

San Francisco Crime Prediction

ww44ss

Oct 2014

Synopsis

Looking at recent San Francisco crime data this analysis addresses the questions:

- What hotspots are there where crimes are most prevalent in San Francisco? - Do particular crimes happen more frequently at different times of day?

This data suggest that police patrols can be optimized to specific districts and locations to focus on particular crimes.

Get Data

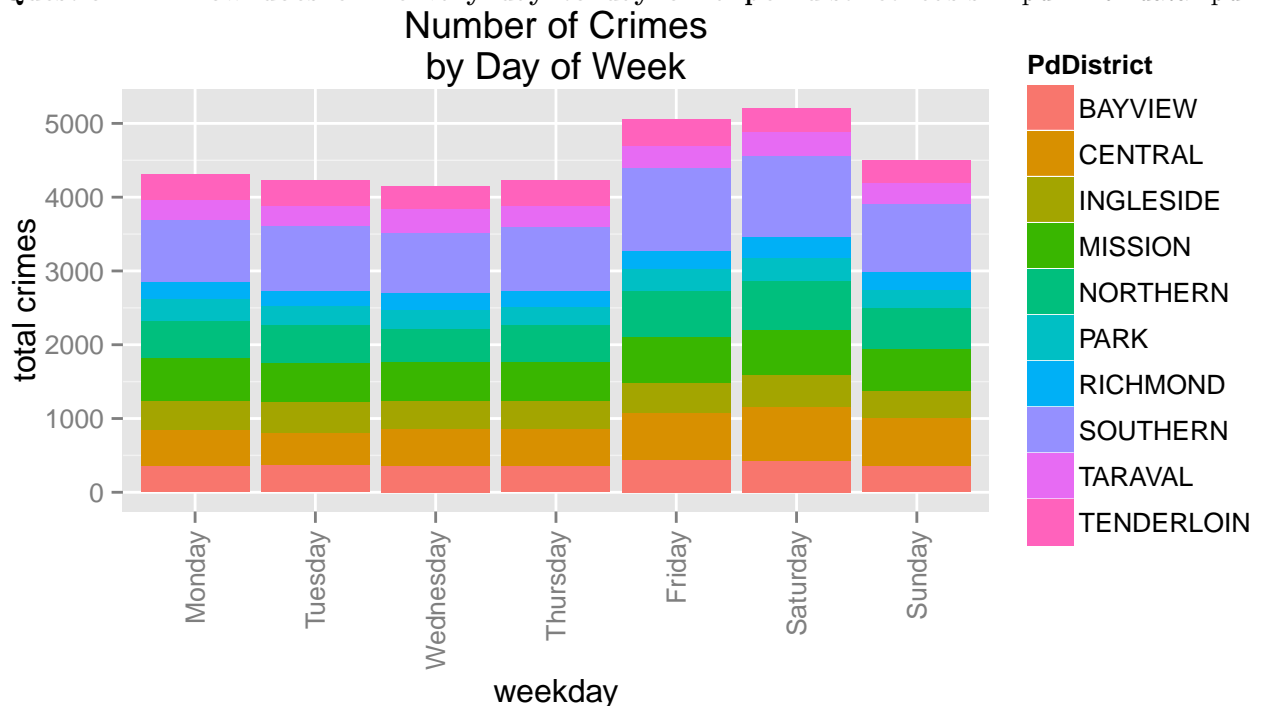
There were 2 files found in the data directory /Users/winstonsaunders/Documents/Crime_Visualization_Challenge.

Data cleaning was pretty straight-forward. Set factors to makes days of week follow standard order (instead of default alphabetical). Convert Date to a r date format. Time I just chose to bucket by hour rather than convert to hh:mm format, which was too fine grained.

The above shows the structure of the data. There are statistics on 31677 crimes in the file datafile.

Analysis

Question 1: How does crime vary day to day on a per district basis? pdf 2 of data2.pdf



Question 2: Does the leading type of crime vary by district? Observing the variability of crime by district its natural to ask whether the nature of crimes show any district by district distinction. The easiest way to get at this is to just pull the data aprt by district and sort. First let’s just look citywide.

```
##              SF
## LARCENY/THEFT 9175
## OTHER OFFENSES 3947
## NON-CRIMINAL  3757
## ASSAULT       2594
## VEHICLE THEFT 1704
## VANDALISM     1668
```

By district the results show some variation.

```
## [1] "MISSION"
```

```
##              ctable
## LARCENY/THEFT    640
## OTHER OFFENSES   559
## NON-CRIMINAL     485
## ASSAULT          420
## WARRANTS         273
```

```
## [1] "RICHMOND"
```

