High School Mathematical Contest in Modeling

Problem B—Curbing City Violence
Prepared by: Team #2561
Date: November 21, 2010

Problem summary

Many people who have never lived in urban areas would associate terms such as boisterous, glorifying and inspirational with the word "city". However, in their picture-perfect image of what a city should look like, these people have failed to understand the dangerous world that is carried out under the radar—the world of crime. Therefore, for everyone's benefit, we need to be able to understand why people commit these crimes in order to stop them, by looking at statistics and analyzing them through math models. This way, we can mathematically find ways to campaign a strategy that will curb the violence for the mayor. These models will be conducted using given data, as well as additional data that has been found in the course of research.

Assumptions

- Unemployment is cyclical, because it is cyclical in the short-term. Therefore, we assumed that it was cyclical in the long-term, which is backed by economic theory.
- Graduation rate, county population and parole violations are causative factors without any other variables, because it's possible to create common sense explanations of why they would increase/decrease. Also, there are no known variables that would infect the data.
- This hypothetical county is in the United States of America, and thus follows USA trends, because it has been established to be similar to Monterey County in California, USA.
- All violent crimes are reported, because there is simply nothing we can do about unreported crimes, and there is no way to determine reporting rates or otherwise estimate unreported crimes with the statistics given.
- The government cannot do much about families in general because that would be interfering with people's personal lives and the factors that deal with families are too complex.
- The words "curbing violence" in the question refers to violent crimes and disregards property crimes because violence is the act of one person harming another, and property crimes does not include those acts.

PART I. Multiple Regression Model

We decided that we should first create a model using only the data provided, as it would create an accurate fit of the hypothetical city given. It was then decided that the multiple regression model would be the best representation, as it would embody most of the data provided. We started by this process by classifying the given variables into four independent groups of explanatory variables:

1. Population (city population and county population)
2. Education (high school enrollment, high school graduation, high school dropout)
3. Unemployment (unemployment and unemployment rate)
4. Incarceration (juvenile inmates, prison population, population released on parole, parole violations, % of population violations)

It was then determined that time and incidents of violence, homicides and assaults were not causative variables. Time is not a causative variable because it is the cause of the explanatory variables, and not a direct cause to the data. The incidents of violence, homicides and assaults are not a causative variable because they are the response variables, and also not a direct cause to the data.

PART IA: Incidence of Violence Model

In order to account for all incidence of violence, we combined assaults and homicides to form the total incidence of violence.

Then, we just used calculator regressions to compare the correlation coefficients (r-squared) of the given variables and the incidents of violence. The correlation coefficients are on the following table:

Table 1.1: Correlations Between Explanatory Variables and Violent Crime

Independent groups of explanatory variables	Explanatory Variables	Incidents of Violent Crime
Population	County Population	.29
	City Population	.24
Unemployment	Unemployment	.011
	Unemployment Rate	.031
Education	High School Enrollment	.34
	High School Dropout	.39
	Graduation Rate	.55
	Juvenile Inmates	.0005
Incarceration	Prison Population	.11
	Parole Releases	.012
	Parole Violations	.13
	% Of Parole Violations	.24

The three highest variables with the highest coefficient correlation were graduation rate, high school dropout and high school enrollment respectively. However, because they are in the same independent group of explanatory variables, we only picked graduation rate, as it was the highest of the three. We did the same process for the other three independent groups.

The reason that we only picked the highest correlation from each of the four independent groups is because ideally, we would want the data that is represented to be as separate and independent as possible. If we picked two variables from the same category, it would cause too much overlap because they are too interrelated, thus creating an inaccurate model.

With that knowledge, we then used the four variables, which were county pop, unemployment rate, graduation rate and % of parole violations, to create a multiple regression model. However, we disregarded unemployment rate because it is irrelevant (as seen in Figure 1.1), as seen because of the lack of any sort of correlation between unemployment rate and total incidences of violent crime. We then made linear regressions for each of the three variables in order to take the average of regression predictions. Finally, when we combined the three linear regressions together, we weighted each of the three equations by their correlation coefficient, and compiled it into one single multiple regression model. The correlation coefficient of the selected variables added up to 1.08 due to slight overlap, so we divided the result by 1.08 to get a multiple regression model to predict how county population, graduation rate and % of parole violation combined influence incidences of violent crime.

