

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 1 files found in the data directory /Users/winstonsaunders/Documents/Crime_Visualization_Challenge.

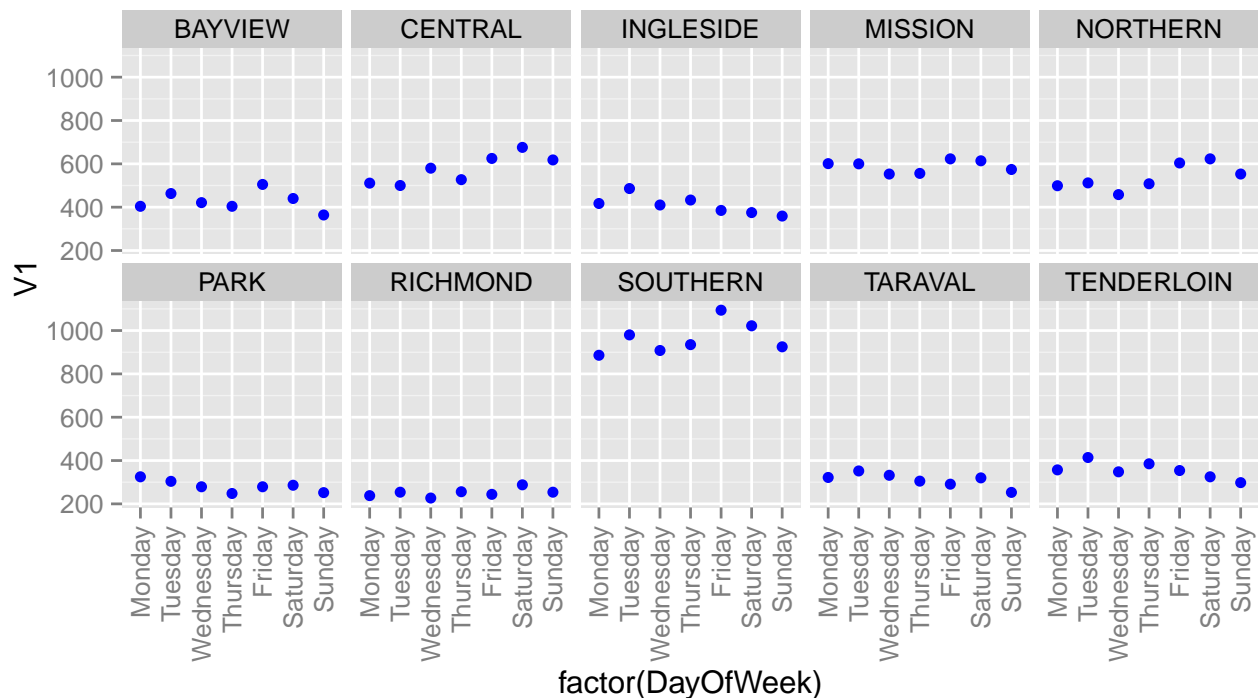
Data cleaning was pretty straight-forward. Set factors to makes days of week follow standard order (instead of default alphabetical). Convert Date to a r date format. Time I just chose to bucket by hour rather than convert to hh:mm format, which was too fine grained.

```
## 'data.frame': 32921 obs. of 12 variables:
## $ IncidntNum: int 140622186 140741225 140593098 140644839 146195066 140662825 140549580 140562902
## $ Category : Factor w/ 36 levels "ARSON","ASSAULT",...: 17 17 33 17 33 21 2 2 17 22 ...
## $ Descript : Factor w/ 418 levels "ABANDONMENT OF CHILD",...: 206 205 245 278 248 196 124 74 277 15
## $ DayOfWeek: Factor w/ 7 levels "Monday","Tuesday",...: 6 3 6 7 7 6 3 1 1 6 ...
## $ Date : Date, format: "2014-07-26" "2014-09-03" ...
## $ Time : num 20 9 18 11 14 7 16 13 9 9 ...
## $ PdDistrict: Factor w/ 10 levels "BAYVIEW","CENTRAL",...: 2 7 9 8 2 6 1 8 9 6 ...
## $ Resolution: Factor w/ 16 levels "ARREST, BOOKED",...: 12 12 12 12 12 12 12 1 12 2 ...
## $ Address : Factor w/ 8867 levels "0.0 Block of 10TH ST",...: 4423 4821 2119 2676 4889 7847 5532 5
## $ X : num -122 -122 -122 -122 -122 ...
## $ Y : num 37.8 37.8 37.8 37.8 37.8 ...
## $ Location : Factor w/ 13201 levels "(37.7080829769597, -122.419241455854)",...: 11424 6975 4396 13
```

The above shows the structure of the data. There are statistics on 32921 crimes in the file datafile.

Analysis

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



To analyze ply the data apart by day and district. At this point the analysis looks at total crime reports only. Later I'll look at types of crimes.

District by district crime rates show variation in the day of the week. Each district has some pretty unique variation. Some of the more interesting ones are listed below. - *Southern* has the highest crime rate by far, strongly peaking on Friday night.

