

Visual exploration for data science course

Alex Ishkin

Saturday, December 12, 2015

The goal of this study is to explore crime patterns in San Francisco and Seattle using visual means. All graphs below are produced using `ggplot2` package in R.

My point of interest was: how are different crime types distributed spatially, across city's districts and neighborhoods. Are there district especially rich in thefts? Burglaries? Something else?

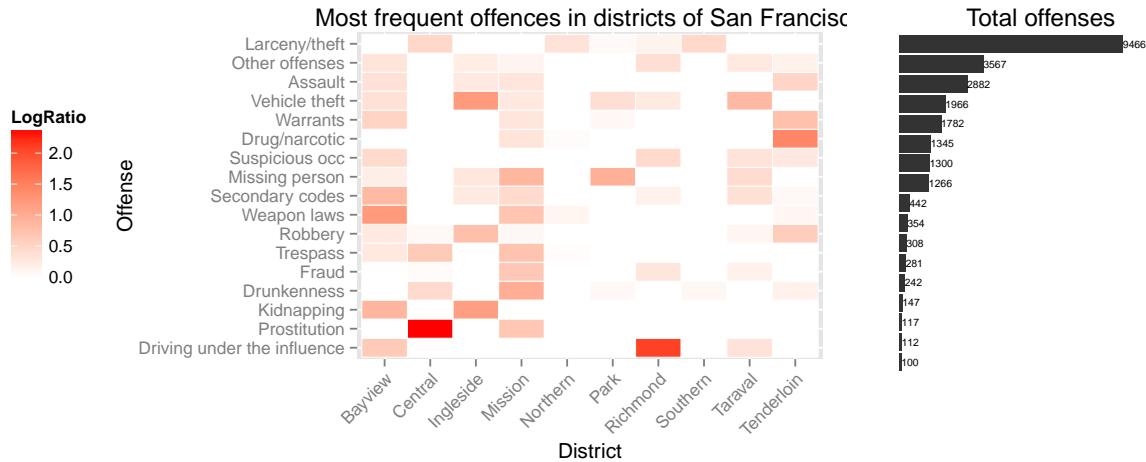
I decided to calculate relative frequencies of crime in each neighborhood. For instance, if thefts constitute 20% of offenses across the city, but 40% of crimes in the downtown, this 2-fold frequency jump is certainly interesting thing.

The overall conclusions are:

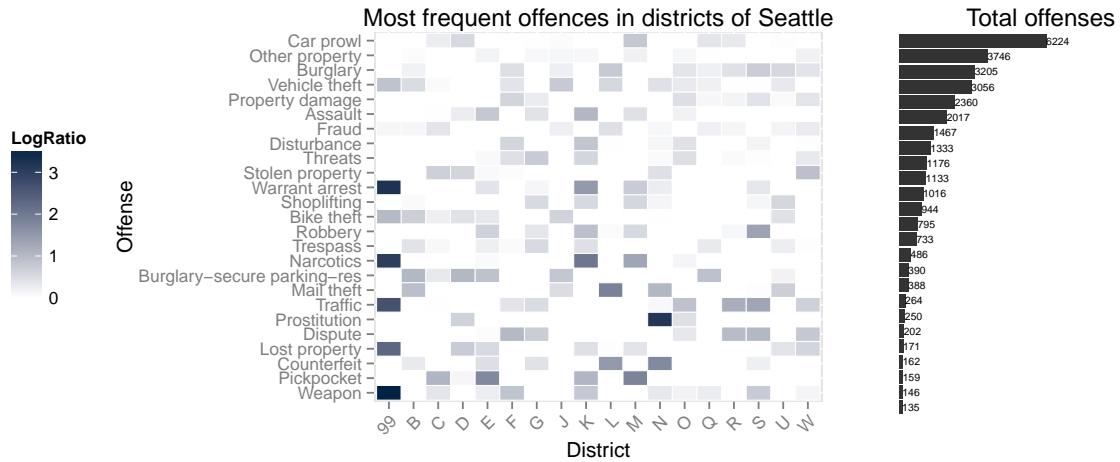
- The biggest problem for both cities appears to be theft in different forms (they are encoded in the data sets differently, but the pattern is clear). There are no areas with drastically elevated frequency of thefts.
- Some of offence types (e.g. prostitution) appear to be largely concentrated in one district; moreover, are limited to certain neighborhoods. This is the case for both cities.

