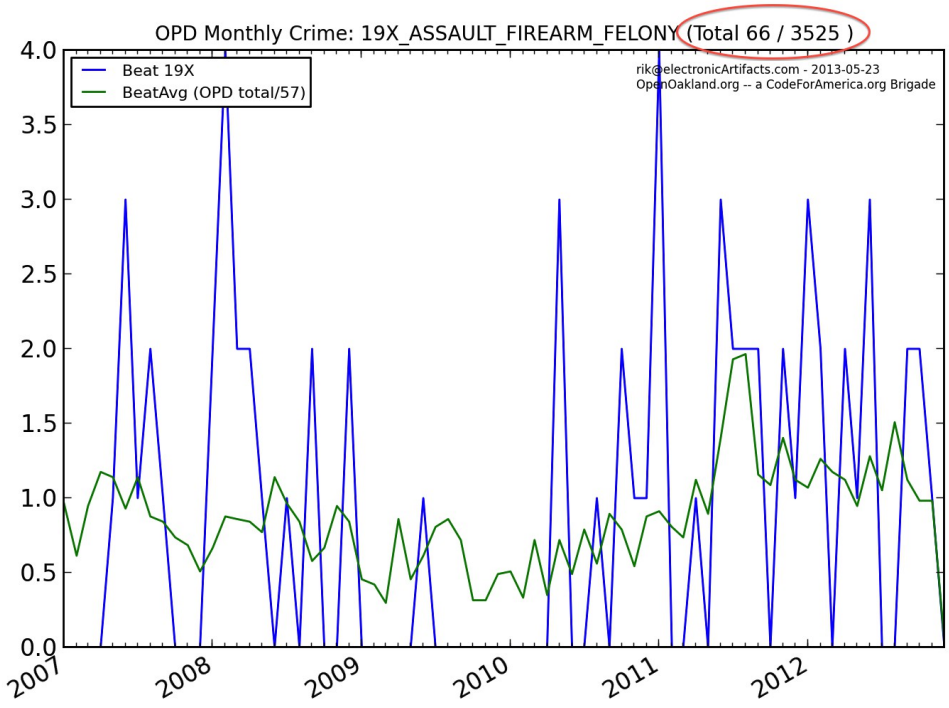
It's ShowCrime! Oakland's criminal past

It's my pleasure to announce the availability of ShowCrime, a visualization tool for Oakland's criminal past.

As loyal readers will recall, I have been working on historical data provided by the Oakland Police Department (OPD) as part of OpenOakland, a CodeForAmerica brigade. Earlier stages have cleaned up this data, and merged it with will contain data analysis done by Urban Strategies Council (USC) In 2012; details of this merge described below.

ShowCrime is designed to provide a baseline of retrospective statistics and analysis, allowing comparisons with new proposals for open data standards now being developed.

Let's begin with the main plot you can get:



This shows shows two curves over the years 2007 through 2012, selecting for two particular features: a beat, and a particular crime type. Beat is a geographic region defined by OPD, and an attribute associated with each crime occurrence. The CrimeCatOAK is with respect to a crime type hierarchy classification system. This hierarchy is also the result of a merge from many sources. The details of

the CrimeCatOAK classification process is also described below.

The first curve shows the number of crimes, aggregated by month, for the selected beat. For comparison, a city-wide "average beat" curve is also shown; this is just the city total, divided by 57 (the number of beats). Early user testing suggests many Oakland residents are interested in how their own beat's statistics compare to city averages, and scaling by the number of beats provides a reasonable baseline.

A couple of fine points:

