Each type of incident shows a different pattern of occurrence during a day.

We examined incidents from June through August in San Francisco. The number of daily incidents fluctuated around a mean value of M~315 with a standard deviation σ~35.6 (Figure 1A). No significant trend like linear increase or decrease was observed during this time. The average of daily incidents remained constant. All values in Figure 1A are used for the box plot in Figure 1B. The median is represented by the red solid line with a value m~308. The four quartiles from the first to the fourth correspond to black flat lines denoted by Q_1, Q_2, Q_3, and Q_4, respectively. The second and the third quartiles (Q_2 and Q_3) are within one sigma range from M–σ to M+σ (blue circles), and appear to be almost symmetrical around the median. The median value of m~308 is nearly the same as the mean value M~315 (red square in Figure 1B). All of these aspects indicate that the incident data can be described by a typical bell-shaped distribution.

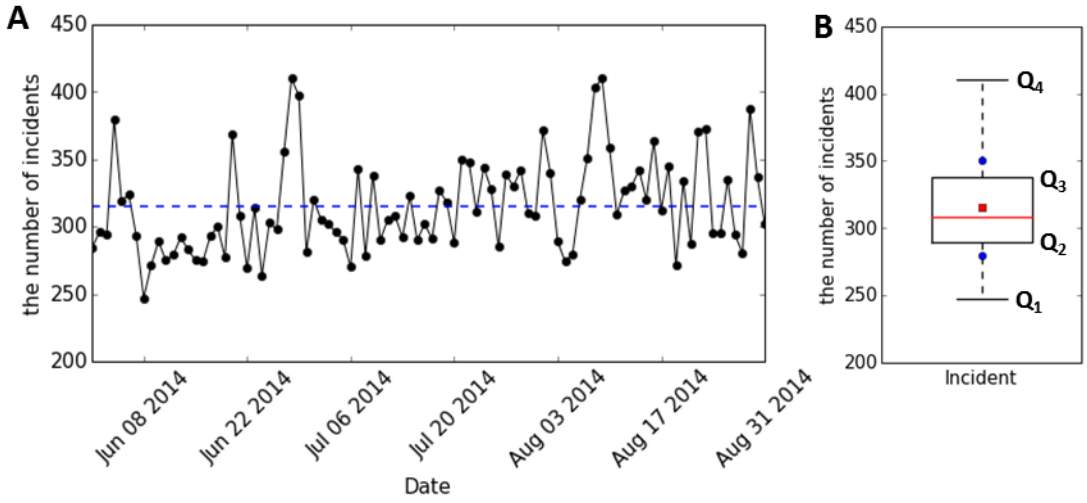

Figure 1. Incident data in San Francisco from 06/01/2014 to 08/31/2014. (A) The number of daily incidents fluctuates around a mean value M~315 (dashed line) with a standard deviation *σ*~35.6. (B) Box plot of daily incidents. Median m~308 (red solid line) is close to the mean (red square). The second and the third quartiles (Q_2 and Q_3) are within one sigma range (blue circles).

Next, we examined the number of incidents in terms of day of the week. Box plots of the incident numbers on each day are shown in Figure 2. The median value on each week day is similar to the value m~308, but it increases as the week progresses. For example, the median values on Friday (328) and on Saturday (318) are higher than the other values. This aspect may be related to increased night-time and economic (paydays) activity. Furthermore, incident numbers on Friday and on Saturday are more widely distributed including the highest number of incidents. This indicates not all weekends are equal in terms of available sports, community and social events affecting the entire community.

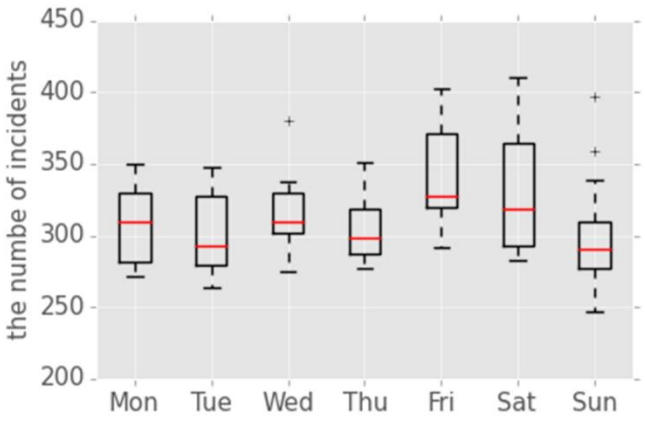

Figure 2. Box plot of the number of incidents in terms of day of the week.