The relative frequencies for all offences with at least 100 cases are calculated as log ratios and visualized as heatmaps (offense vs. district) below. Offenses are sorted by number of total cases.

As we can see, top crime (theft) in San Francisco doesn't have regional preference. Surprisingly, the prostitution is heavily localized in Central district (there are few other curious spots, but I do not analyze them in detail)



Seattle shows pretty much the same pattern, with top crime not showing preferences and a couple of bright spots including high frequency prostitution in area N. The area 99 is also curious, with a lot of hot spots on it.



Next, the spatial patterns can be explored in more detail using coordinates of events. As the following plot shows virtually all of Seattle prostitution cases are localized along one street, north of the downtown.



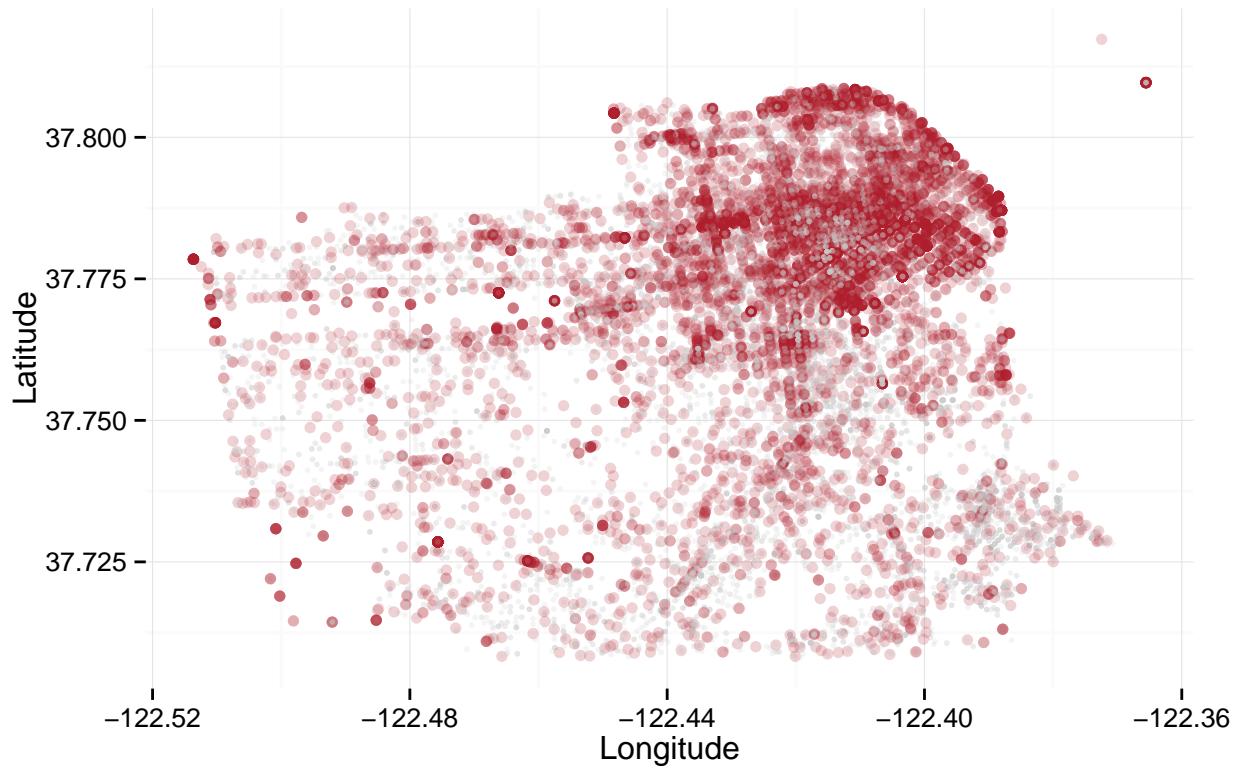
In San Francisco, there appear to be two major foci of prostitution crime: one located in the midst of the downtown, and another one slightly to the south:

Prostitution in San Francisco



The theft, a top offence category in both cities, is prevalent across the board, without any area drastically standing out. The following image shows theft cases in San Francisco. The downtown and Embarcadero areas appear to be theft-heavy, but these areas appear to be crime-prone overall.

Theft in San Francisco



Seattle appears to be even more theft-egalitarian. One should interpret this image with caution, as the Seattle's map only includes car prowls, with other theft offences being separated in other categories such as 'Bike theft' or 'Pickpocket'.

Theft (car prowls) in Seattle

