Introduction

As a former resident of San Francisco, I am very interested in this assignment, and want to explore the data set fully. However given the timescale between the course opening and the holiday period, there hasn't been time to explore it in as much detail as I would have liked.

While resident, my car was burgled several times, and a friend’s car was stolen, so my exploration will focus on this subcategory, in particular the subcategories of:

- Theft and attempted theft of Vehicles
- Theft and attempted theft from Vehicles

Dataset

I noted the availability of the original data set covering crime from 2003 onward in the assignment. Given that this covers the period of time where I was resident, I've taken the liberty of using this data set instead of the smaller sample.

Initial Queries

Some of my initial questions were:

- Do vehicular burglaries vary with time of day, day of the week, or month/season?
- Is there any trend in the number of reported vehicular burglaries since 2003? How does this compare with the overall trend in numbers of reported crimes?
- What are the trends for this by neighbourhood (and how does this compare with my own experience)?
- What are the most common outcomes for these crimes?

Reproducability

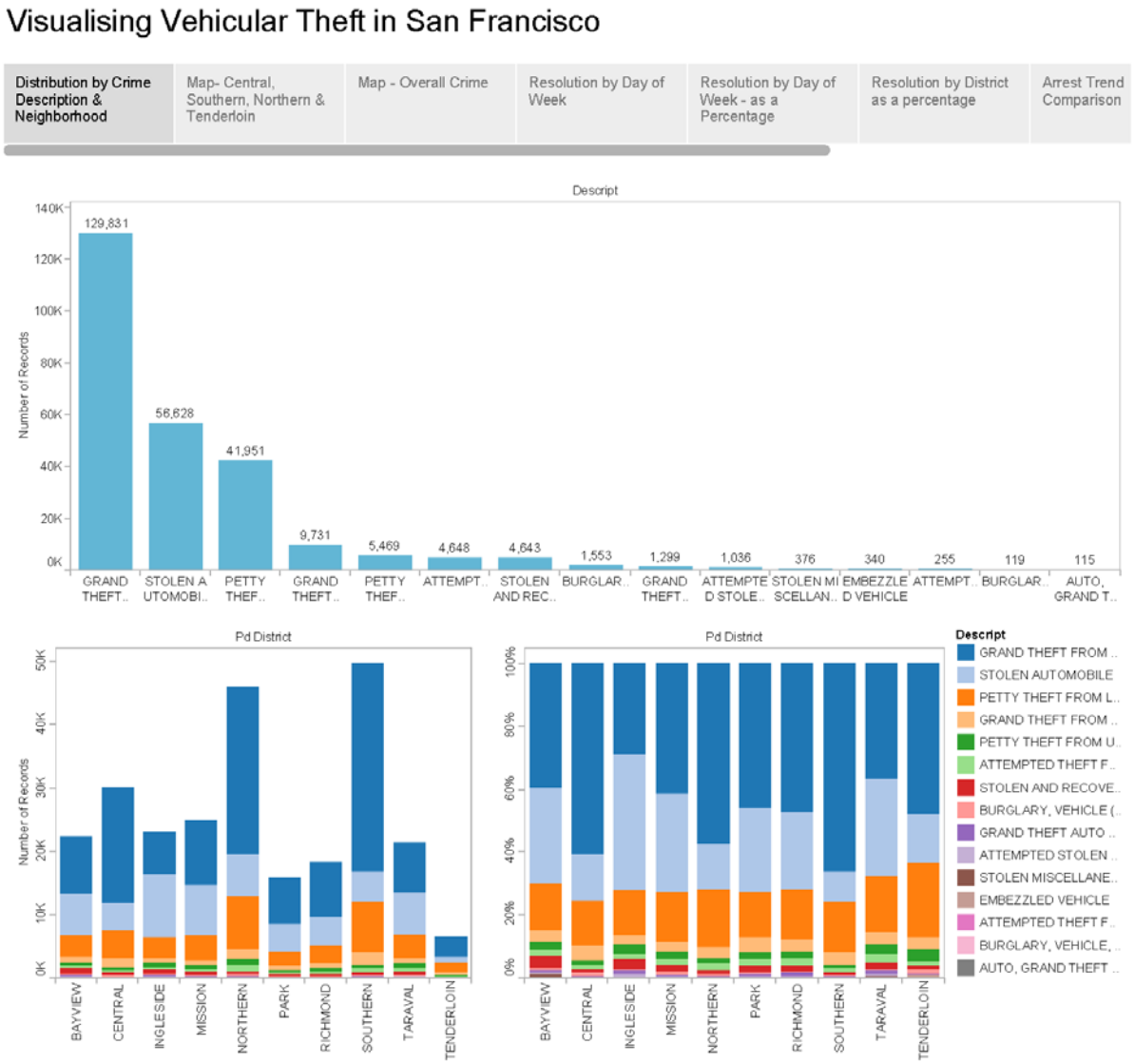
I have included a packaged Tableau workbook. I'm still on the Python learning curve, so given the timescales I chose to use a tool which is much more accessible for someone with limited coding knowledge. Also included is the CSV file downloaded.

