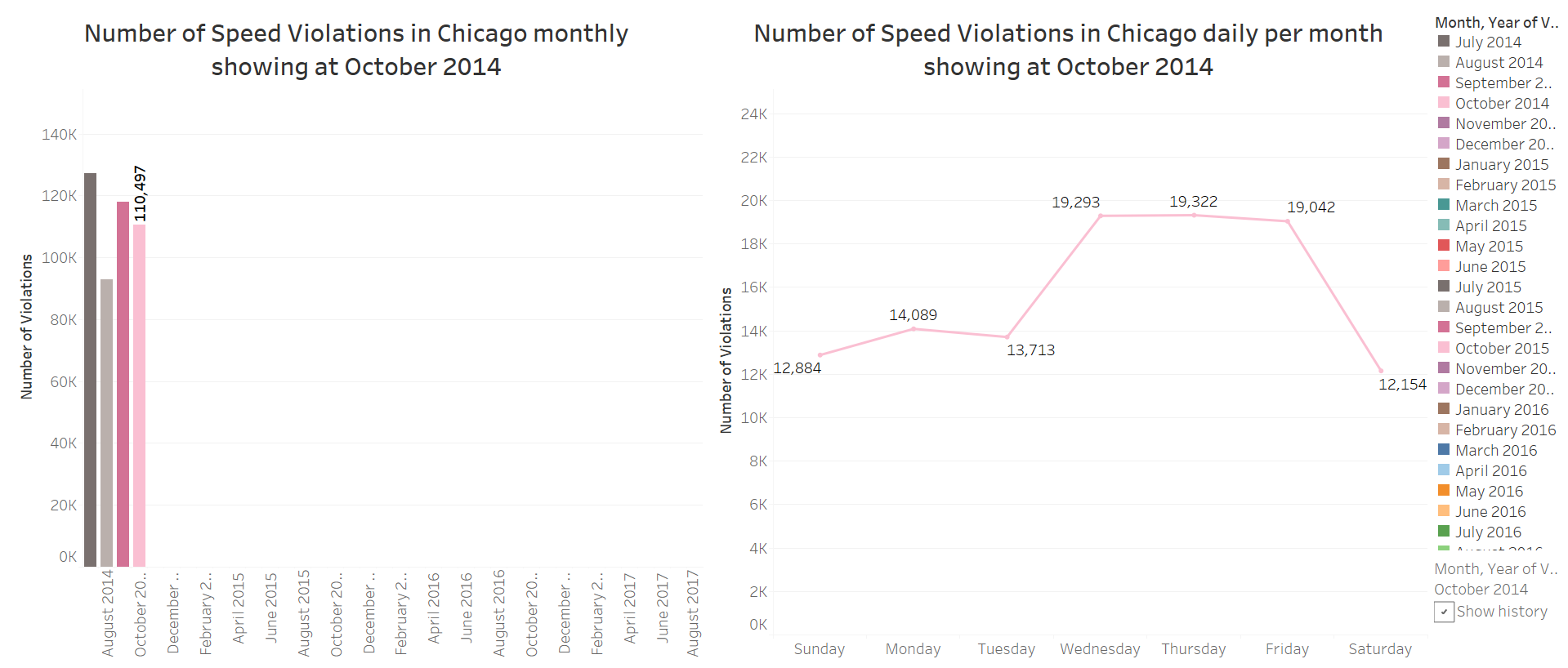


Figure 3

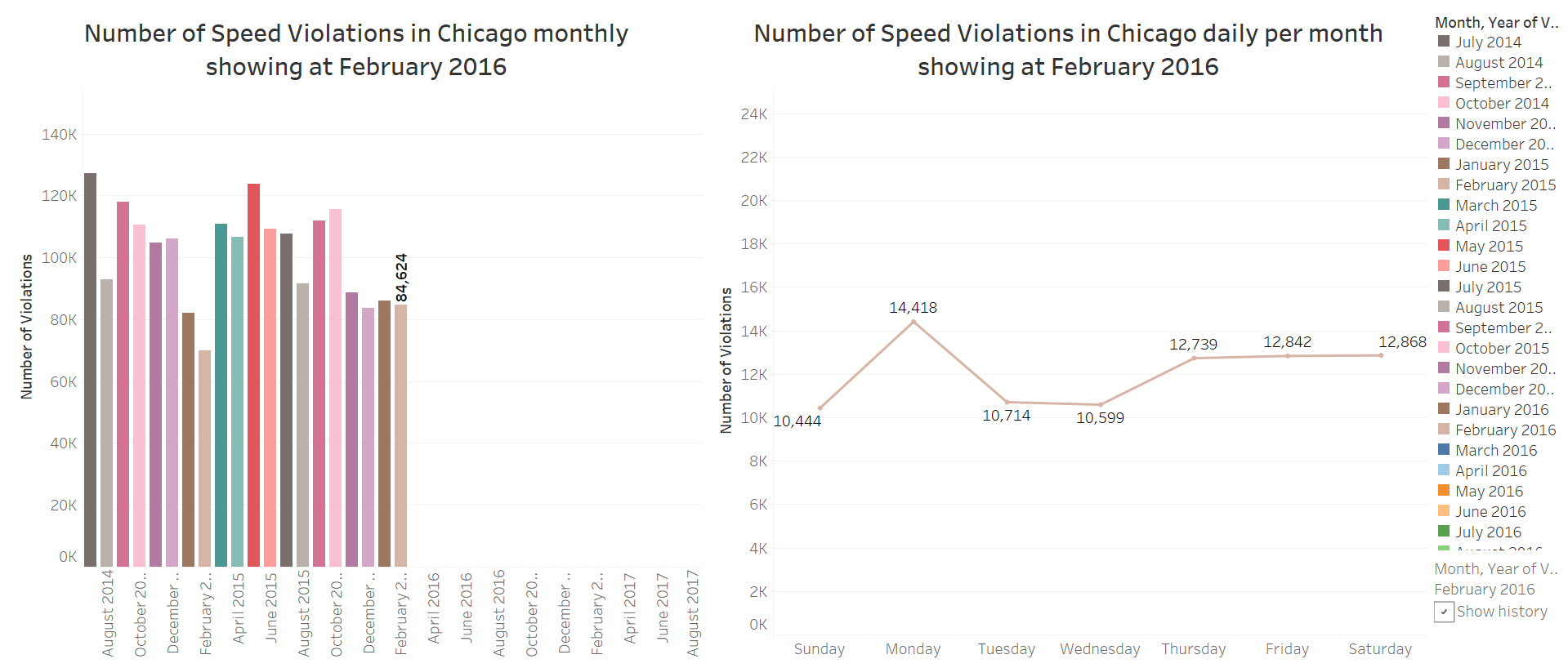


Figure 4

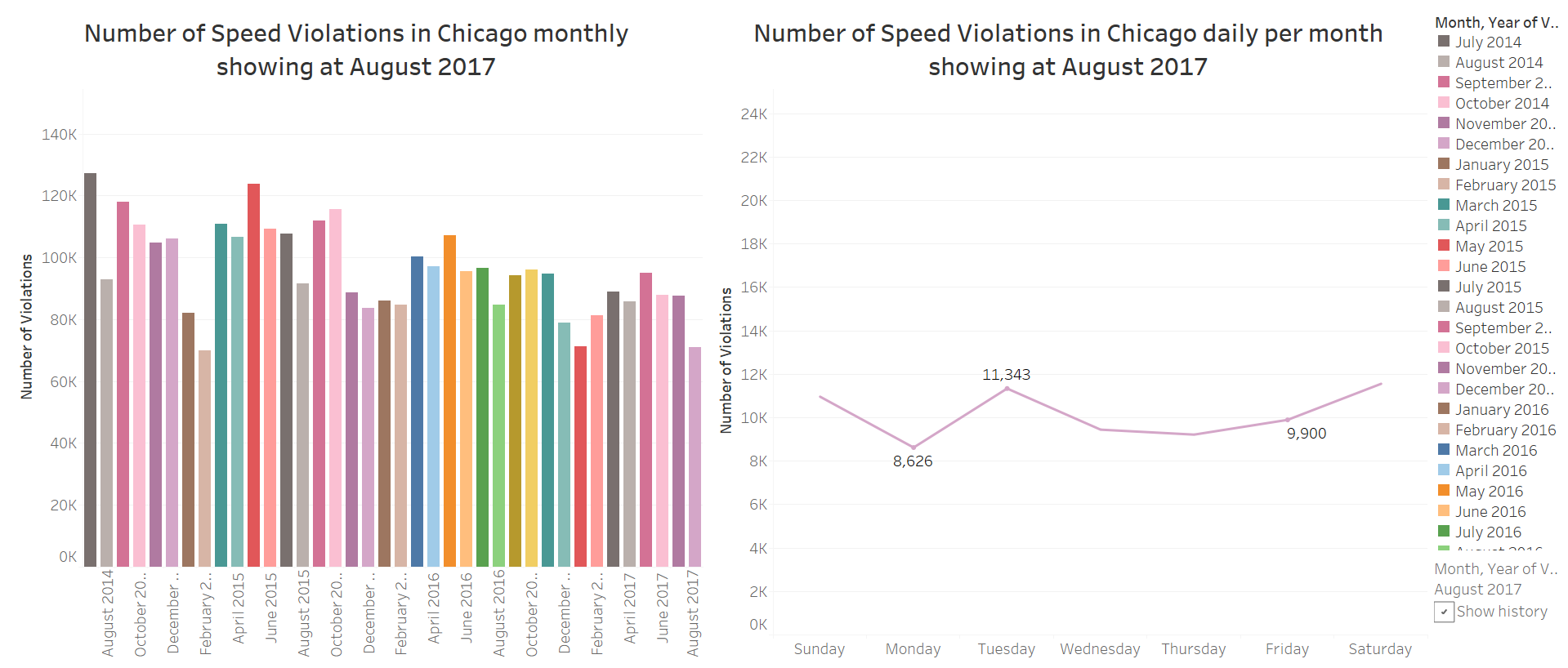


Figure 5

Figure 3, 4 and 5 illustrate 3 random months of 36 months from July 2014 to August 2017 when speed violations occurred. The bar graph presents the total number of speed violations monthly while the line graph is broke down into detail by days in week. Viewers can use both graphs to instantly gather the information about total number of speed violations in each month and the performance of speed violations frequency per day in a specific month; or, they may be interested in comparing among those months to see the big picture of speed violations in Chicago.

In overall, the lowest number of speed violations is found in February 2015 and highest in July 2014