- *Bayview* is mostly flat, but seems to show a higher rate on Friday nights.
- *Central* shows a strong upward trend on the weekends, with Friday and Saturday night showing about 20% increase in crime.
- *Mission* while having an overall fairly high crime rate, shows little variation.
- *Tenderloin* shows an apparent drop in the crime rate.

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 9262
## OTHER OFFENSES 4241
## NON-CRIMINAL  3846
## ASSAULT       2691
## VANDALISM     1775
## VEHICLE THEFT 1762
```

By district the results show some variation.

```
## [1] "MISSION"

##          ctable
## LARCENY/THEFT   657
## OTHER OFFENSES  589
```

```
## NON-CRIMINAL      490
## ASSAULT           413
## WARRANTS          297
```

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

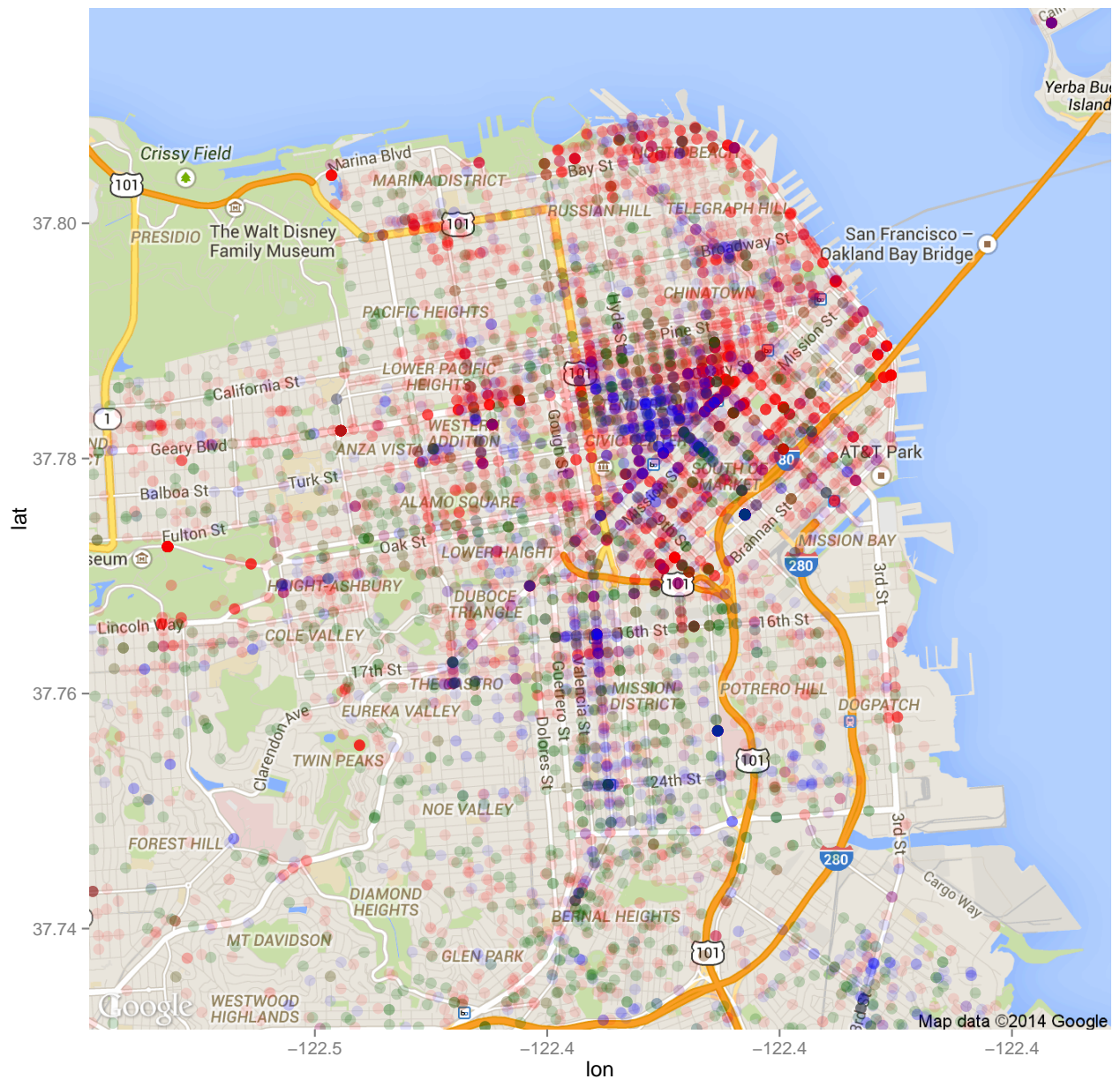
```