Where dates are used, missing dates have been excluded. When year-on-year trend are visualised, results from 2015 have also been excluded (so that only full years are included)

Results

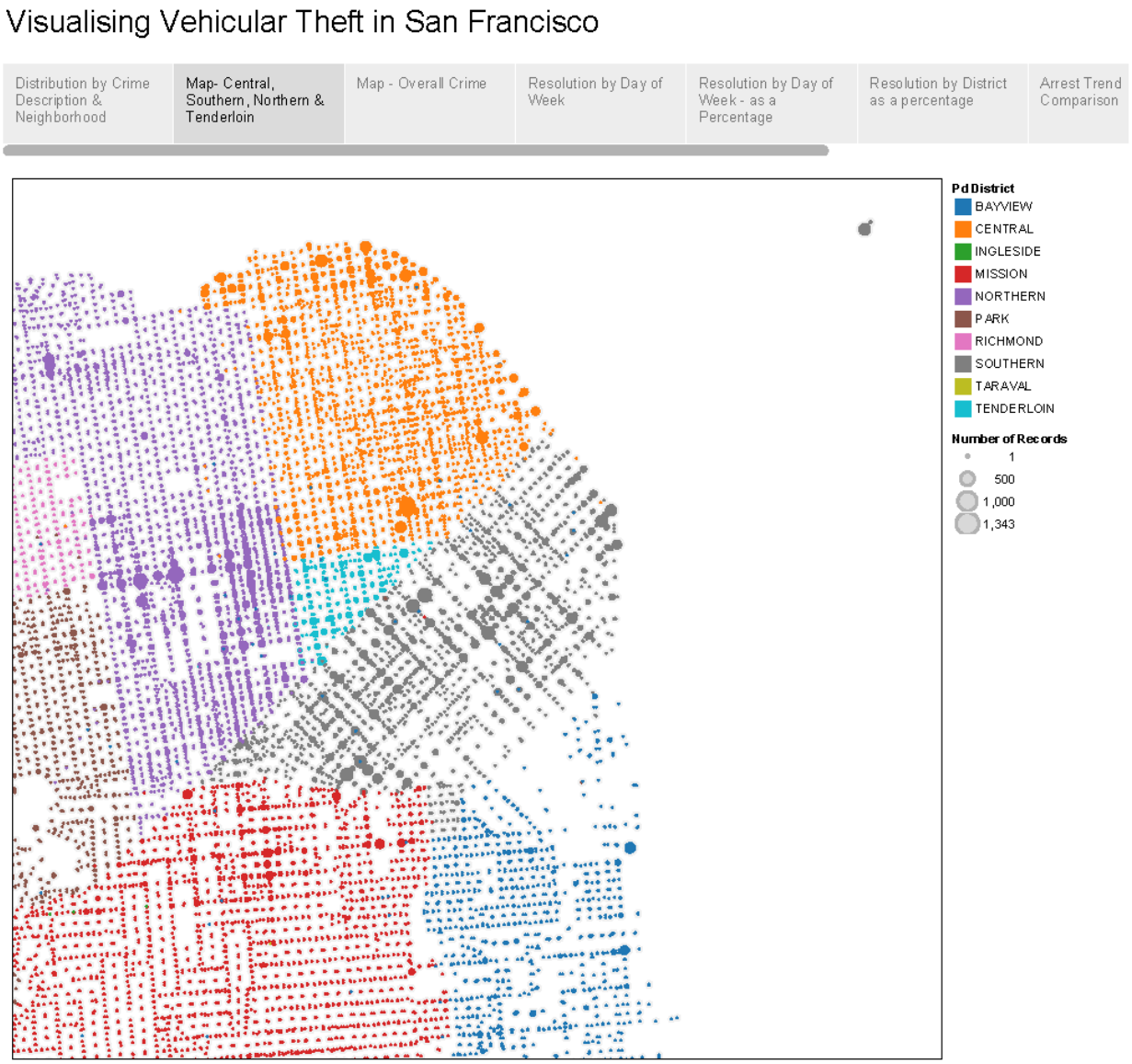
I started by looking at the frequency of occurence for the descriptions included in my defined Vehicle Theft (VT) group. The most common type is Grade Theft from a Locked Vehicle,

followed by Stolen Automobiles. Looking at the various neighborhoods, it is apparent that some neighborhoods have more vehicular theft crimes than others - in particular the Northern, Southern and Central districts have the highest frequencies, while the Tenderloin has few.