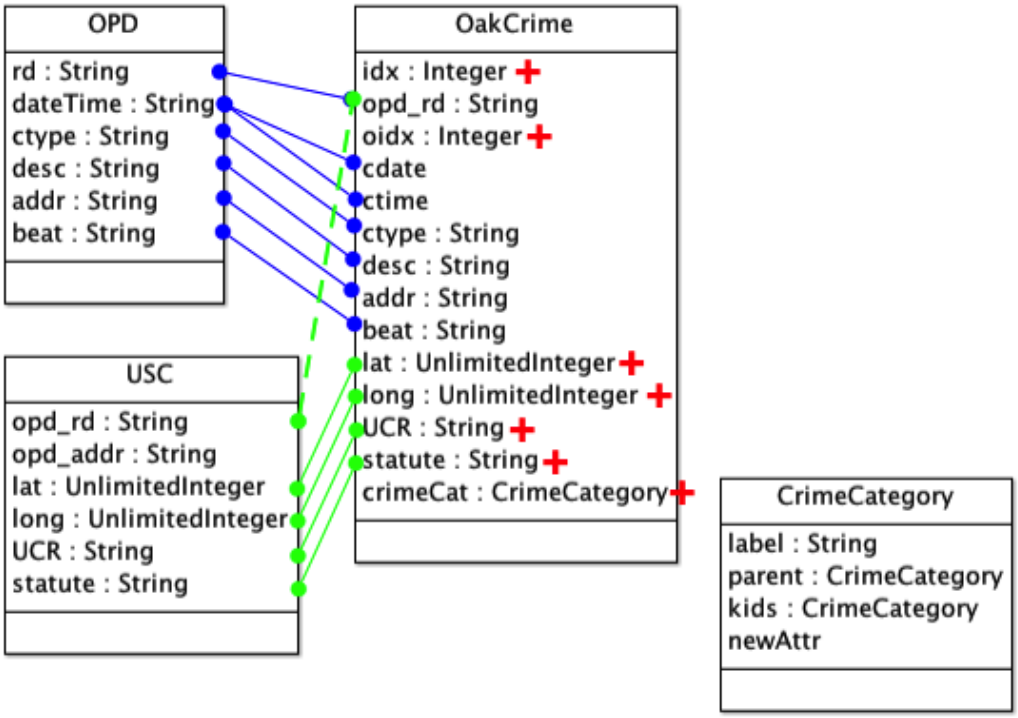
- The plot's title mentions both the beat and crime type, followed by two additional numbers in parentheses (circled in red). These are the total number of the crime type over the 2007-2012 period, within the beat and city-wide.
- Also note that some plots have a fractional number of crimes labeling the vertical (Y) axis. That's because there are averages being taken on a monthly basis, and infrequent (but often very serious) crime types may well happen at low enough rates that fractional values are appropriate.

Data model

A key assumption made in this project is that city's official OPD police distribution is and must remain the primary provider of record, of base crime data for the City of Oakland. The range of third-party vendors building from this source continues to grow, and we are optimistic that some of these will also make their data publicly available.

For example, USC has done great work in the past analyzing OPD data. This project has made use of that work, and some results would not be possible without USC input. In particular USC provided geo-coding of crime locations, both before the termination of their OPD contract, as part of a follow-up analysis in April, 2013, and by virtue of "extrapolation" techniques (when the same street address was mentioned more than once).

The figure below shows the primary OakCrime data type defined by this project, and its relation to data provided by OPD and USC.



The blue lines capture the dominant role played by OPD's data. USC's data provided latitude/longitude data, and also some UCR and statute labeling for some OPD data. This has been included in the merged data set, but was not used for CrimeCatOAK classification

Crime categorization

Before describing the CrimeCatOAK system, it is important to acknowledge that it is far from perfect. The perfect "Crime categorization system" would be provided by OPD, relating their itemized crime reporting to the aggregate statistics they provide elsewhere on their website, to the FBI, via the California Attorney General. But because no such classification system has been provided by Oakland, CrimeCatOAK has been developed.

CrimeCatOAK refers to the following hierarchy of crime types:

1. ARSON
2. ASSAULT
    1. ASSAULT_FIREARM
        1. ASSAULT_FIREARM_FELONY
        2. ASSAULT_FIREARM_MISDEMEANOR
    2. ASSAULT_HANDS-FISTS-FEET
        1. ASSAULT_HANDS-FISTS-FEET_FELONY