Figure 1.1: Incidents of violent crime as a function of unemployment rate

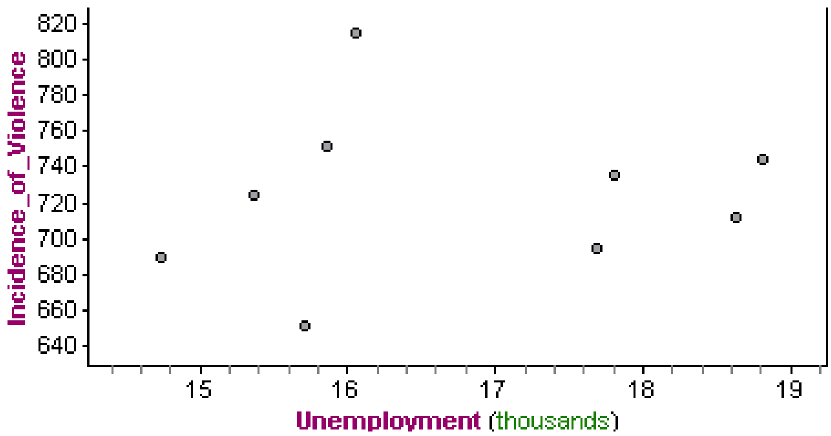


Figure 1.2: Multiple Regression Model—Attempt #1

(3 Variables: county pop, graduation rate, and % of parole violation)

<u>Equation:</u>

Incidences of violence =

$$\frac{1}{1.08}\left(0.29\left(-0.002894 CountyPop+1932\right)+0.55\left(-1234 GraduationRate+1813\right)+0.24\left(878\%ofParoleViolation+123\right)\right)$$

<u>Graphs</u>:

Figure 1.2.1: (Predicted model)

Elastic multiple regression model for incidence of violence

as a function of time (using the given statistics for each variable)

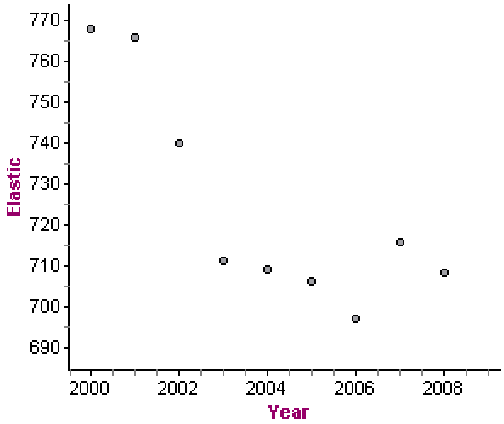


Figure 1.2.2: (Actual model)

Incidence of violence as a function of time

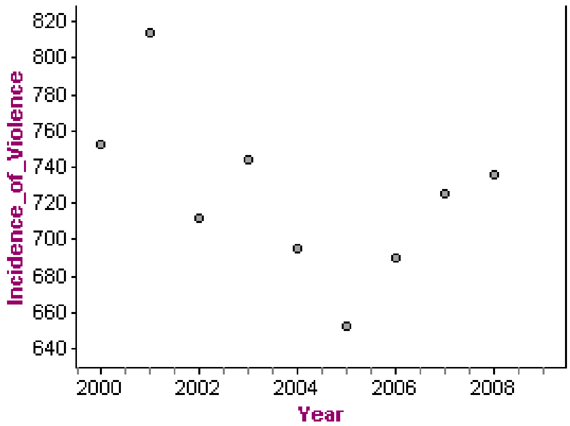
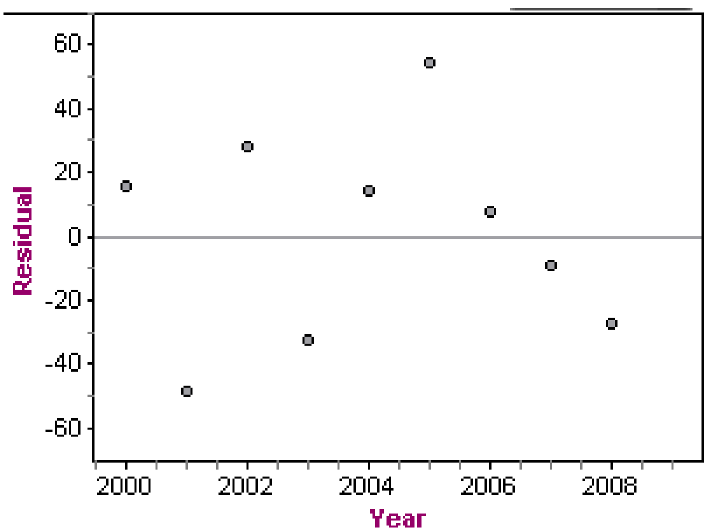