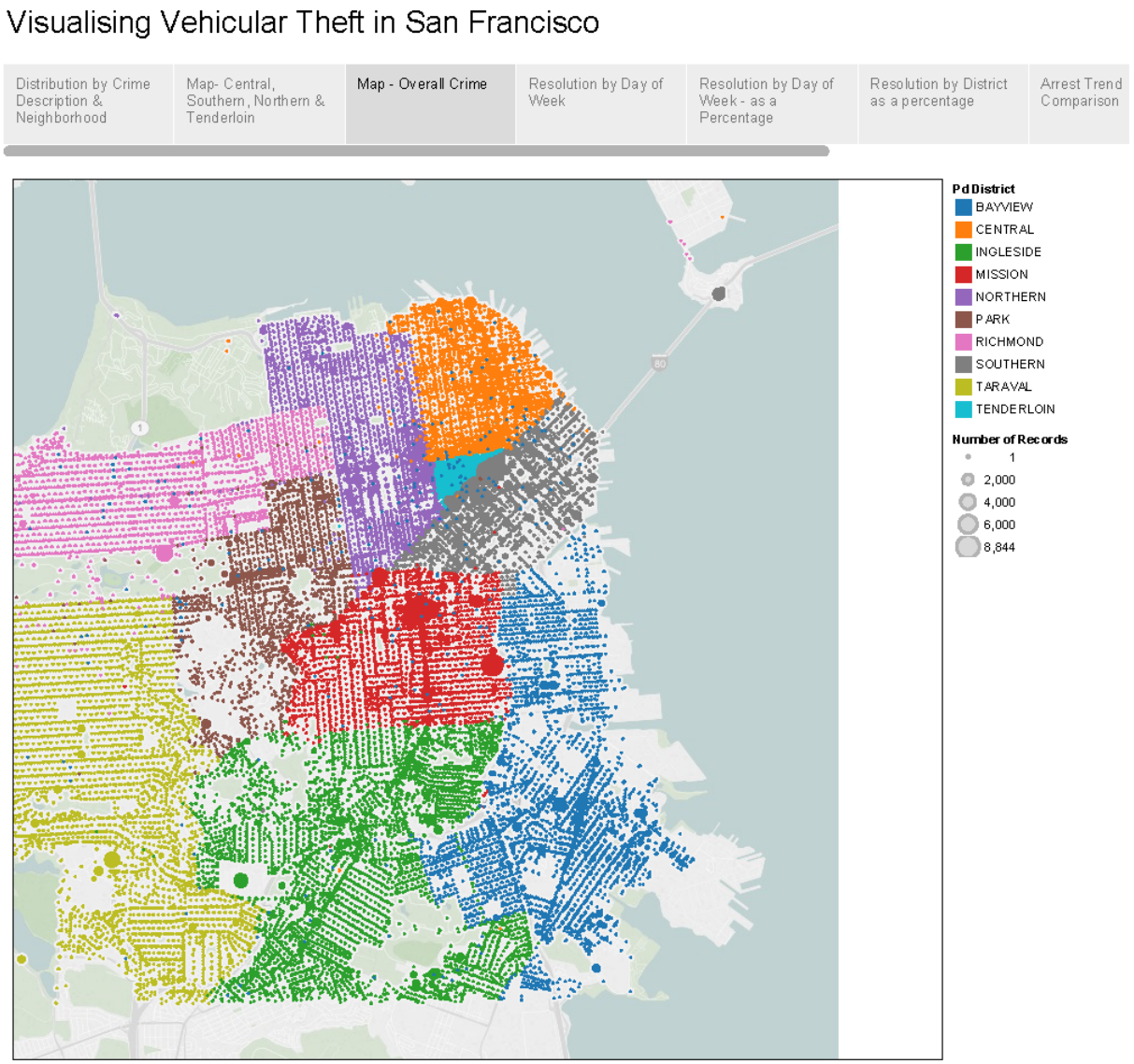


In order to better see the trends, I mapped this, and zoomed in on these four neighborhoods. I immediately noticed a huge cell on Bryant between 6th and 7th. However, a little investigation shows this corresponds to the Hall of Justice, and is likely an error. This disproportionately large cell also reduced visibility of other cells, and I therefore excluded it from the map.