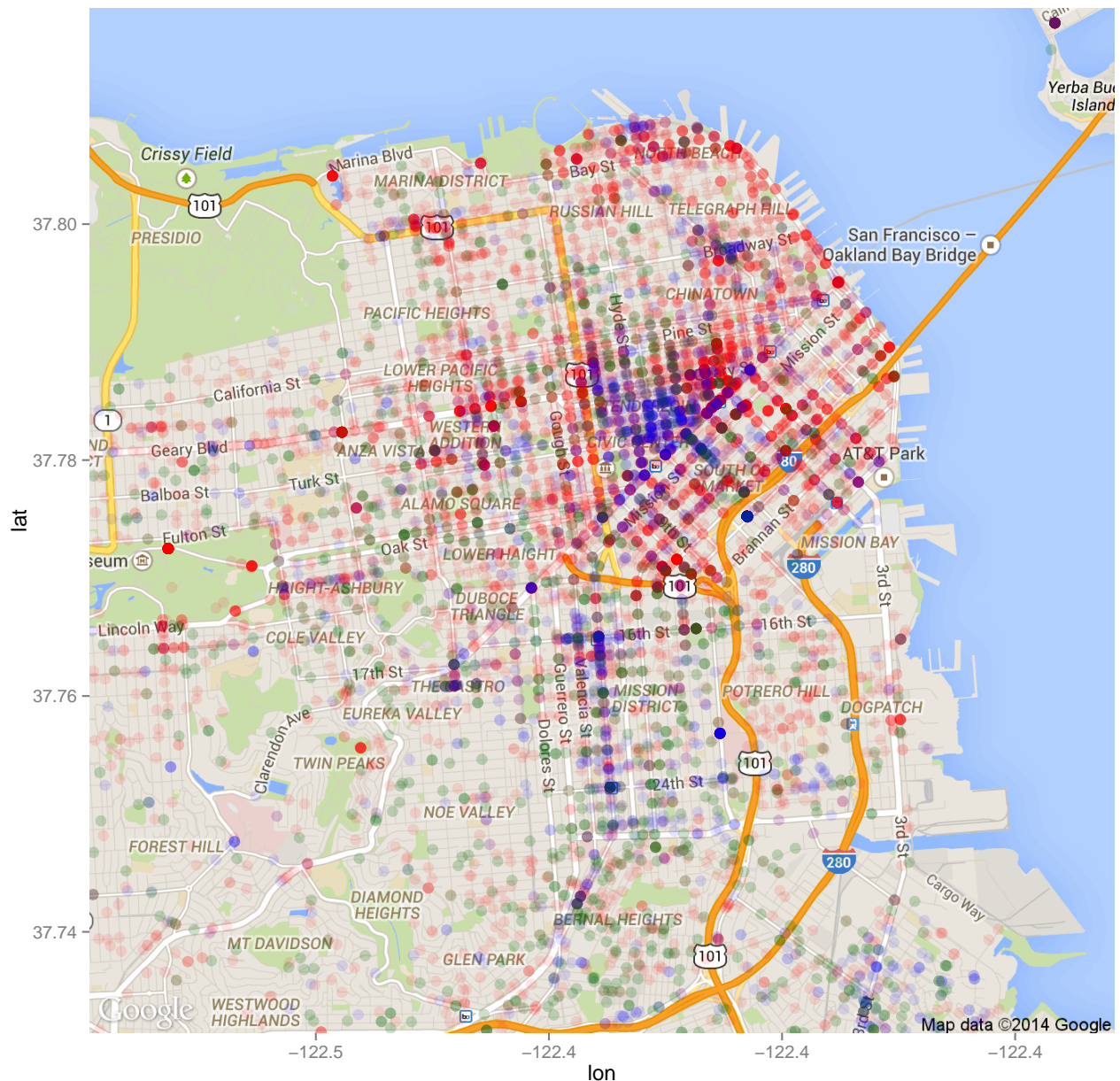
```
##              ctable
## LARCENY/THEFT    533
## NON-CRIMINAL     245
## OTHER OFFENSES   181
## VANDALISM        103
## ASSAULT          92
```

This starts to show some of the richness of the data. For instance in the *Mission District* while Larceny/Theft is the most prevalent item, assault and drugs/narcotic violations together account for more total crime than the does Larceny/Theft.

In the *Richmond District*, by contrast, Assault is not among the top six items, while vandalism and vehicle theft together account for less than half of the leading crime, again Larceny/Theft.

Hence, although the leading type of crime does not vary by district, the top crimes shows marked variation depending on the district.

