Crime Analytics: Visualization of Incident Reports

A- Overview

In this notebook, I analyze criminal incident data from Seattle and San Francisco to visualize patterns, contrast and compare patterns across the two cities. This notebook is a visual narrative consisting of a series of visualizations with sufficient descriptive text to make a convincing argument.

B- Project Ideas

- For either city, how do incidents vary by time of day? Which incidents are most common in the evening? During what periods of the day are robberies most common?
- For either city, how do incidents vary by neighborhood? Which incidents are most common in the city center? In what areas or neighborhoods are robberies or thefts most common?
- For either city, how do incidents vary month to month in the Summer 2014 dataset?
- For either city, which incident types tend to correlate with each other on a day-by-day basis?
- (Advanced) What can we infer broadly about the differences in crime patterns between Seattle and San Francisco? Does one city tend to have more crime than the other, per capita? Do the relative frequencies of types of incidents change materially between the two cities? (NOTE: The two datasets do not have the same schema, so comparisons will require some work and some assumptions.)
- (Advanced) For either city, do certain crimes correlate with environmental factors such as temperature? (To answer this kind of question, I use external data sources)

C- Data

This project uses the real crime data from Summer 2014 of the two US cities: Seattle and San Francisco.

These datasets do NOT agree on schema, and they do NOT agree on categories or descriptions for specific crimes. To draw comparisons and contrasts across cities, some assumptions about correspondences should be made. For example, LARCENY/THEFT is a category in San Francisco, but Seattle uses codes of the form THEFT-CARPROWL or THEFT-BUILDING. I need to make some reasonable assumptions to compare these data, and explain and justify these assumptions. Scale.

The data is a subsets of all three datasets during the period Summer 2014.

D- Tools

Python with Bokeh and matplotlib