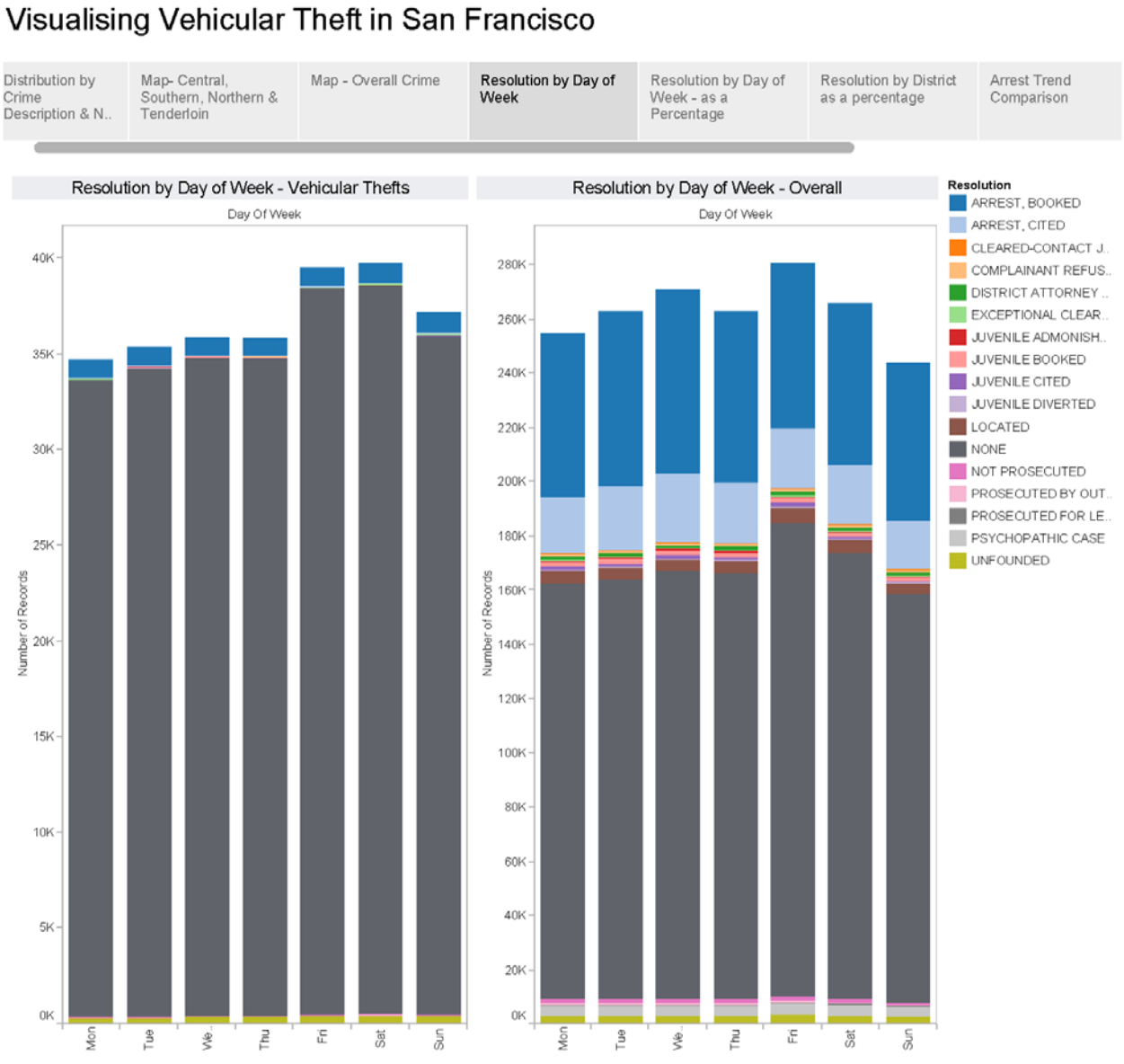
With this map, I can see some clear trends: there is a more concentrated level of crime along Geary (primarily Northern District but extending into Central and Park Districts), and along the Embarcadero (Central District) through to Fishermans' Wharf (northern part of the Central District). In the Southren District, VT crime appears more evenly distributed (barring the node excluded on Bryant).



