Is Theft in Louisville on the Rise?

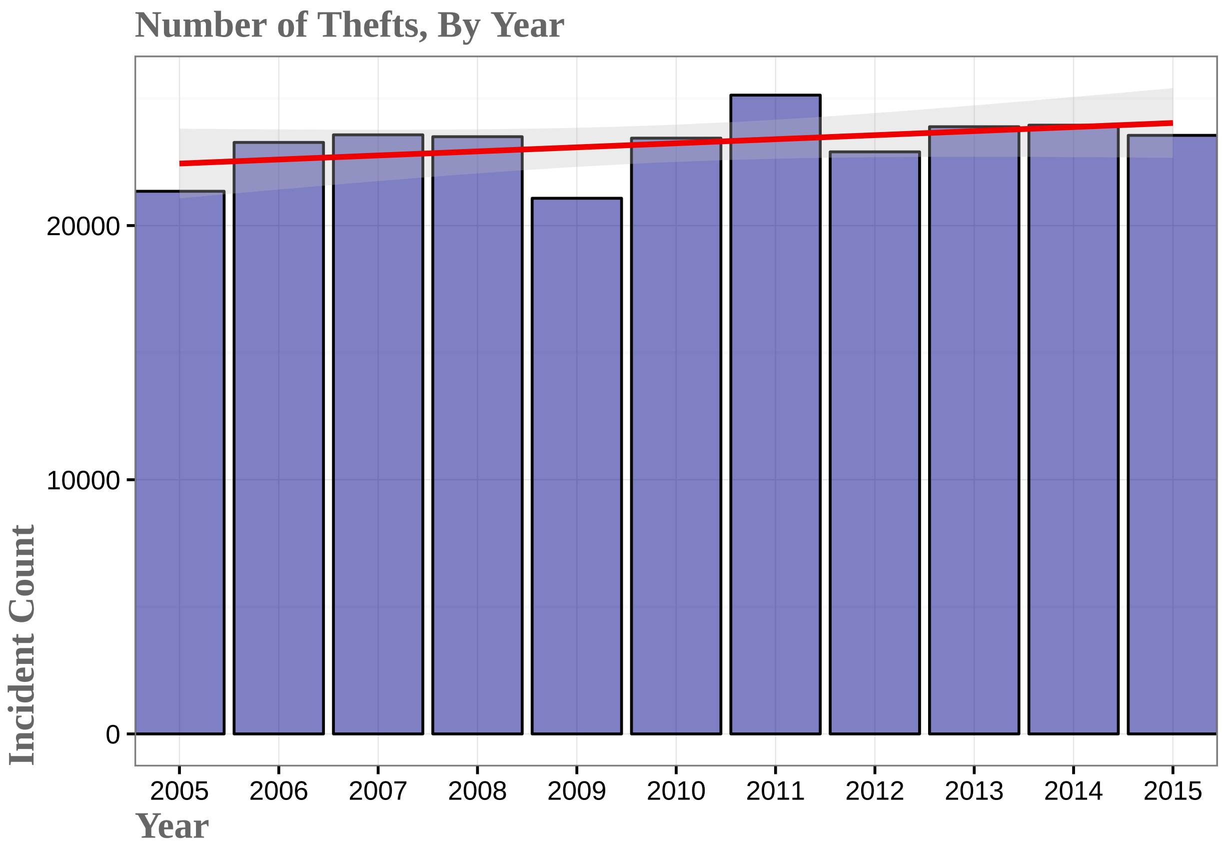
Introduction

Last time, we dove into Louisville's open crime data set to explore the city's violent crime. Today, we examine the vast world of theft/larceny crimes in Louisville, from shoplifting to motor vehicle theft. In this post, I will explore in more detail the idea of crime dispersion. I would like to quantify more rigorously the idea of whether crime, specifically theft in this post, is becoming worse everywhere or whether we are only seeing a worsening in localized areas, with a stablization (or even a decrease) throughout the rest of the region. To do this, I will borrow ideas from economics and geospatial crime analysis. Without further ado, let's get started.