Figure 1.2.3:

Residual graph of model



While the residuals comparing the actual graph and the predicted graph and are significant, they are fairly random and are within a reasonable margin considering how unpredictable real life statistics often are. From there, we attempted to solve for each of our three causative functions as a function of time in order to create a projection of the incidences of violence over time at current rates.

We started off by using matrices to solve for polynomial functions that would pass through every point on the statistics given using the following method

[A]:	[B]:
[1^0 1^1 1^2... 1^8]	[Value for 2000]
[2^0 2^1 2^2... 2^8]	[Value for 2001]
[3^0 3^1 3^2... 3^8]	[Value for 2002]
...	...
[9^0 9^1 9^2... 9^8]	[Value for 2008]

Performing the operation $[A]^{-1}[B]$ gave nine coefficients [a,b,c,d,e,f,g,h,i] which could be entered into the following function:

$f(x) = ax^0+bx^1+cx^2+dx^3+ex^4+fx^5+gx^6+hx^7+ix^8$

This gave the following equations where x is time (in years since 1999):

A. County Population =
421679-61236.870x+69128.652x^2-37738.153x^3+11951.481x^4-2256.281x^5+248.507x^6-14.696x^7+.360x^8
B. Graduation Rates =
.62+.681x-.793x^2+.443x^3-.130x^4+.021x^5-.002x^6+.00007x^7-.00000079x^8
C. % Parole =
2.786-5.105x+4.809x^2-2.339x^3+.659x^4-.113x^5+.012x^6-.00065x^7+.000016x^8

These octic functions in theory should fit the points given perfectly, but in practice they did not always fit the trends, since there was a lot of rounding involved. The rounding was significant enough that the equations were often nowhere close to the actual. This is especially important since this model goes up to the 8th degree, so even the ten-thousandths place is very significant. Furthermore, these functions were simply too cumbersome for practical use. As a result, for purposes of simplicity, a much simpler and direct cubic or quartic regression function was determined for all three variables.

Equations where x is time (in years since 1999):

A. County Population=

$$87.733x^3-1621.357x^2+11591.862x+390800.802 \quad (R^2=.984)$$

B. Graduation Rates=

$$.000161x^4-.00337x^3+.0213x^2-.032x+.850 \quad (R^2=.842)$$

C. % Parole=

$$-.000423x^4+.0085x^3-.0607x^2+.147x+.610 \quad (R^2=.753)$$

However, % Parole did not work well as a function of time, and the newly created violent crime projection as a function of time was much less accurate than the original elastic equation, a disparity only increased over time. As such, for the purposes of creating a function of time we replaced % Parole with the total parole violations, and from there, we created a new multiple regression model.

We then used the same steps as the previous multiple regression model to formulate this new one. This involved calculating the linear regressions for each of the three variables, then taking the average regression prediction and finally combining the three equations by weighting each of them by their respective correlation coefficient. The correlation coefficient of the selected variables added up to 0.97 and so we divided the result by 0.97 instead of 1.08 to get an accurate multiple regression model.

Figure 1.3: Multiple Regression Model—Attempt #2

(3 variables: county population, graduation rates, and parole violations)

Equation:

Incidences of violent crime =

$$\frac{1}{0.97}\left(0.29(-0.002894\,CountyPop+1932)+0.55(-1234\,GraduationRate+1813)\right.$$
$$\left.+0.13(0.00266\,ParoleViolation+495)\right)$$

Graph:

Figure 1.3.1: (Predicted model)