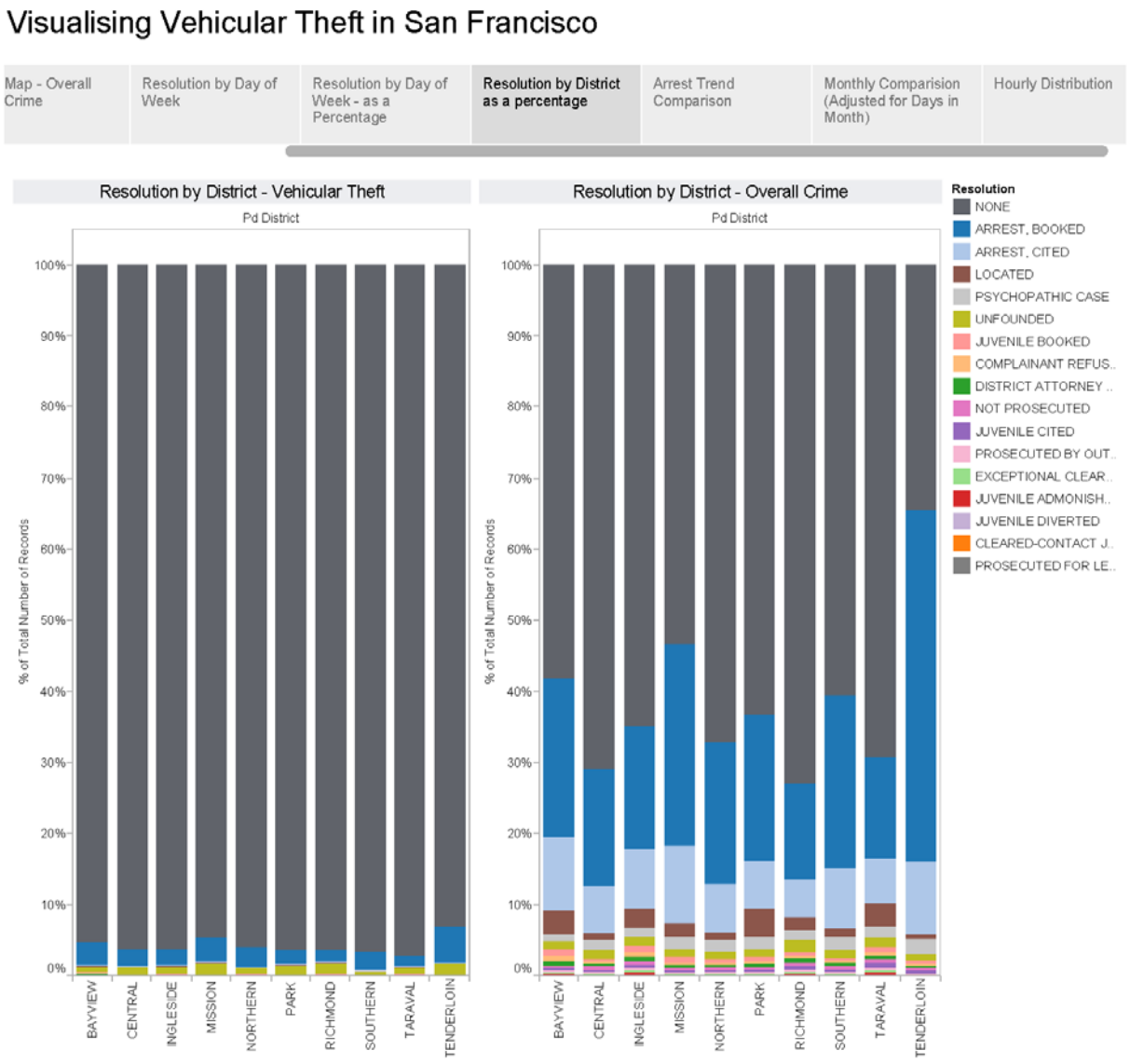
When compared to the overall map of crime in SF (again excluding the Halls of Justice), it is apparent that there is a very different geographic distribution for VT crime as compared to overall trends; overall crime is more concentrated on Market Street (Southern), the Tenderloin, Geary again, Haight Ashbury (Park), and on Mission Street (Mission District). This makes some degree of sense - these areas are busier, with considerable foot traffic so are less likely to be great locations to (successfully) break into a car.