    2. ASSAULT_HANDS-FISTS-FEET_MISDEMEANOR
  3. ASSAULT_KNIFE
    1. ASSAULT_KNIFE_FELONY
    2. ASSAULT_KNIFE_MISDEMEANOR
  4. ASSAULT_OTHER-SIMPLE
    1. ASSAULT_OTHER-SIMPLE_FELONY
    2. ASSAULT_OTHER-SIMPLE_MISDEMEANOR
  5. ASSAULT_OTHER-WEAPON
    1. ASSAULT_OTHER-WEAPON_FELONY
    2. ASSAULT_OTHER-WEAPON_MISDEMEANOR
  6. ASSAULT_PEACE-OFFICER
    1. ASSAULT_PEACE-OFFICER_FELONY
    2. ASSAULT_PEACE-OFFICER_MISDEMEANOR
  7. ASSAULT_THREATS
3. COURT
  1. COURT_CONTEMPT
  2. COURT_PROBATION
  3. COURT_WARRANT_FELONY
  4. COURT_WARRANT_MISDEMEANOR
4. CURFEW-LOITERING
5. DISORDERLY-CONDUCT
6. DOM-VIOL
  1. DOM-VIOL_BATTERY-SPOUSE
  2. DOM-VIOL_COURT-ORDER
  3. DOM-VIOL_DOMESTIC-DISPUTE
  4. DOM-VIOL_PHONE
7. DRUG
  1. DRUG_ALCHOHOL
  2. DRUG_POSSESSION
    1. DRUG_POSSESSION_MARIJUANA
    2. DRUG_POSSESSION_NARCOTICS

        3. DRUG_POSSESSION_OTHER
    3. DRUG_SALE-MFCTR
        1. DRUG_SALE-MFCTR_NARCOTICS
8. HOMICIDE
9. LARCENY
    1. LARCENY_BURGLARY
        1. LARCENY_BURGLARY_AUTO
        2. LARCENY_BURGLARY_COMMERCIAL
        3. LARCENY_BURGLARY_OTHER
        4. LARCENY_BURGLARY_RESIDENTIAL
    2. LARCENY_FORGERY-COUNTERFEIT
    3. LARCENY_FRAUD
    4. LARCENY_THEFT
        1. LARCENY_THEFT_GRAND
        2. LARCENY_THEFT_PETTY
        3. LARCENY_THEFT_VEHICLE
            1. LARCENY_THEFT_VEHICLE_AUTOS
            2. LARCENY_THEFT_VEHICLE_CAR-JACKING
            3. LARCENY_THEFT_VEHICLE_OTHER
            4. LARCENY_THEFT_VEHICLE_TRUCK-BUS
10. MENTAL_ILLNESS
11. OTHER
    1. OTHER_FOUND
    2. OTHER_LOST
    3. OTHER_MISSING-PERSON
12. PROSTITUTION
13. RAPE
14. ROBBERY
    1. ROBBERY_FIREARM
    2. ROBBERY_INHABITED-DWELLING
    3. ROBBERY_KNIFE

    4. ROBBERY_OTHER-WEAPON
    5. ROBBERY_STRONG-ARM
15. TRAFFIC
    1. TRAFFIC_DUI
    2. TRAFFIC_HIT-RUN
    3. TRAFFIC_MISC
    4. TRAFFIC_TOWED-VEHICLE
16. VANDALISM
17. WEAPONS

An assignment of crimes to these categories was built from OPD `ctype` and `desc` attributes. An analysis of the distribution of these two attributes taken independently, as well as interactions between the variables between these two fields, was performed. The result is a process beginning from a data table capturing dominance by a crime type only (n=11), dominance by description only (n=145) or interactions between these two attributes (n=186). The data table (`crimeCatMap_130506.csv`) driving this classification process is available to anyone interested.

94% of all OPD records can unambiguously be assigned to a unique CrimeCatOAKe in this matter. A small number (~ 0.5%) of crime items' `ctype`+`desc` combinations result in two different assignments; the most frequent of these are listed in the following table.

CType	Desc	Freq	Found
HOMICIDE	MENTAL ILLNESS	626	['HOMICIDE', 'MENTAL_ILLNESS']
THREATS	DISTURB THE PEACE	432	['ASSAULT_THREATS', 'DISORDERLY-CONDUCT']
DOMESTIC VIOLENCE	CONTEMPT OF COURT:DISOBEY COURT ORDER/ETC	393	['COURT_CONTEMPT', 'DOM-VIOL_COURT-ORDER']
WEAPONS	VIOLATION OF PROBATION-FELONY	337	['WEAPONS', 'COURT_PROBATION']
DOMESTIC VIOLENCE	VIOLATION OF PROBATION-FELONY	297	['COURT_PROBATION', 'DOM-VIOL_COURT-ORDER']
VANDALISM	DOMESTIC DISPUTE	284	['VANDALISM', 'DOM-VIOL_DOMESTIC-DISPUTE']
DOMESTIC VIOLENCE	VIOLATION OF PROBATION-MISD	146	['COURT_PROBATION', 'DOM-VIOL_COURT-ORDER']
PROSTITUTION	VIOLATION OF	124	['PROSTITUTION',

