

More Crimes on Weekends?

An Investigation Using Raleigh Open Data

Anylytics

January 17, 2015

1 Introduction

The Open Raleigh websites gives us all of the police reports from 2005 until now. The data can be found through the following link <https://data.raleighnc.gov/category/public-safety>. The data provided in the open data portal gives us the following information for each police report

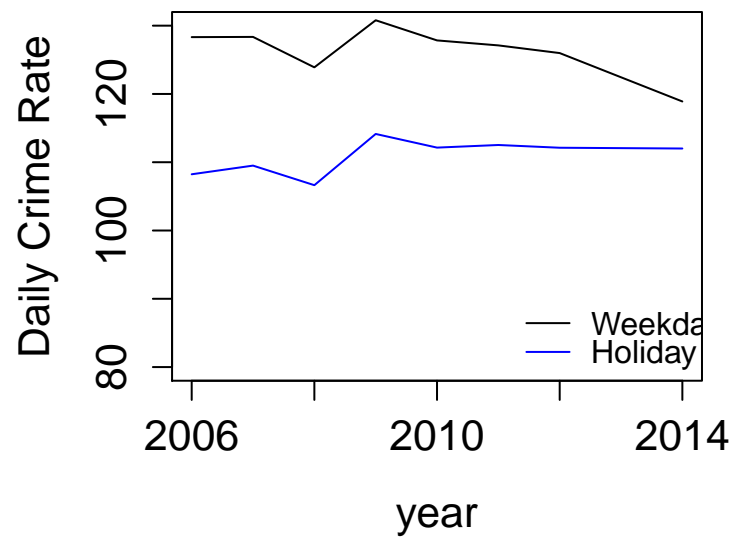
- Date and Time
- Type of Crime
- Location of the crime
- The Police Beat covering the area of crime

Even with just this data, we can do some interesting analysis. In this article, we will try to compare crime in Raleigh on workdays and on holidays. We will use a standard linear regression framework for the trend analysis. It is fairly basic but still insightful. As we collect more data from other sources, we will be able to enhance the analysis and make it more meaningful and robust.

Caveat: Data is available from 2006 to 2014. We have, for the time being, omitted the data from 2013 because we saw an odd spike in the 2013 data that we could not justify. This is discussed in detail in Appendix A.

2 All Crimes

Using the data, we can also see if there is more or less crime on the weekends and holidays. On the plot, we compare the crime rate between workdays and holidays. We see that there is a fairly constant difference between the amount of crimes committed on workdays, specifically that there is more. In 2014, though, there seems to be a smaller difference.



We also ran a linear regression, and it gives us significant results. It tells us that on average, there are 16.8 fewer crimes on holidays rather than days that are not. It also shows that there 14.28 fewer crimes on weekends than weekdays. If we pool together holidays and weekends, we see that there are 15.5 fewer crimes. All in all, the police report data seems to suggest that there are more crimes happening on workdays than not.

3 Separating by Type of Crime

3.1 Methodology

Now that we know that there is a difference between workdays and not, we see if the same result holds for specific types of crimes. Knowing such information can be of use to the police districts; if there are clearly different trends for the different types of crimes, then officers should have different assignments on workdays vs non-workdays. We use the following linear regression framework to find out the

$$DailyCrimeRate_{jt} = \alpha + \beta_j TypeofDay_t * \sum_j^J Crime_{jt} + \gamma_j \sum_j^J Crime_{jt} + \epsilon_{jt} \quad (1)$$

where t represents a day, $TypeofDay_t$ represents whether day t is a weekday or holiday, $Crime_{jt}$ is the amount of crimes of type j on day t , and ϵ is the unobserved error term. There are J different types of crimes and j represents one of many types of crimes. We "interact" the the Type of Day with the type of crime in our regression; this is the multiplicative term in Equation 1. The $TypeofDay_t$ is a binary: 1 if it is non-workday and 0 if it is not a workday. When we run a regression with this specification, we get estimates and standard errors for coefficients above. The estimates are denoted $\hat{\beta}_j$ and $\hat{\gamma}_j$. For a given j , i.e. for a given crime category, $\hat{\beta}_j + \hat{\gamma}_j$ is the coefficient for crime for non-workdays and $\hat{\gamma}_j$ is the coefficient for crime for workdays. So to compare the two, we take the difference; thus $\hat{\beta}_j$ tells us the difference between non-workday and workday daily crime rate for a particular crime j . We run the regression above with three difference specification, one for weekend, holiday or non-workday. A non-workday is any day that is a weekend or a holiday.

3.2 Results

The following is the table of estimated coefficients to Equation 1. The coefficients represents how many more or less crimes of that type happen on that day versus not; these are the $\hat{\beta}_j$ from the methodology section. For example, the weekend coefficient for Driving is 2.768. This means that there were 2.768 more reported driving crimes on weekend days than weekdays. The amount of stars represent the significance level of the

estimate. One star is 68 percent confidence level, two stars represent 95 percent confidence and three stars represent 99 percent confidence. For example, two stars implies that zero is within two standard errors of the estimate, i.e.” I can say with 95 percent confidence that the estimate is non-zero”.