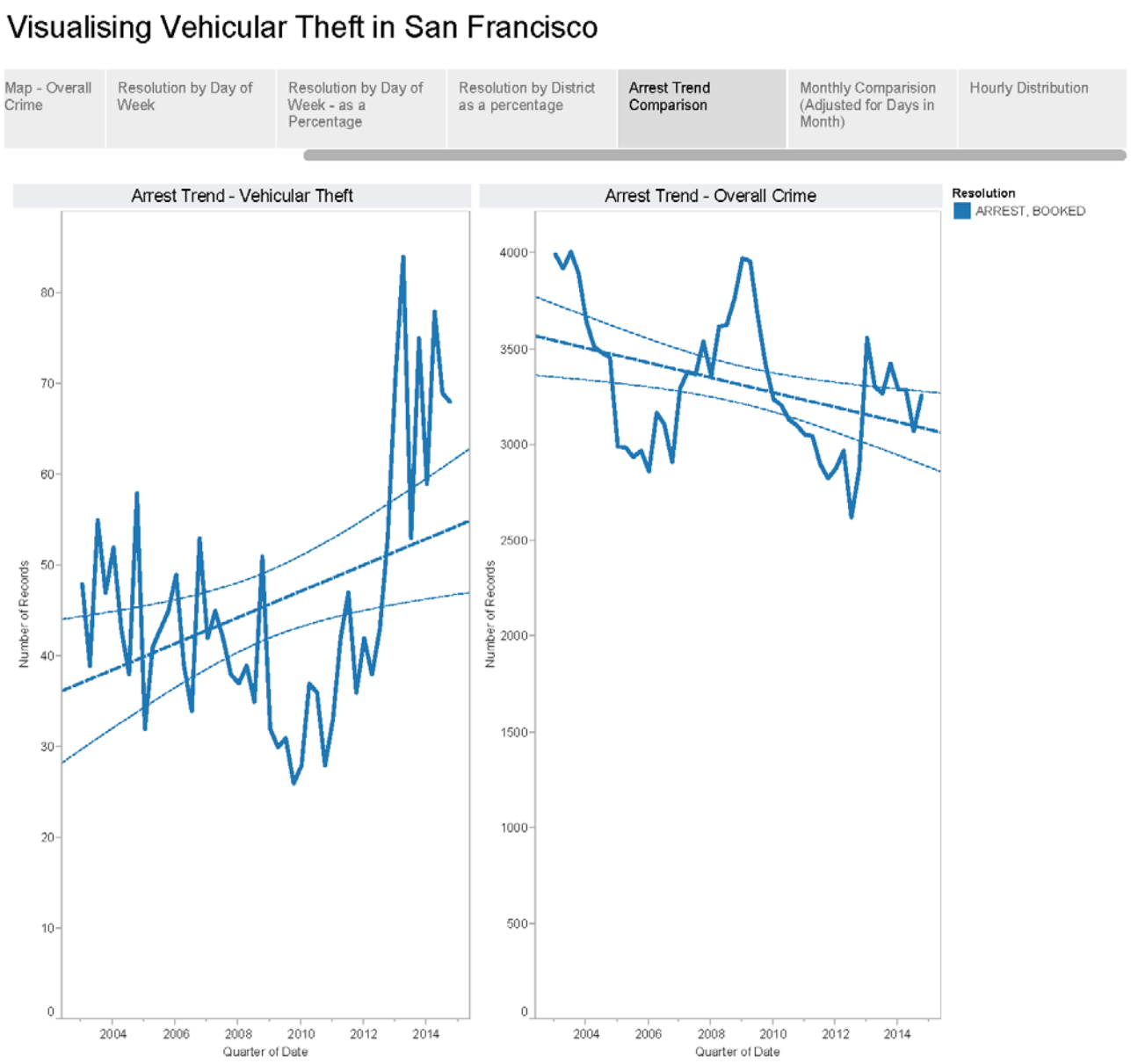
I wondered whether there were any days of the week which witnessed "spikes" in VT crime activity. Visualising this (both for VT crime, it looks like the weekend (Friday - Sunday) sees a small spike in activity while overall crime does not. With VT crime, it's apparent that the vast majority (~96%) have no resolution. Overall, while the majority also see no resolution, there is a relatively higher proportion of arrests made.