Elastic multiple regression model for incidence of violence

as a function of time

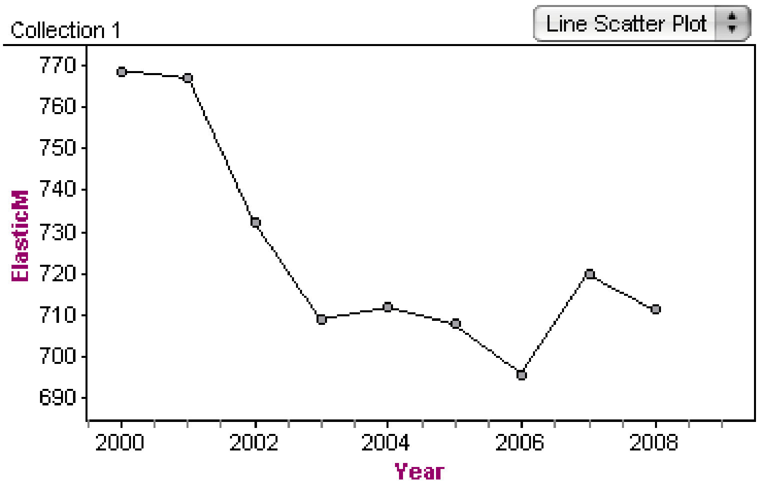


Figure 1.3.2: (Actual model)

Incidence of violence as a function of time

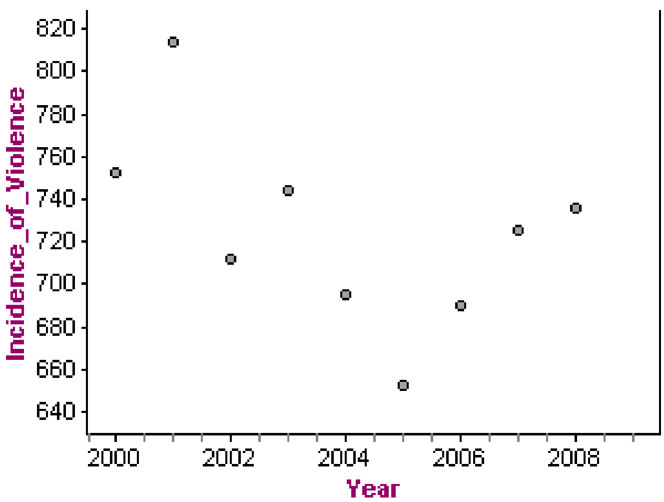
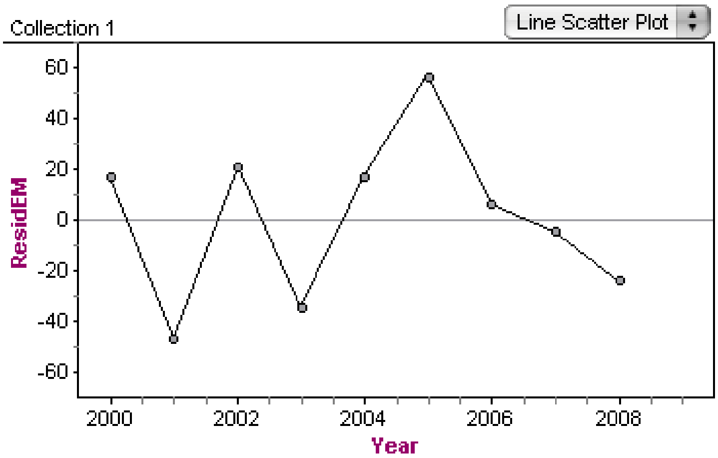


Figure 1.3.3: (Residuals)

Residual graph of predicted model against actual model



We found a new quartic model, after discovering that the octic model was not logical reflecting time against the parole violation count. Like in the previous procedure, we used cubic and quartic equations to model each of the three variables as a function of time. However, this time, all three variables including parole violation were good fits as a function of time. We then substituted these three models for the variables in the final equation. From there, we simplified the final equation to get incidence of violence as a function of time. (Please refer to Appendix III for the calculations.)