Overview of Theft Offenses

While violent crime had a very narrow scope consisting of 4 major offenses, theft offenses cover a much wider range. Depending on how we want to define it, theft could cover crimes as diverse as fraud, motor vehicle theft, bad checks, shoplifting and indentity theft. The National Incidence Based Reporting System (NIBRS) classifies 8 crimes(pocket-picking, purse-snatching, shoplifting, theft from building, theft from coin-operated machine or device, theft from motor vehicle, theft of motor vehicle parts or accessories, and all other larceny) as theft/larceny offenses. The separate crimes of motor vehicle theft and stolen property offenses seem to get labeled as theft by the city of Louisville, so we will include those offenses in our analysis as well. Fraud, while obviously related to theft, is given separate treatment in the NIBRS reports and I will defer to them in this instance.

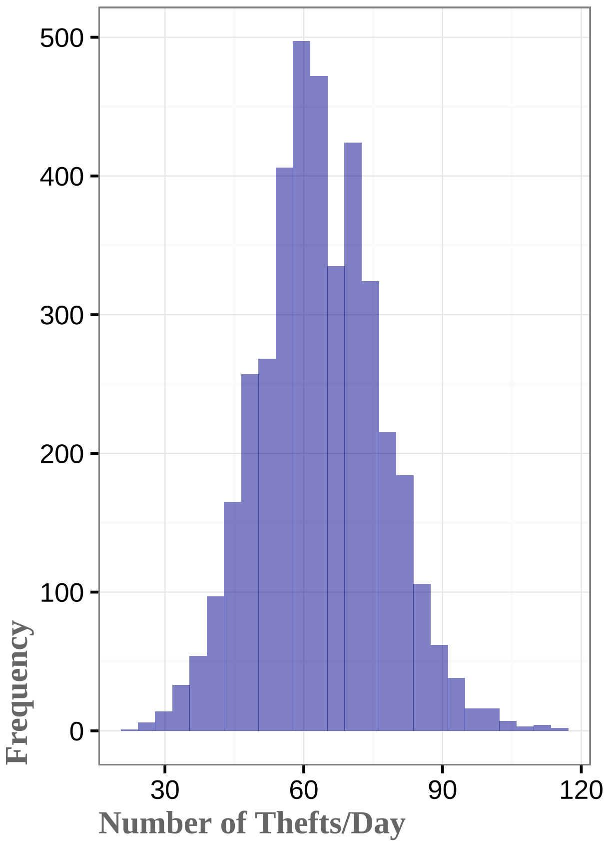
Let's start in the same manner as last time by taking a broad view of theft over the years.

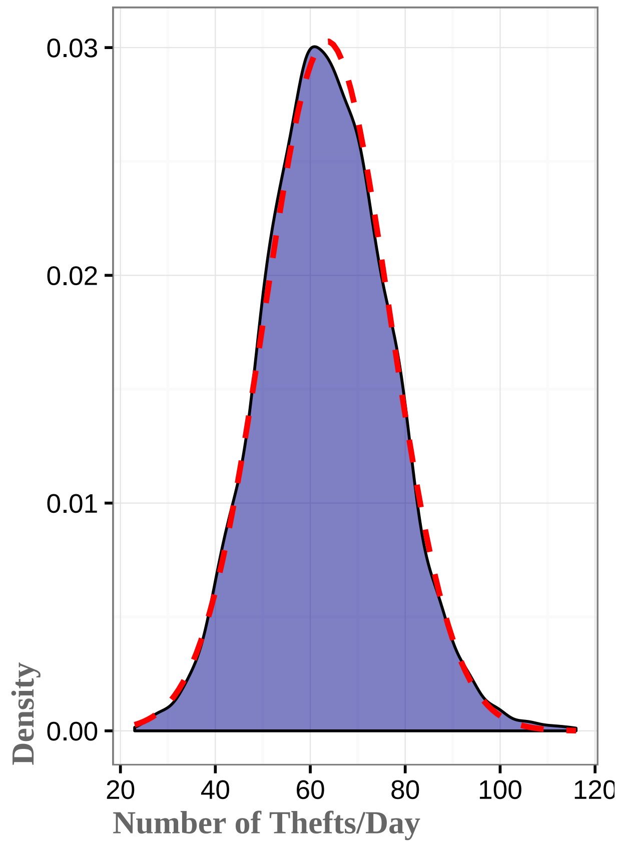


As with violent crime, at first glance there appears to be a minor upward trend, although without the red trend line it is very subtle. However, unlike violent crime we actually had a very slight decrease in theft during 2015. Though, as you can see in the table below, we are still up about 1.3% from the mean for the last 10 years.