We calculated the average number of incidents per hour on each day of the week. The values are shown by hour from 0 to 23 in Figure 3. Each dashed line corresponds to day of the week. The average values for the week are shown as a black solid line. The average curve showed a pattern of 24 hour periodicity which the curves of day of the week followed with an exception of late night on Friday and Saturday. The number of incidents from 20 to 23 hours on Friday and on Saturday diverged from the average due to late night activity. In addition, we observed a peak at 12, probably due to a lag between the incident event and the discovery of the event with people reporting morning incidents at lunch time. A similar peak might occur at 18 as people returned home from work.

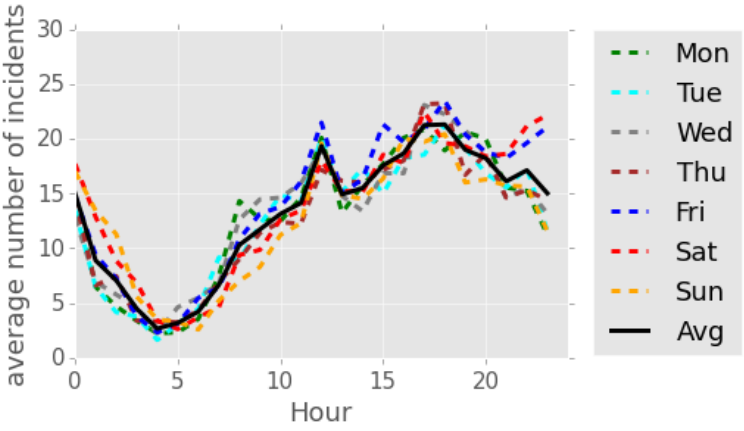

Figure 3. Average number of incidents per hour. Dashed lines are for day of the week. Their average is shown by a solid line.

We identified the four most frequent incidents which are larceny/theft, other offenses, non-criminal, and assault. The procedure for Figure 3 was applied to each of these types (Figure 4). Larceny/theft pattern was similar to the average pattern of Figure 3 because this category dominated the whole data; it occurred nearly three times more than other offenses. These are crimes against property and are most often discovered after the event occurred. Compared to larceny/theft, the other three types showed a more evenly distributed pattern during day-light hours. But assault occurred more frequently at night and on weekends when conflict between people may happen more.

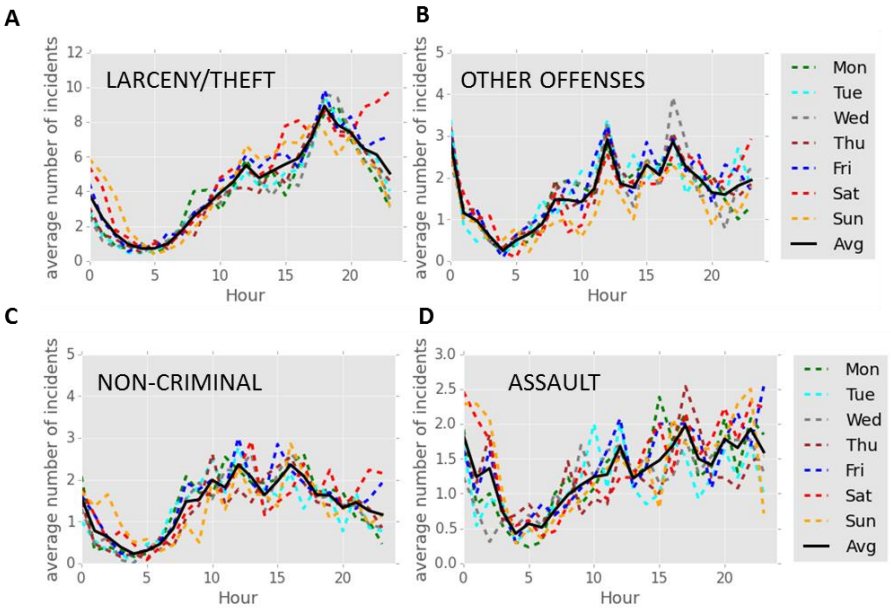

Figure 4. Average number of incidents per hour for the top four types of incidents. (A) Larceny/Theft (B) Other Offenses (C) Non-Criminal (D) Assault.

In summary, the number of incidents per day could be described by a bell-shaped distribution. It correlated with human activity characterized by a 24 hour periodicity. However, each type of incidents showed slightly different patterns depending on time because they occur in different ways.

This result was produced in python using pandas, numpy and matplotlib.