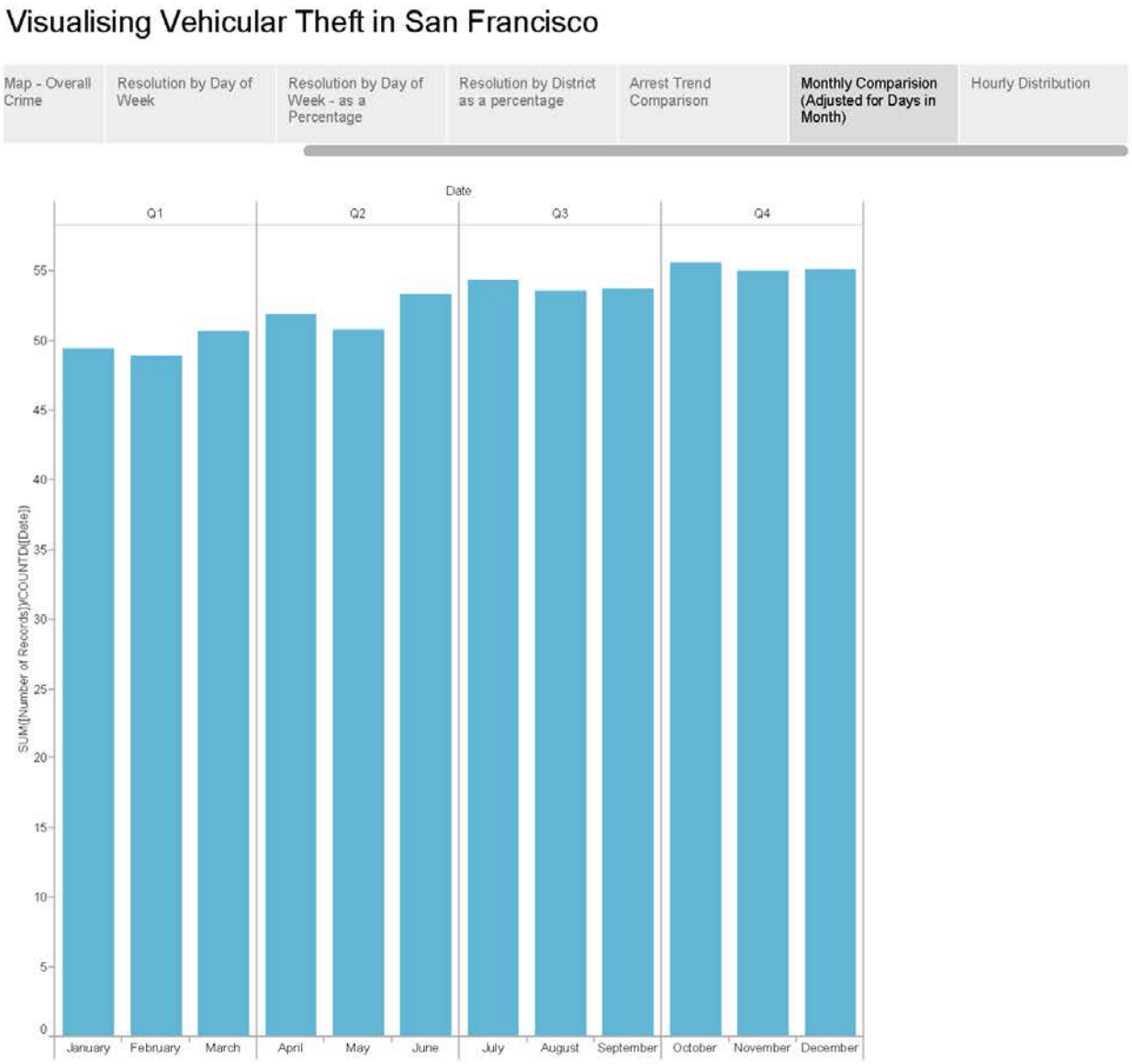
On a district level, VT crime resolution is relatively constant, with only the Tenderloin arresting in more than 5% of cases. Overall, the Tenderloin is again the most likely to arrest - with just under 60% of cases resulting in arrest. On one hand this is surprising (as the Tenderloin is notorious for being a "bad" neighborhood), but it is a very small district, which is heavily policed and has residents that are "used to" a certain level of petty crime (e.g. drug use and prostitution).