The final column is the most important one to look at; it represents any day that is not a workday. The first two columns combine to make the final column. It is clear that there is a lot of variation in sign of the coefficient. Even though it is true that there are fewer crimes reported on the off-days, that cannot be said about all types of crimes. Driving, Assault, Robbery and Weapons have significantly higher rates on workdays than not. Most of the types of crimes also have significant results. There were others in the data that are not shown in the table below; we chose not to because there were a low amount of occurrences and thus will not provide any significant results.

	Type	Weekend	Only Holiday	Non-Workdays
1	General	-14.28***	-16.24***	-15.533***
2	Driving	2.768***	-0.607	2.589***
3	Larceny	-6.0451***	-5.9273***	-6.4250***
4	Assault	4.1446***	1.79288**	4.1528***
5	Drugs	-1.72634***	-3.01060***	-1.97802***
6	Burglary	-2.24984***	-0.20183	-2.16830***
7	Property Damage	0.84086***	0.31633	0.83693***
8	Robbery	0.33910***	-0.25847	0.29562 ***
9	Weapons	0.43901***	0.26368	0.44860***
10	Sex Offense	-0.35464***	-0.45569**	-0.38331***
11	Humane	-0.24768***	-0.4211*	-0.28206***
12	Disorderly Conduct	-0.00013	-0.3536*	-0.03076
13	Homocide	0.04218	-0.03311	0.03808

Let's visualize the information given in the previous table. Below we have a set of figures that compare the Daily Crime Rate for non-workdays vs workdays split by different types of crimes. The blue line represents days that are either holidays or weekends. The black one represents a workday.

The graphs bring insight into a couple important aspects of reported crime in Raleigh. Firstly, it shows us that different types of crime behave differently on workdays and non-workdays. For example, there are many more reported larcenies and burglaries on workdays, whereas the opposite is true for assault, driving and weapons-related crimes. For the crime categories that are not shown in the graphs, there does not seem to be a consistent difference; for certain years the workday/non-workday difference is positive and negative for others. For these crimes, either there might not be a truly consistent difference or there are not enough data points to come to a significant conclusion. An example of the latter is homicide; since there have only been 193 reported homicides in Raleigh in the past decade, it is more difficult to come to a statistically significant result.

The workday/non-workday comparison for these crimes is further reassuring since they do not change from year to year. In all of the plots in the figure below, the workday/non-workday difference is consistently positive or negative for a certain crime category.

The graphs also give us time trends for types of crimes. Though we see in the figure above that crime overall has for the most part stayed fairly constant, this is certainly not true when we split by crime categories. For example, reported driving incidents have constantly increased over time whereas the opposite holds for property damage. In fact, reported property damage have drastically decreased from approximately eleven a day to four a day. This could be due to legitimate decrease in crime or it could represent police diligence changing in that specific field.