crime_type	Change_from_2014	Change_from_Max	Change_from_Mean
theft_offenses	-1.7%	-6.3%	1.3%

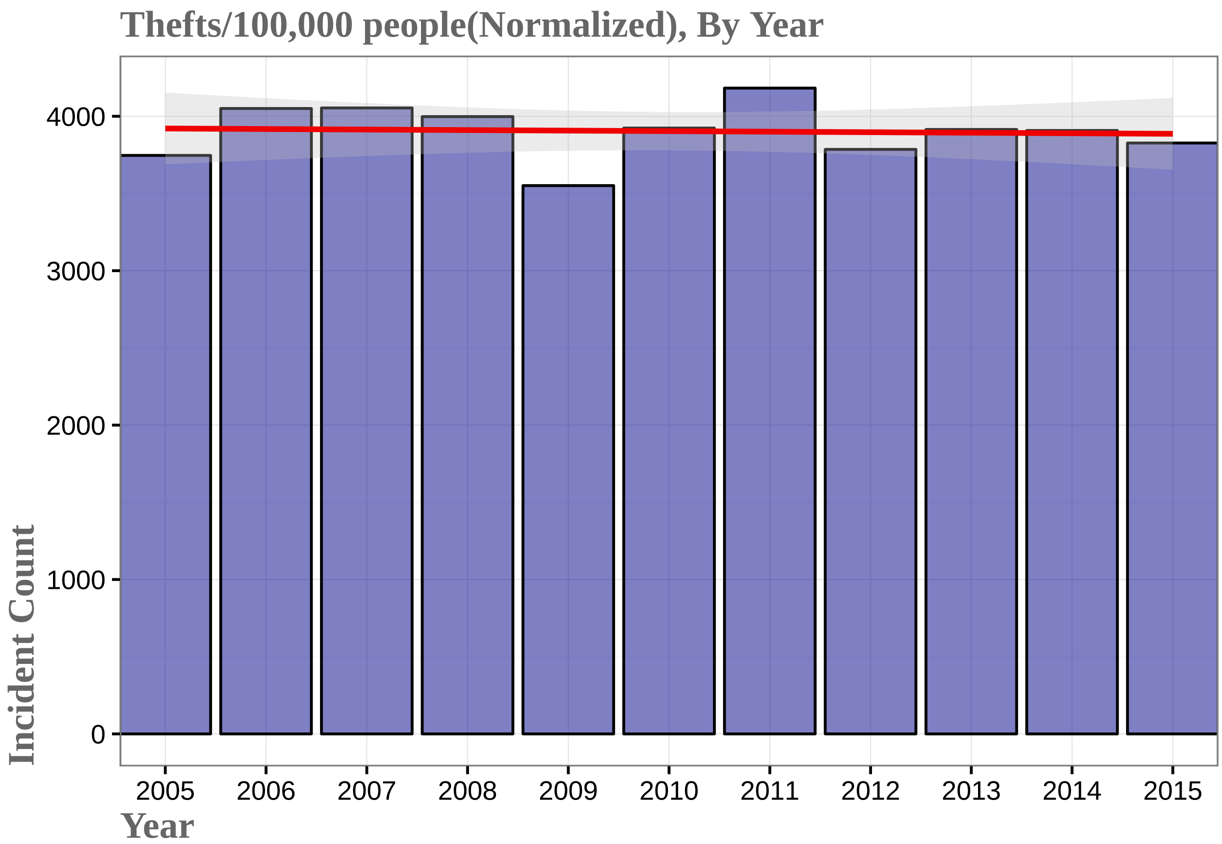
As usual, it is very informative to view the distribution of thefts, in this case by day, to check for anything unusual.