Question 3: Are there crime hotspots? Here the hypothesis is there are “hot spots” where specific crimes tend to be localized. We can answer this by plotting crime types geographically. The easiest way to see this is to map the results. To speed up analysis I’ve chosen to focus only an a few “top” crimes from the lists above. Namely Larceny/Theft, Vehicle Theft, and Assault.



```
## pdf
## 2
```

Clear hotspots are visible

The Map shows locations of crimes,

red data points correspond to thefts: these appear to be localized to mainly tourist areas.

blue data points representing Assault appear localized in the Tenderloin, Mission, and Broadway areas.

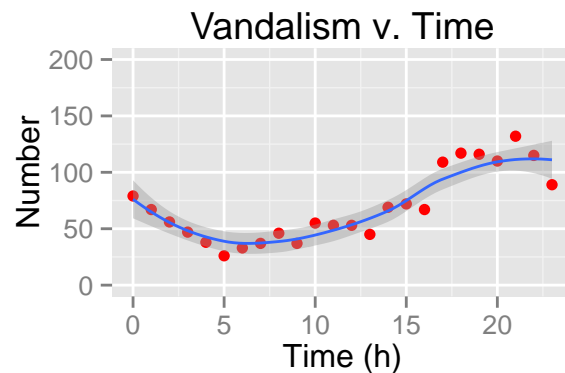
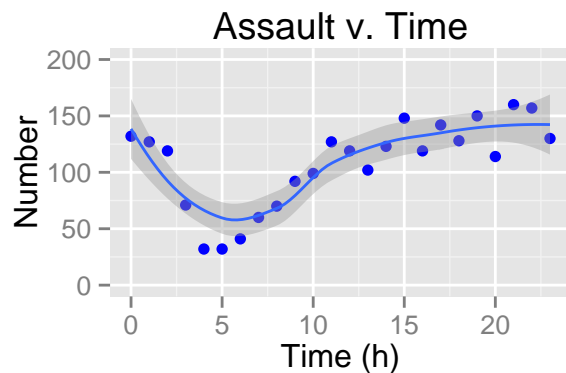
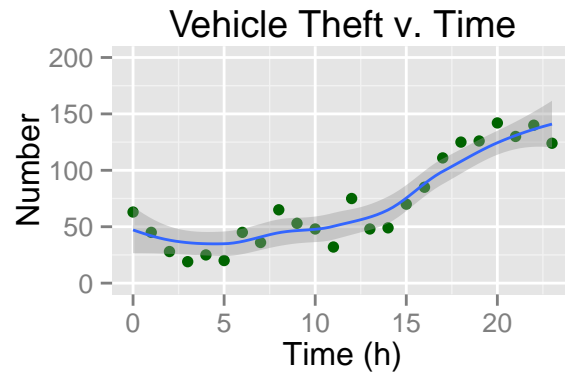
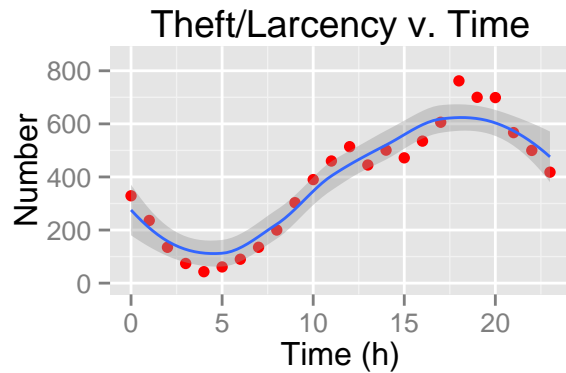
DarkGreen data points representing Vehicle Theft are more spread across the City but appear most prevalent in residential areas.

Question 4: Do crimes have unique time correlation signatures? Let's first look at Larceny data:

This looks very different from the Assault data below.

And vehicle theft shows an even more pronounced behavior.

```
## Loading required package: grid
## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to cl
## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to cl
## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to cl
## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to cl
```



```
## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to cl
## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to cl
## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to cl
## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to cl
```

```
## pdf
## 2
```

```
## NULL
```

Crimes seem to show distinct time behavior. For instance Theft and Larceny appear to be low during morning hours, but peak around 6 pm. Vehicle theft , on the other hand, picks up only after about 6 pm and drops off after midnight.

Conclusions

This quick exploratory analysis found that crime frequency and type vary strongly by location in the city and also by time of day. Taking the data at face value, it suggests that plic patrols could be optimized for time and location, especially when targeting specific crimes.

There is some interesting analysis that could be done as a follow-up. For instance looking deeper at the time/location correlation of specific crimes. This data could be used to test the effectiveness of particular patrol and enforcement strategies.