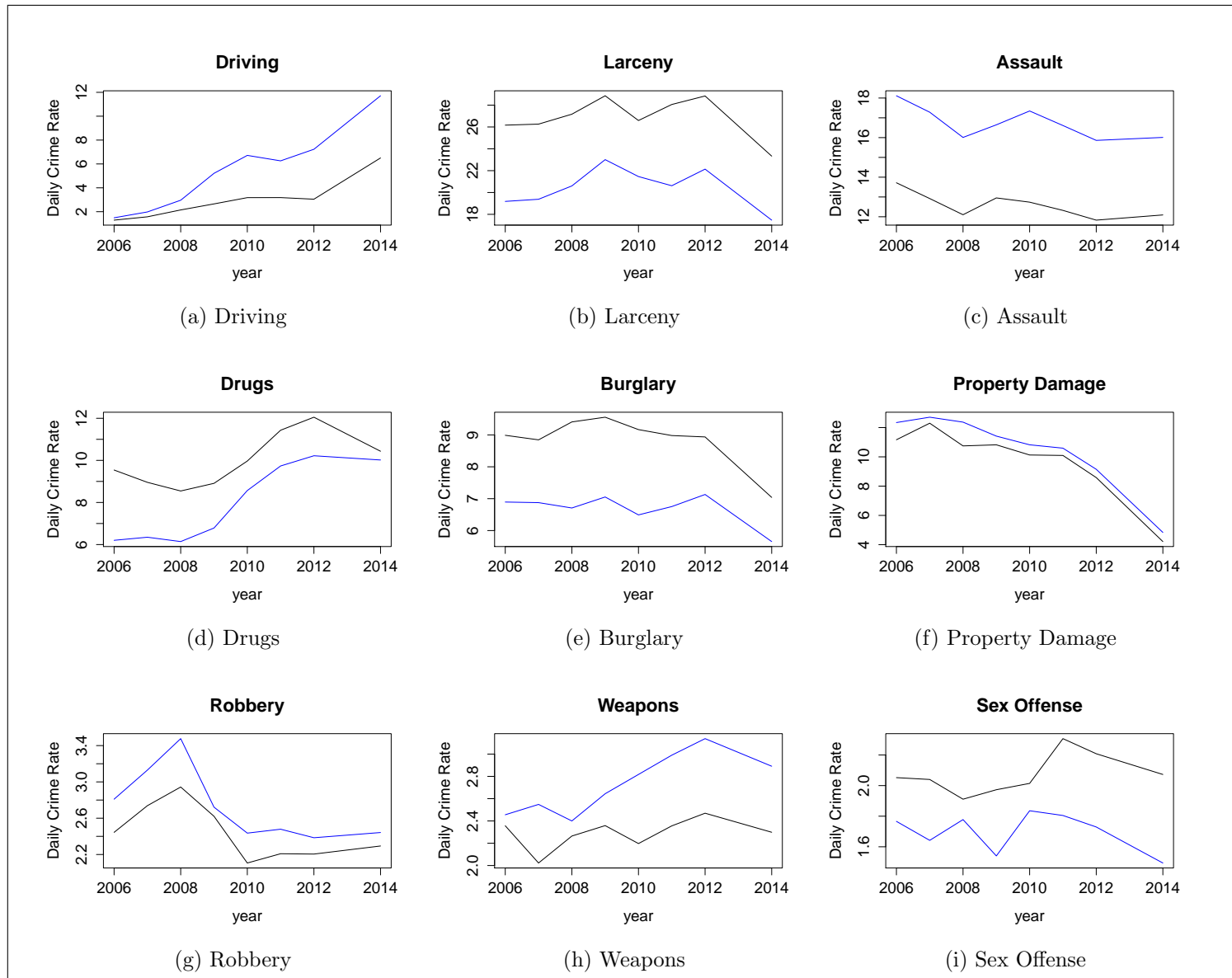


Figure 1: In each of the plots, the black line represents a workday and the blue line represent a holiday or weekend

4 Conclusion and Future Work

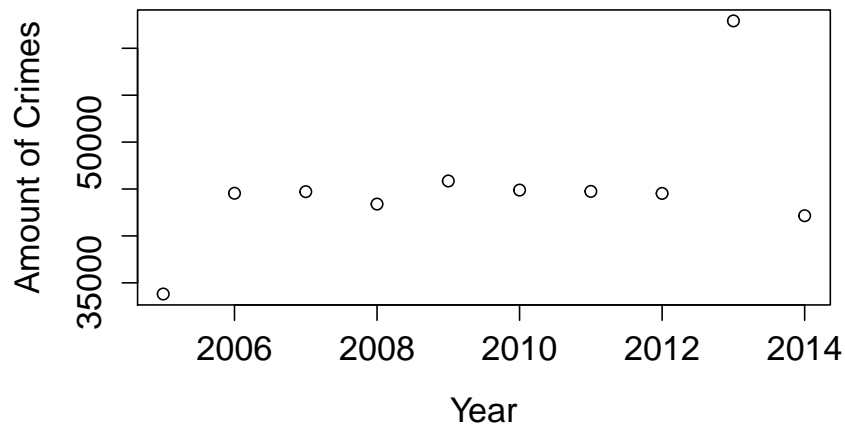
In this paper, we have shown that there are clear difference in crimes when comparing workdays and not. Given that police officers are assigned shifts and specific beats, if these trends can be verified with more rigorous analytics using more data, police scheduling can be strategically shifted to address these trends. For instance, since we know that larceny is much more prevalent during workdays, police officers should increase diligence accordingly. Further analysis could be done to find out which locations have higher instances of which types of crimes. With that knowledge, we can assign police officers specific location on specific days to address the current crime climate.

Within this day to day analysis, we can also start to incorporate other data sources like weather data. If in regression represented by Equation 1, we added a weather variable e.g. temperature, precipitation etc., we can bolster our predictive analysis. Other possible data sources that we can utilize includes subdivision data, traffic data, police officer assignment, etc. Essentially, this paper gives a strong start for our analysis of Raleigh crime; our future work will build on this and provide usable and trustworthy predictive analysis.

Appendices

A Yearly Analysis

There seems to be a discrepancy in 2013 compared to 2006 - 2012 and 2014. Approximately 65000 in 2013 but an average of around 45000 in the other years (except for 2005)



It's clear that there is an unexplained jump in the amount of crimes reported in 2013. We suspect that record taking might be different in 2013 as compared to other years, thus resulting in measurement. Due to this issue, we will drop 2013 from the data set. It is not much of an issue that we do since we have a lot of data from the other years.