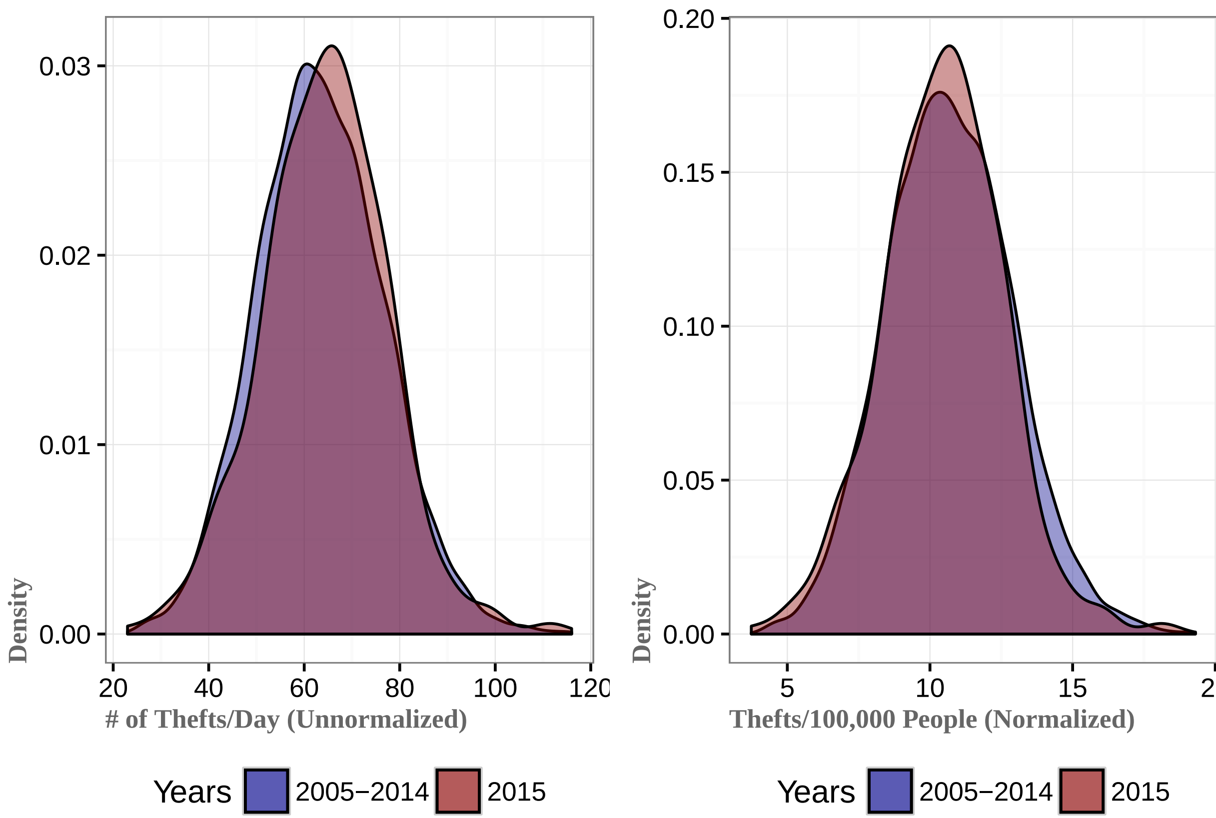
From the plots above, we can see that the data is very close to being normally distributed. There is still some right-skew to the data, indicating that there are more high theft count days than we would expect if the data was perfectly normal, but it is close enough that we won't have to worry too much about model assumptions for later posts.

The distributions also give us our first sense of how prevalant thefts are with an average of 63.54 thefts per day. Looking at two specific years, say 2010(the last available census year) and 2015, we see an increase from 64.18 to 64.42 thefts per day. However, if we normalize for the population increase using US census data, and view it in more relateable units, Louisville has actually seen a decrease in the number of thefts per 100,000 people per day–10.75 in 2010 to 10.47 in 2015. This ignores many of the complexities inherent to normalizing data for population(e.g., at what level should you normalize? City wide? Zip Code? City Block?) but does provide a glimpse into just how nebulous this sort of analysis can be. In fact, when we normalize the yearly theft counts we see a plot with a minor trend down, opposite from what we saw without normalization.



Has there been a significant change?

As you may have guessed, our decision on whether to normalize the data for population changes will have a major influence on the question of significance. First, compare the normalized versus the unnormalized distributions for 2015 vs 2005-2014.



Notice the minor shift right of the 2005-2014 distribution relative to the 2015 distribution. This indicates that after population adjustment, we may have seen a decrease in thefts. When t-tests are performed on both versions, you get results corresponding to what the distribution plots show. The first t-test tests for a significant change in the mean of the unnormalized data. As you can see from the results, with a p-value of .15, we cannot reject the null hypothesis that the means are equal. When you look at the confidence interval, you can see that the 2015 mean of 64.42 is contained within the interval, indicating that they very well could be the same. Simply put, with unnormalized data there is no significant change in the mean number of thefts per day.