PROBATION-MISD 'COURT_PROBATION']

Comps

One way to look at what CrimeCatOAK provides is compare it to an alternate. CrimeMapping.com is paid by the City of Oakland (?? how much ??) to provide a similar analysis. An experiment was performed taking data provided by OPD for the dates 18 - 30 April, 2013. OPD listed 1167 incidents during this period. CrimeMapping reports a total of 496 during this same period, i.e., only 43% of what OPD provides! A summary of the crimes most frequently missed by CrimeMapper's analysis is shown in the table below.

Ctype	Desc	Freq
STOLEN VEHICLE	VEHICLE THEFT - AUTO	78
BURG - AUTO	BURGLARY-AUTO	53
BURG - RESIDENTIAL	BURGLARY-FORCIBLE ENTRY	50
PETTY THEFT	THEFT	38
MISDEMEANOR ASSAULT	BATTERY	36
VANDALISM	VANDALISM	34
DOMESTIC VIOLENCE	BATTERY:SPOUSE/EX SPOUSE/DATE/ETC	32
	VEHICLE THEFT - AUTO	31
ROBBERY	ROBBERY-FIREARM	22

But there's more: CrimeCatOAK provide more refined analysis than CrimeMapping as well. The table below shows how CrimeMapping's "gross" categories are broken up into more refined crime categories within CrimeCatOAK.

CrimeMapper category	Freq	CrimeCat	CCFreq
theft-larceny	153	LARCENY_BURGLARY_AUTO	84
		LARCENY_THEFT_PETTY	36
		LARCENY_THEFT_GRAND	19
		<unclass>	14
assault	93	DOM-VIOL_BATTERY-SPOUSE	32
		ASSAULT_OTHER-SIMPLE_MISDEMEANOR	25

		ASSAULT_THREATS	12
		<unclass>	11
		ASSAULT_OTHER-WEAPON_FELONY	3
		ASSAULT_OTHER-WEAPON_MISDEMEANOR	3
		ASSAULT_FIREARM_FELONY	2
		VANDALISM	2
		WEAPONS	1
		ASSAULT_FIREARM_MISDEMEANOR	1
		LARCENY_THEFT_VEHICLE_AUTOS	1
motor-vehicle-theft	84	LARCENY_THEFT_VEHICLE_AUTOS	84
robbery	44	ROBBERY_FIREARM	22
		ROBBERY_STRONG-ARM	17
		<unclass>	2
		ROBBERY_OTHER-WEAPON	1
		ROBBERY_KNIFE	1
		ROBBERY_INHABITED-DWELLING	1
burglary	44	LARCENY_BURGLARY_RESIDENTIAL	33
		<unclass>	5
		LARCENY_BURGLARY_COMMERCIAL	3
		LARCENY_BURGLARY_OTHER	2
		LARCENY_BURGLARY_AUTO	1
vandalism	32	VANDALISM	31
		<unclass>	1
drugs-alcohol-violations	23	DRUG_POSSESSION_NARCOTICS	10
		DRUG_POSSESSION_OTHER	9
		<unclass>	2
		DISORDERLY-CONDUCT	1
		DRUG_POSSESSION_MARIJUANA	1
disturbing-the-peace	22	DISORDERLY-CONDUCT	18
		<unclass>	3
		DOM-VIOL_PHONE	1
homicide	1	HOMICIDE	1

Summary

ShowCrime is designed as a place-holder: an example of what is possible using even old, dirty data. As I've said before, Oakland and OPD are getting ready to change their internal data collection procedures, and their policies for providing data to the public through data.oaklandnet.com. Certainly a much

better version of ShowCrime should be possible, especially if the city winds up paying vendors to help.

In the interim, ShowCrime could use lots of help. Here are some of the first:

- viewing two CrimeCatOAK crimes on the same plot
- (client-side JS, …) polish and flash
- geographic smarts far beyond beats. Much of the data is geo-tagged, and all we need is an industrial-grade geo-server as part of OpenOakland.org, or similar.
- With geographic support, much more relevant analyses based on census tracts, existing and proposed OPD districts, etc. can be performed
- The details of the CrimeCatOAK could and should be connected to OPD's reporting procedures much more directly.

Come help Openoakland.org make it better!