Graph:

Figure 1.3.4: (Final multiple regression model)

Incidence of violent crimes as a function of time

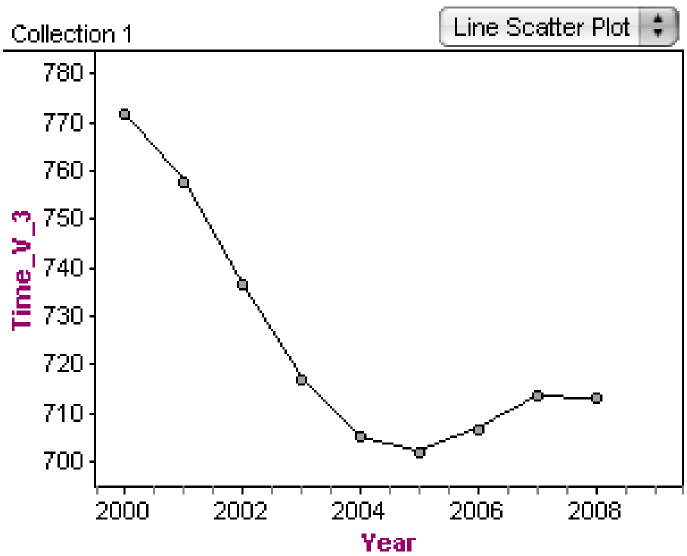


Figure 1.3.5: (Actual model)

Incidence of violence as a function of time

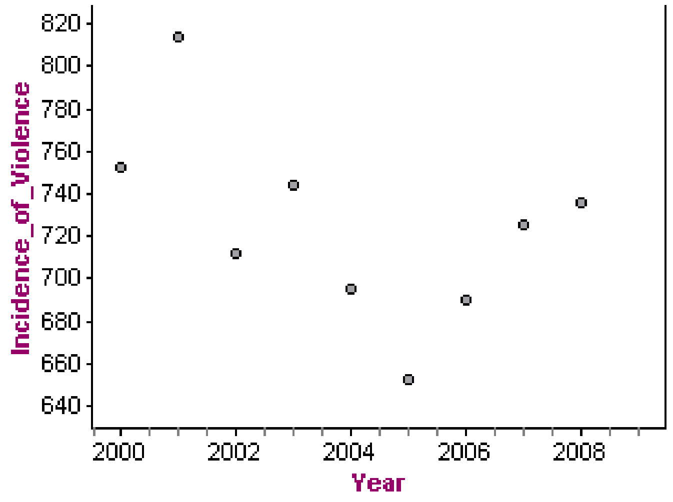
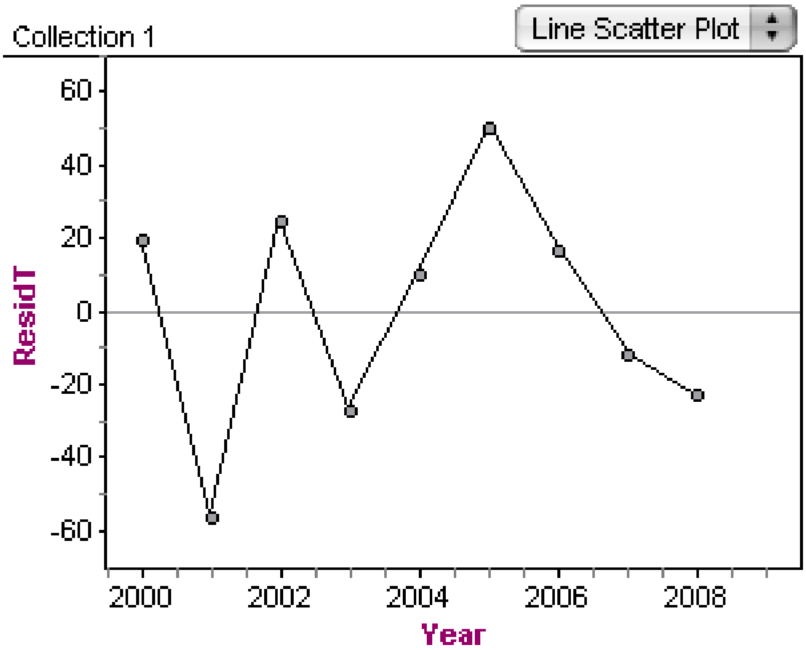


Figure 1.3.6: (Residuals)