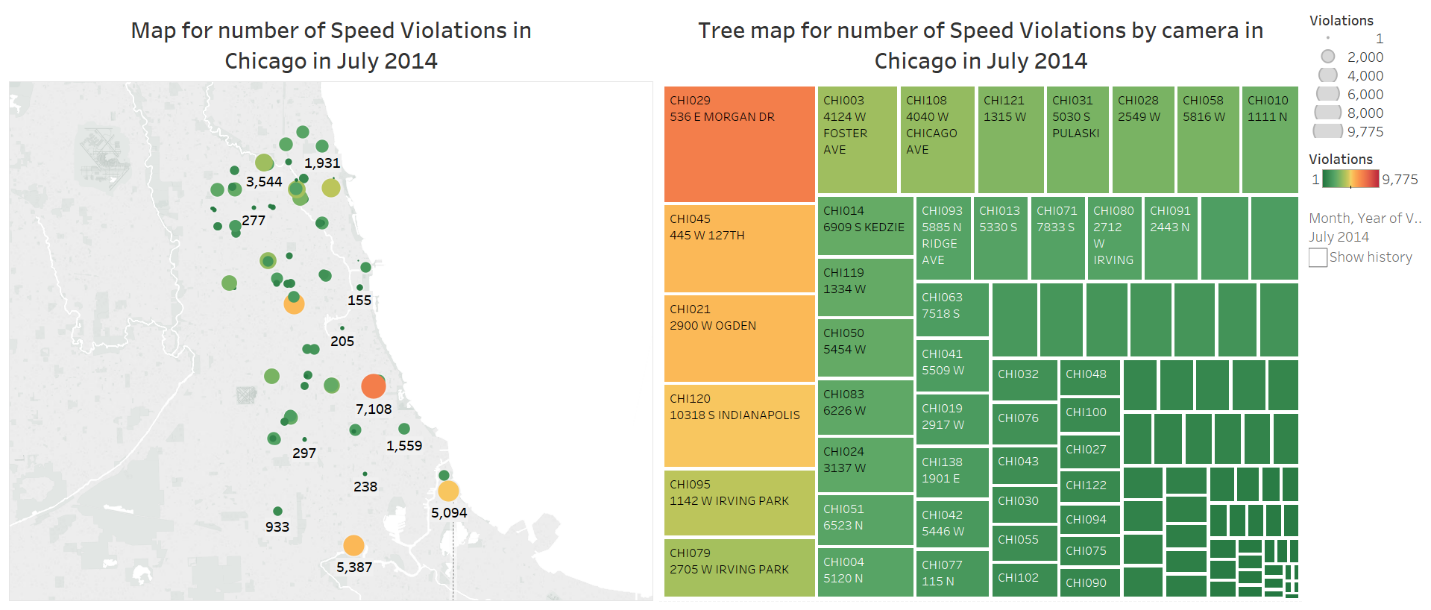


Figure 6

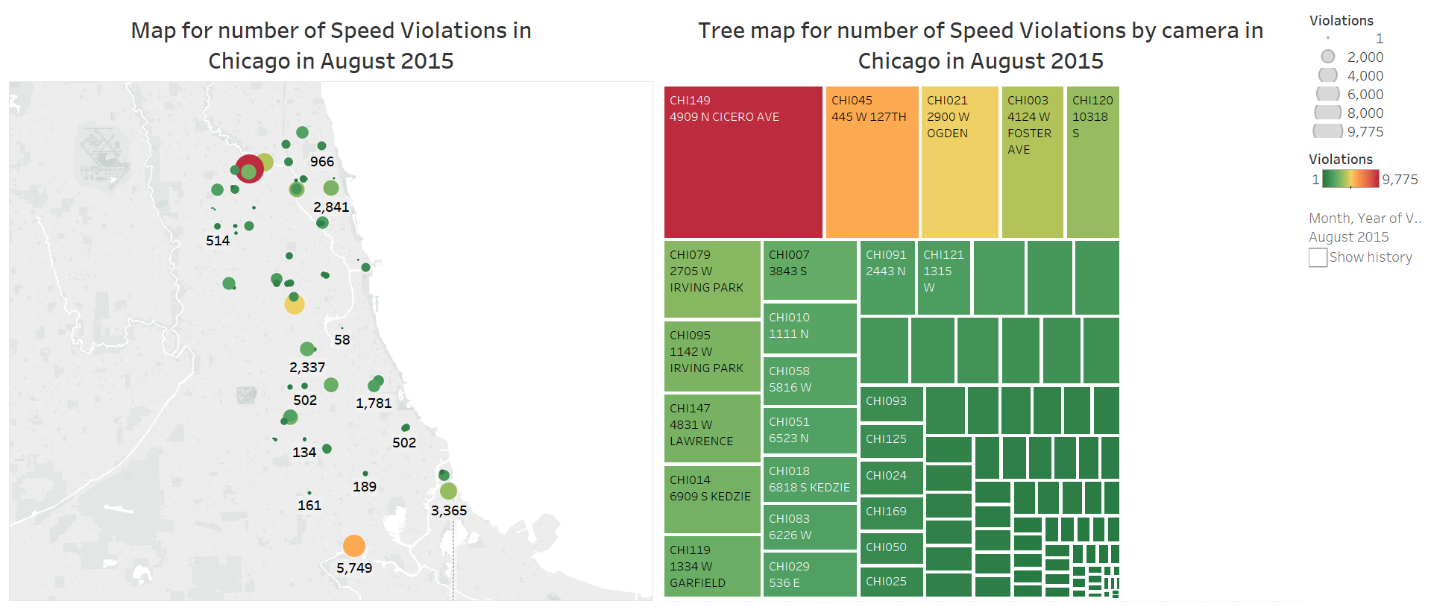


Figure 7

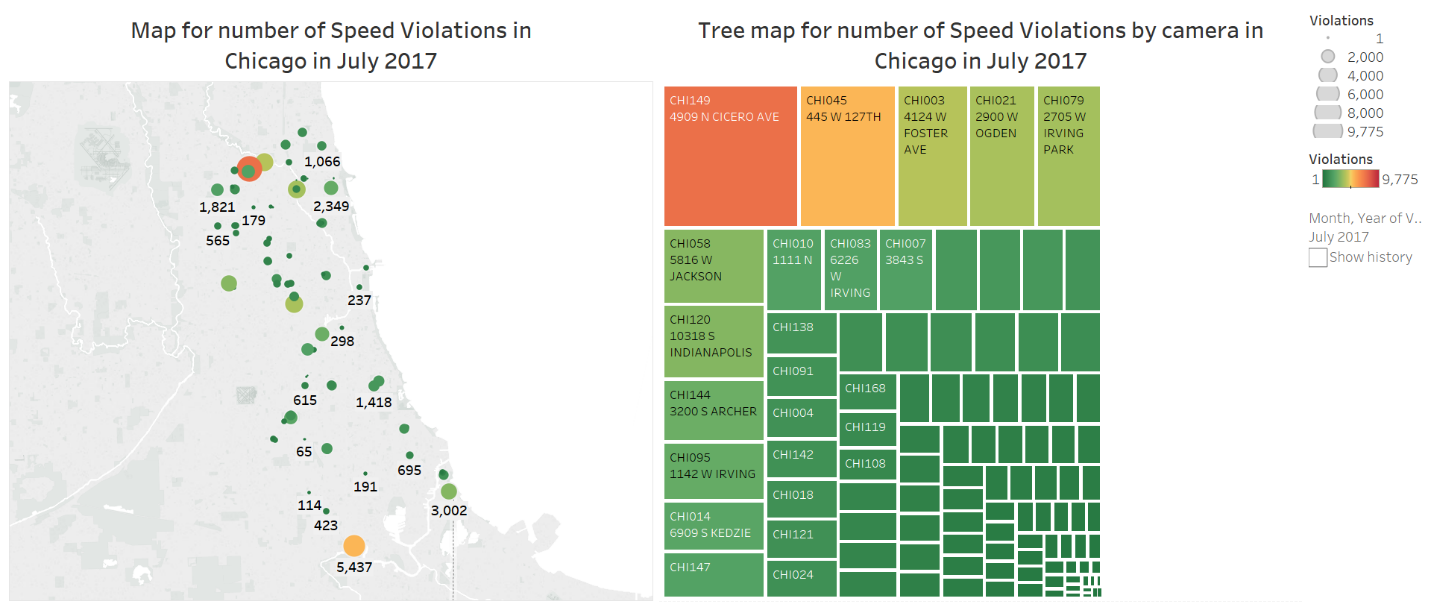


Figure 8

Figures 6, 7, and 8 shows a map with plotted points indicating camera locations and a tree map presenting the number of speed violations captured by the cameras in the city of Chicago in the different months: July 2014, August 2015, and July 2017, respectively.

Each plotted point on the map represents a camera location. Both size and color of those points are used to indicate the number of violations captured by that camera in a month. The bigger the point the more speed violations occur in the neighborhood. Similarly, a diverging color palette changing from green to red shows the variation of violations from less frequent to more frequent. By looking into the map, we can also visualize the distribution of the cameras on over the city. The density of cameras is high in the North and sparser to the South of the city. There are two cameras that appear to have the most speed violations constantly across the three periods: one located in the North, and the other one in the South. We can retrieve more information of those cameras from the tree map.

The tree maps show the cameras along with their address. Size and color have the same scale and present the same information as the map. With the tree map, we can easily identify which camera captures the most violations every month. By looking the tree map in Figure 8, we can see that they are cameras #149 and #045, which are located at 4909 N Cicero Avenue and 445 W 127 TH, respectively.

**3. Conclusion:**

By viewing the visualizations above, the viewers can obtain and extract more informative insights about the cameras and their operations in the city of Chicago. It provides from basic information such as geography locations on the map, number of speed violations in the neighborhoods, or which cameras captured the most violations in each period. In addition, they also deliver more advanced knowledge including the distribution of cameras, the variations and seasonal patterns in number of speed violations for different regions throughout the years. It allows one to make inference and prediction about the locations that have the higher frequency of speed violations, and from that being able to make appropriate decisions.