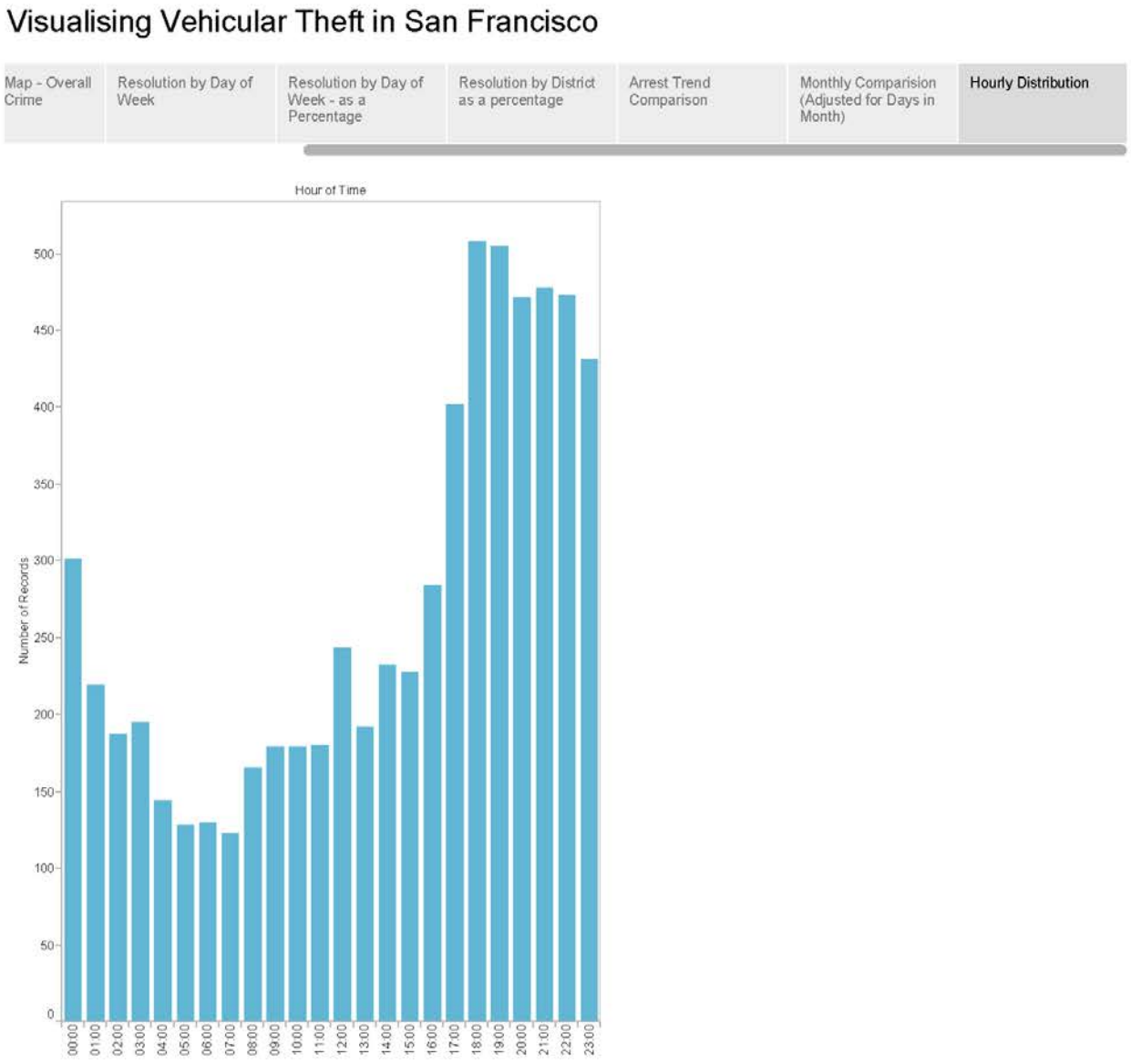
Looking at the time trends, it appears that VT crime arrest rate is increasing, while the overall crime arrest rate is decreasing.



I wondered if there were any significant trends across the seasons and months. For this visualisation, I divided the sum of the crimes by the number of days (such that February wouldn't show a drop in crime just because it has fewer days). It appears there are fewer VT crimes reported in the winter, and more in the autumn, but it's not a huge variation.



On an hourly basis, there is a clear trend; VT crime appears most likely between 18:00 - 01:00, with little activity in the small hours of the night. I'd hypothesise that this decrease could be a result of victims being asleep at the time the crime was committed and not reporting it until they've discovered the crime, though there is no morning “spike” in reports so this could use further
investigation.



If I were to have further time for this, I would like to overlay demographic information onto this - for example, population density, median wages, etc, but that is beyond the scope of this assignment.