##              ctable
## LARCENY/THEFT    560
## NON-CRIMINAL     248
## OTHER OFFENSES   219
## VANDALISM        111
## VEHICLE THEFT    100
```

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.



Clear hotspots are visible

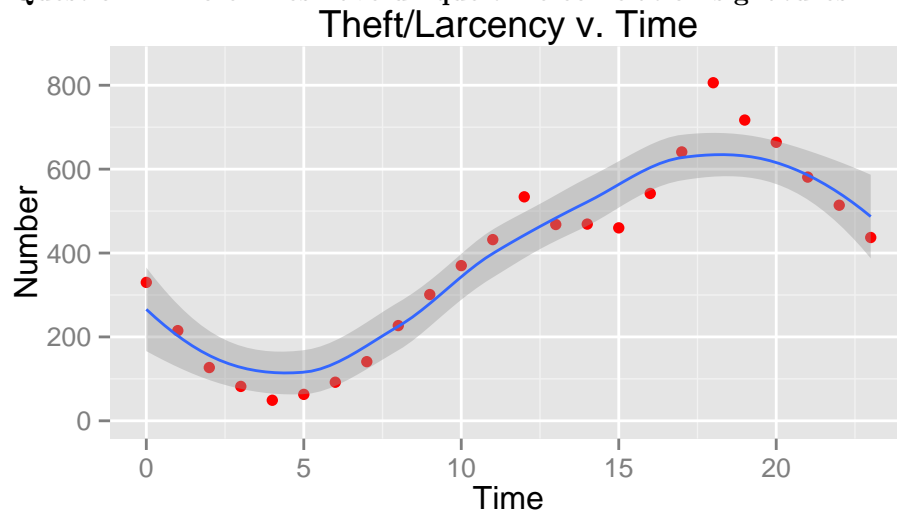
The Map shows locations of crimes,

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

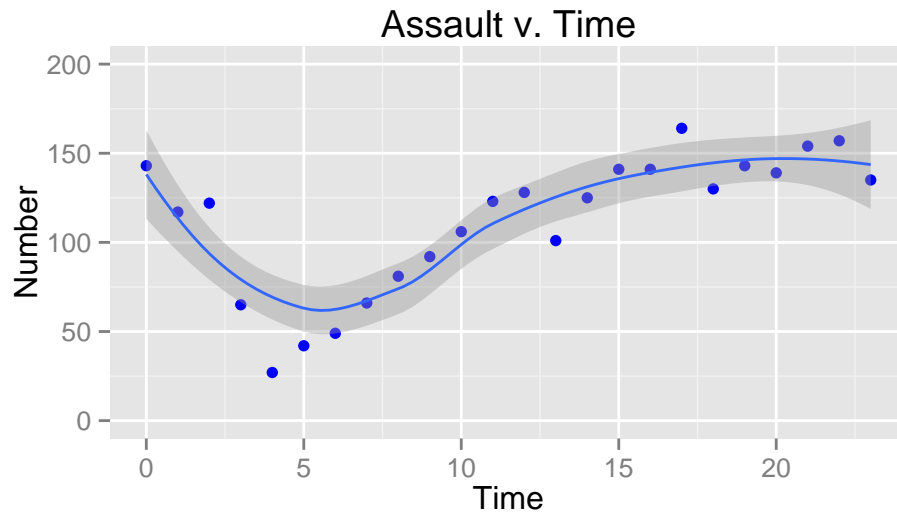
blue data points representing Assault appear localized in the Tenderloin, Mission, and Broadway areas.

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

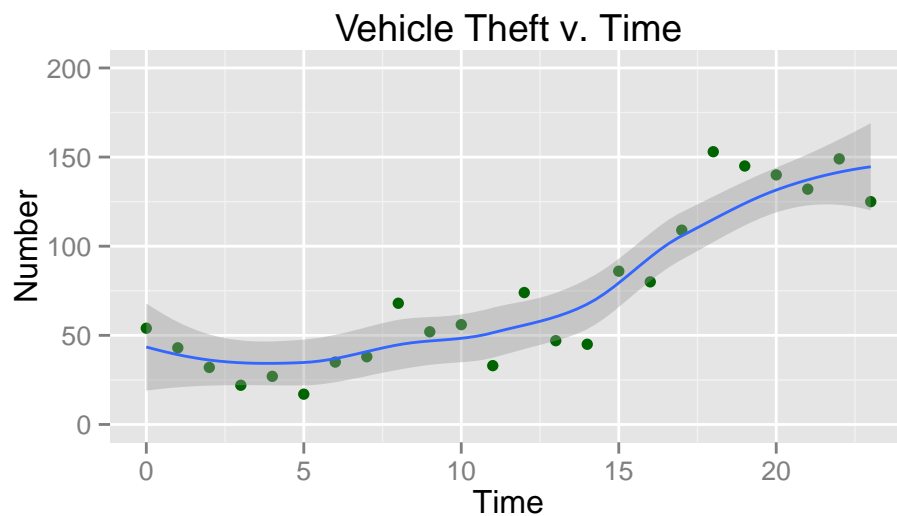
Question 4: Do crimes have unique time correlation signatures? Let's first look at Larceny data:



This looks very different from the Assault data below.



And vehicle theft shows an even more pronounced behavior.



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.