The second t-test, dealing with normalized data, does find a significant result, albeit a very borderline one. Here, there is evidence that the 2015 mean is not equal to the 2005-2014 mean and we reject the null hypothesis that they are equal. In fact, given that our confidence interval lies completely below the 2005-2014 mean of 10.69, there is evidence that the theft level has declined slightly.

```
##
##  One Sample t-test
##
## data:  sums_2015$count
## t = 1.4098, df = 363, p-value = 0.1594
## alternative hypothesis: true mean is not equal to 63.45387
## 95 percent confidence interval:
##  63.07118 65.77497
## sample estimates:
## mean of x
##  64.42308
```

```
##
```

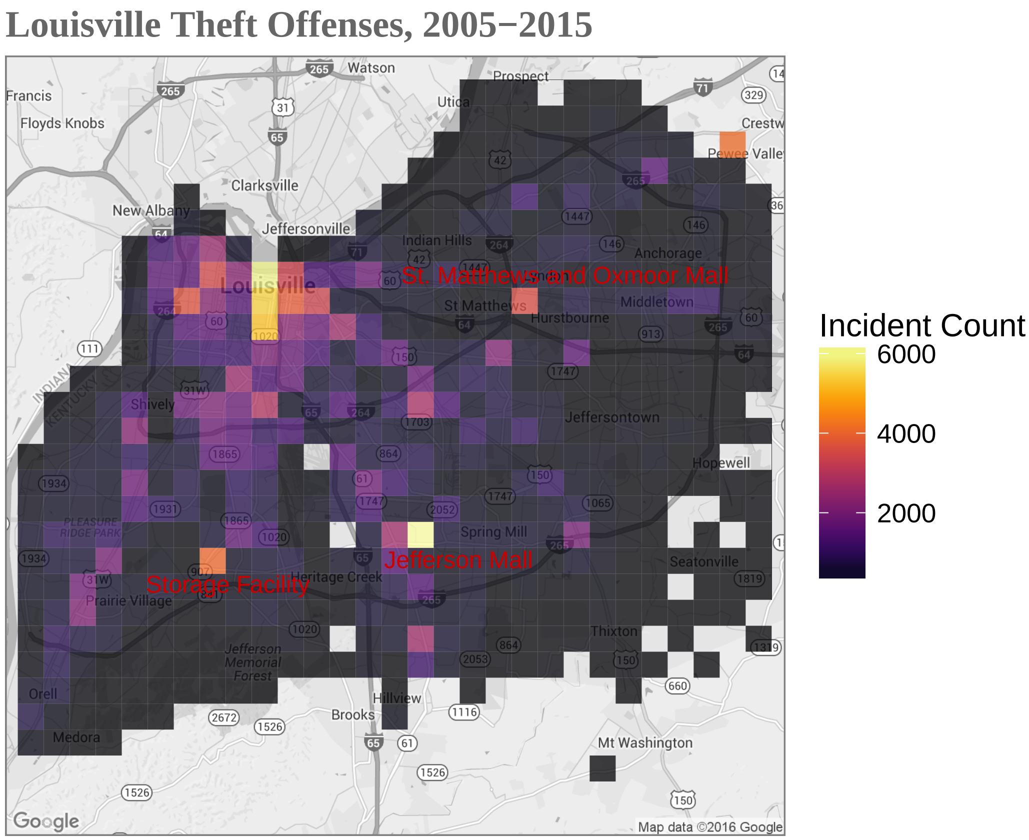
```
##  One Sample t-test
##
## data:  sums_2015$pop_adjusted
## t = -2.025, df = 363, p-value = 0.0436
## alternative hypothesis: true mean is not equal to 10.69529
## 95 percent confidence interval:
##  10.24938 10.68876
## sample estimates:
## mean of x
##  10.46907
```

This set of tests is only a first, rough, step, but it does give us an idea of what we are dealing with. In the worst case, we haven't seen any change at all. And depending on the accuracy of our population estimates, we might have seen a slight decline in theft levels.