Residual graph of final regression model against actual model



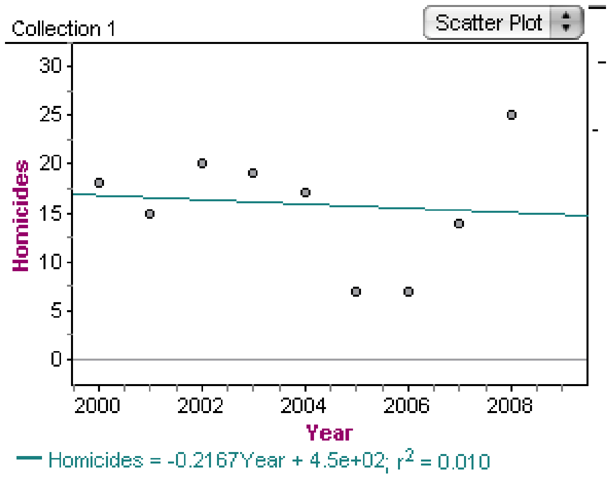
The residuals are fairly random, and so we concluded that this model is a fairly good fit for the data given.

PART IB: Homicide Model

As homicides are more violent than assaults, we decided to utilize the same procedure to find a model specifically for homicide rates.

Step 1: Plotting the given data

Figure 2.1: Actual Homicides as Function of Time



Step 2: Derive the correlation coefficient for each variable compared to homicides

Table 2.1: Correlation coefficients of each variable VS homicides

Independent groups of explanatory variables	Explanatory Variables	Homicides
Population	County Population	0.003
	City Population	0.0083
Unemployment	Unemployment	0.56
	Unemployment Rate	.51
Education	High School Enrollment	.058
	High School Dropout	.082
	Graduation Rate	.10
	Juvenile Inmates	.0043
Incarceration	Prison Population	.054
	Parole Releases	.00093
	% Of Parole Violations	.059

Step 3: Determine the most important variable from each of the four independent groups and graph them against homocides.

a. City population
b. Unemployment
c. High School Graduation Rate
d. % of Parole Violations

Figure 2.2.1: Homicides as a Function of City Population

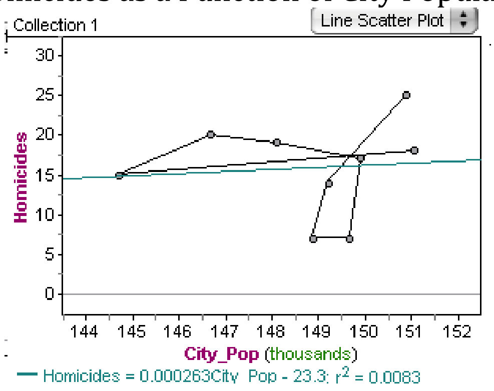


Figure 2.2.2: Homicides as a Function of Graduation Rate

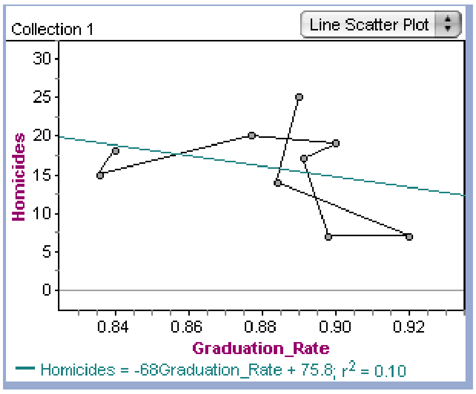


Figure 2.2.3: Homicides as a Function of Parole Violation

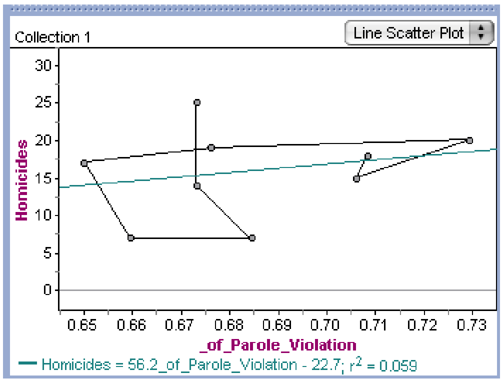
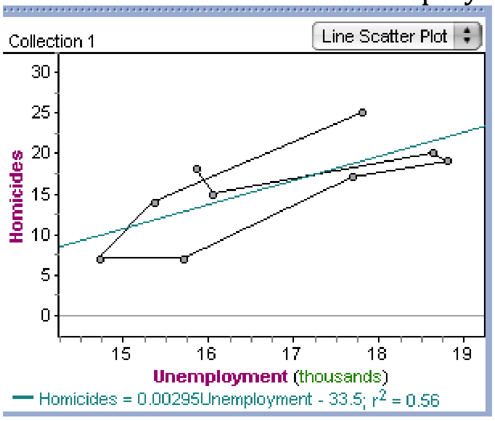