Mapping Theft Locations

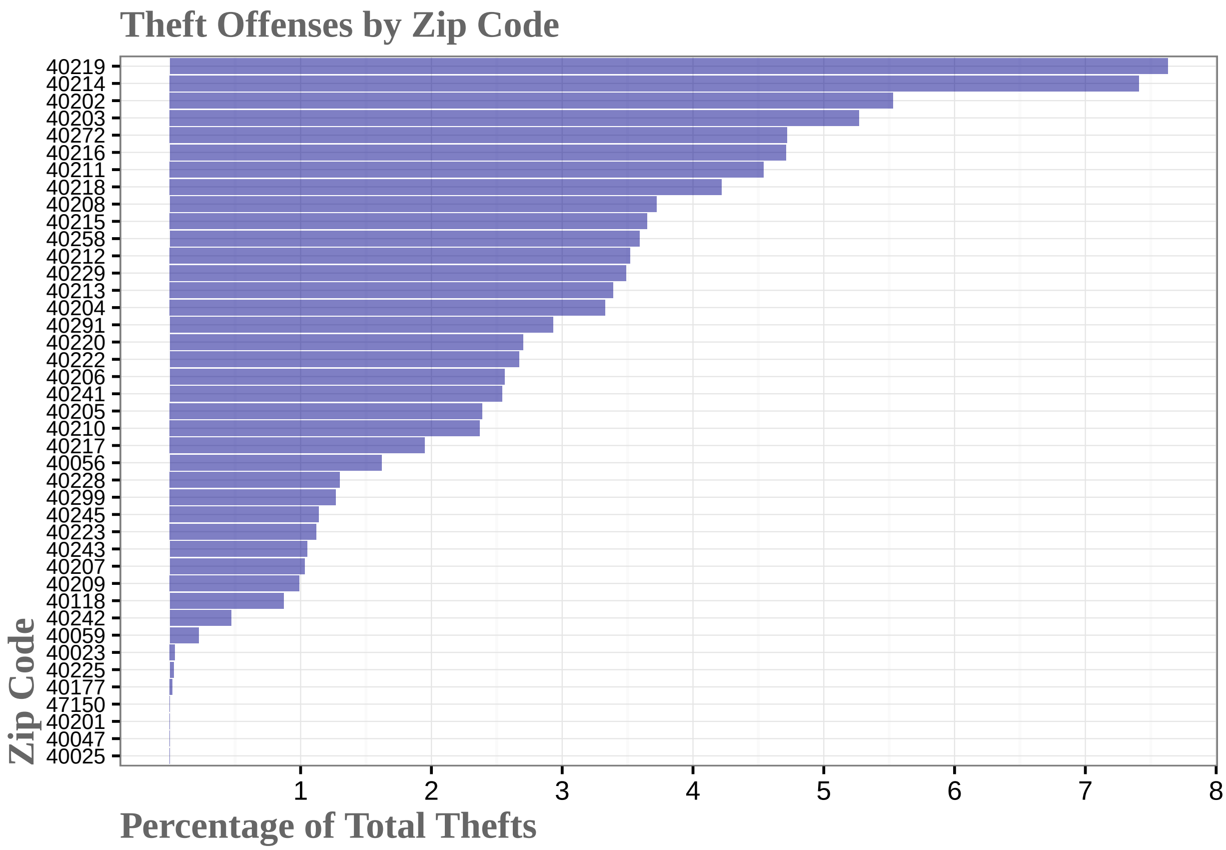
Mapping theft locations provides a good understanding of the distribution of theft throughout the city. We can immediately see that, as with violent crime, the major hot spot is centered in downtown. Presumably this has something to do with the population density being higher downtown, but I will not delve into that issue today.

Instead, I want to briefly look at the dispersion throughout the city. If you look carefully at the map, you will notice that despite the obvious hotspot downtown, we still have color feathering into the far corners of the city. There are also several isolated 'blocks' of high activity (indicated by the yellows and orange/salmon colors) that are located well outside what I would consider downtown Louisville. Two of those–labeled in red–are locations of shopping malls which, understandably, having high levels of shoplifting/theft. The other two areas are less obvious. After some investigation, the high count block west of Heritage Creek seems to be a high theft storage facility. The block up by Pewee Valley seems to be a result of lazy record keeping. Instead of specific block addresses for crimes in that area(zip code 40056), most of the entries were just given a 'community at large' address. When the geocoding was done, this resulted in thousands of identical coordinates which then spawned a high incidence count block.



Analysis of Theft Dispersal

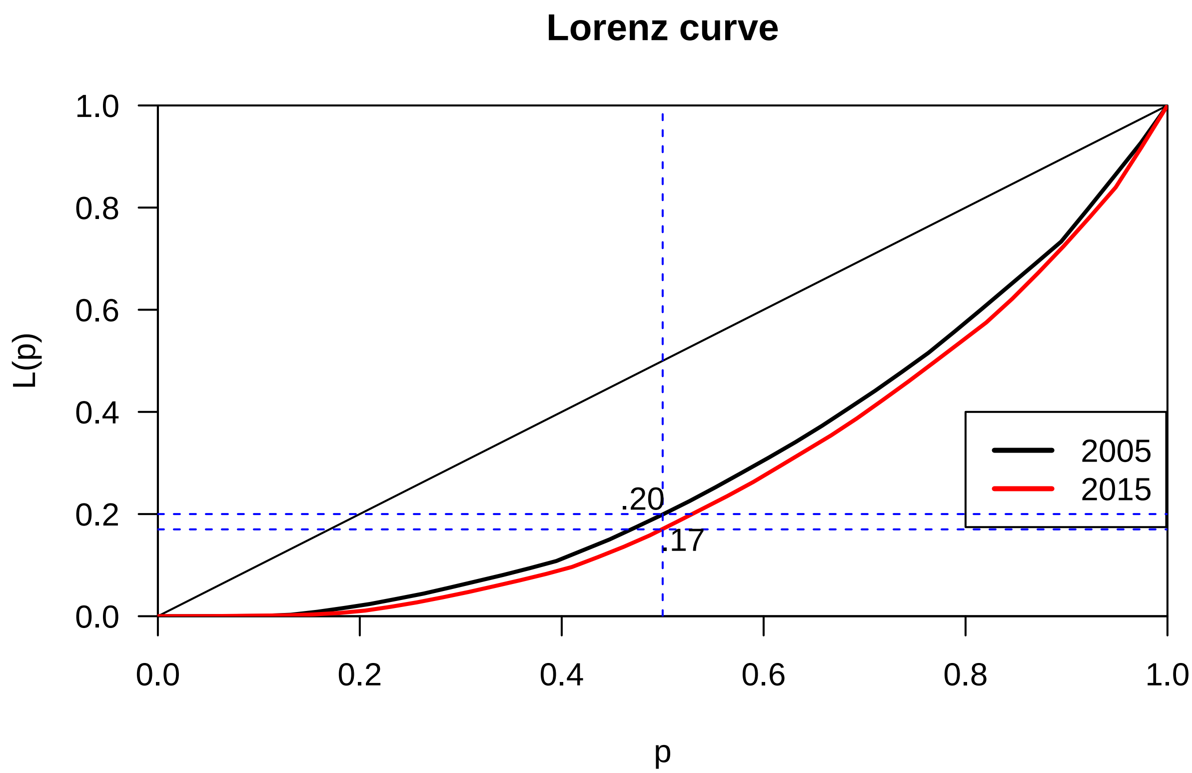
Since our map seems to indicate a wider dispersal of thefts, looking at theft counts by zip code should be informative.



Surprisingly, the highest crime zip codes are not in downtown Louisville at all, but are more in the south and west end of Louisville. There are still several downtown zip codes– particularly 40202 and 40203 which are located at the bright yellow area of our heat map– near the top of the plot, but we can see evidence that thefts are more evenly dispersed throughout the community. While just 6 of Louisville's 38 zip codes accounted for 48.75% of all violent crime, it takes 9 to cover 47.75% of thefts.

But this sort of discussion is not particularly rigorous and, frankly, is a little unsatisfying. Luckily, there are entire fields devoted to the study of dispersion. In economics, it is common to measure income related inequality using the Gini coefficient and Lorenz curve(a good introduction can be found here and here). These techniques have also been applied to identify unequal distributions in crime frequencies. For our purposes, we will measure the inequality in the frequency distribution of thefts by zip code. A Gini coefficient of 1 indicates maximum inequality(in our case this would be all of the thefts taking place in one zip code) while a coefficient of 0 indicates complete equality (thefts are evenly dispersed throughout all zip codes).

Calculating Gini coefficients for 2005 and 2015, we get 0.41 and 0.45 respectively. This increase indicates that theft offenses have become more unevenly distributed(i.e., thefts are becoming more highly concentrated in certain zip codes). We can vizualize this with a Lorenz curve. If crimes were completely evenly distributed, they would fall on the diagonal 'line of equality'. The further below the curved line bows, the more unequal the data. As you can see, the 2015 curve bows ever so slightly more, indicating higher inequality. So, for example, we can see that the bottom 50% of zip codes account for 20% or 17% of the thefts, depending on the year. This solidifies the idea we saw above where a fraction of zip codes accounted for a disproportionate percentage of the total thefts.



While this measure is a good start, it is a global measure of unequal distribution when what we would like is something more localized. A more crime specific measure that is a step in this direction is called the Offense Dispersion Index(ODI). When looking at the crime increase between two years, we first calculate the difference in each area–in our case zip code areas. Then, we order these differences from highest to lowest change. Finally, we remove the highest ranking area and recalculate the crime rate with the remaining areas. Then we take the second highest ranked area and remove it, again recalculating, but for the n-2 areas. This continues until only one area is left.

From this procedure we can calculate the ODI, which is just the proportion of areas that must be removed from the calculation before the increase in crime turns into no-change or a decrease in crime. So as you can see in the table below, it takes the removal of 5 zip codes worth of data to change the increase in crime rate from 2005 to 2015 into a decrease. So the ODI is just 5 divided by the total number of zip codes, 38, or 0.13. ODI's range from 0 to 1 with values close to zero indicating a low crime increase dispersion factor. In other words, a value closer to 1 suggests suggests a problem across many areas, rather than something very localized. To compare, I calculated the ODI for drug crimes in Louisville to be 0.56. This suggests that thefts crimes have not dispersed much and remain essentially localized while drug crimes are now a problem in many areas of Louisville.

zip_code	citywide_thefts_2005	citywide_thefts_2015	citywide_change_after_zip_removal
all	21348	23537	10.25
40214	19967	21615	8.25
40216	19169	20286	5.83
40211	18266	18915	3.55
40291	17815	18068	1.42
40272	16866	16817	-0.29

zip_code	citywide_drug_crimes_2005	citywide_drug_crimes_2015	citywide_change_after_zip_removal
all	10393	13900	33.74
40203	9230	12231	32.51
40212	8512	11100	30.40
40211	7702	9882	28.30
40272	7476	9258	23.84
40214	7025	8489	20.84

One additional measure worth mentioning is the non-contributory dispersion index(NCDI). The NCDI is the proportion of areas that showed crime increases divided by the total number of areas. Unlike the ODI ratio–which only uses areas that contributed to the overall crime increase–the NCDI looks at all areas that showed increases and can thus be used to show the spread of areas showing increases. For our theft and drug data, we see 0.55 and 0.85 NCDI ratios, respectively. The lower NCDI of thefts, when combined with the low ODI, suggests that theft crimes are localized to a few problem spots. On the other hand, the higher NCDI of drug offenses suggests drug crime may be an emerging issue for the entire region.

It must be noted that both ODI and NCDI measures are first, rudimentary, steps into quantifying crime dispersion. Much more advanced techniques which take full advantage of spatial analysis can provide significantly more detail on localized variation in crime patterns. These are far beyond the scope of this post, but with a little more work the ODI measures we calculated here can be combined with other techniques for more specific analysis. If you are interested, a good starting point is located here. This whitepaper was the inspiration behind much of this analysis and a full aknowledgement of the use of its ideas is essential.

Conclusion

This first look into theft in Louisville provides a little more hope for ordinary citizens that crime is not shooting through the roof. While many factors were ignored or, admittedly, overlooked, it appears that theft offenses are stable or maybe slightly down for the region as a whole.

The more interesting, and simultaneously worrying, issue is the increasingly localized nature of thefts. A tiny fraction of the city bears the burden of the vast majority of thefts and it only appears to be getting worse. On the one hand, this could be an effect of enforcement policies. It seems entirely possible that, especially given the tightening metropolitan budgets, instead of policing within hotspot regions it is easier to police the zone around them in an attempt to quarantine crime to these small regions. This could potentially be more cost effective for the city.

On the other hand, it could be that patrols do police these areas frequently, but there are tangible demographic, infrastructure and economic factors overrepresented in these regions that drive up crime. For instance, an interesting next step could be to look at the spatial correlation between theft crime and drug crime to see if, perhaps, drug addiction is driving up thefts. If this is the case, then, unfortunately, the fix is more difficult than something simple like increasing police presence in the area.

Next time we delve into how Louisville's notoriously fickle weather effects the crime rate. This will lead directly into the building of several machine learning models designed to predict the count of various crimes throughout the city.

Till next time!