Figure 2.2.4: Homicides as a Function of Unemployment



Step 4: Using these variables to create an accurate multiple regression model:

$$\text{ElasticH} = \left(\frac{1}{0.7273}\right)\left(0.0083\,(0.000263\,\text{City\_Pop} - 23.3) + 0.1\,(-68\,\text{Graduation\_Rate} + 75.8) + 0.059\,(56.2\,\text{\_of\_Parole\_Violation} - 22.7) + 0.56\,(0.00295\,\text{Unemployment} - 33.5)\right)$$

We then plotted the residuals to see the function's accuracy, and compared it to the original data by plotting a two-variable graph.

Figure 2.3.1:

Homicides (predicted in blue, actual in grey) as a function of time

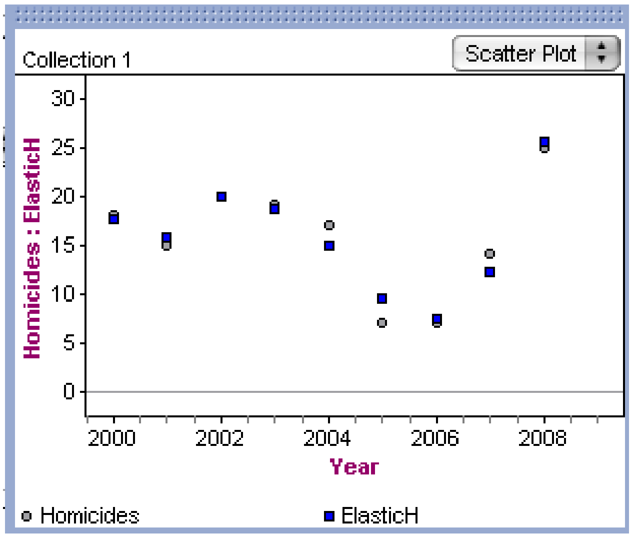
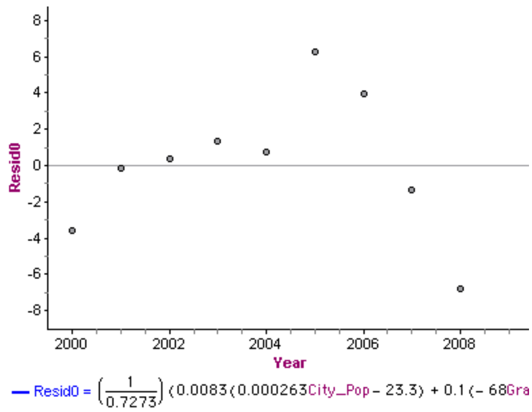


Figure 2.3.2: Residuals of Multiple Regression Model of Homicides



The residuals looked very systematic, as they seem to embody a perfect quartic function, which indicates that the model is not an accurate fit. So we then made a regression for the residuals, and subtracted it from the original function so that it would be a more accurate projectile.

Step 5: By changing the original function to adjust for known inaccuracies in the model, we made a new multiple regression model:

$$
\begin{aligned}
ElasticH = &\left(\frac{1}{0.7273}\right)(0.0083\,(0.000263\,City\_Pop - 23.3) + 0.1\,(-68\,Graduation\_Rate + 75.8) \\
&+ 0.059\,(56.2\_of\_Parole\_Violation - 22.7) + 0.56\,(0.00295\,Unemployment - 33.5)) \\
&- (-0.0234438721\,(Year - 2000)^4 + 0.2678457404\,(Year - 2000)^3 \\
&- 1.101750172\,(Year - 2000)^2 + 3.153249553\,(Year - 2000) - 3.264802607)
\end{aligned}
$$

We then graphed this new multiple regression model with the actual homicides to see how accurate of a fit this new model was. We did a color-coded two-variable graph so that it would be easier to compare the predicted data and actual data.

Figure 2.4.1:

Homicides (predicted in blue, actual in grey) as a function of time

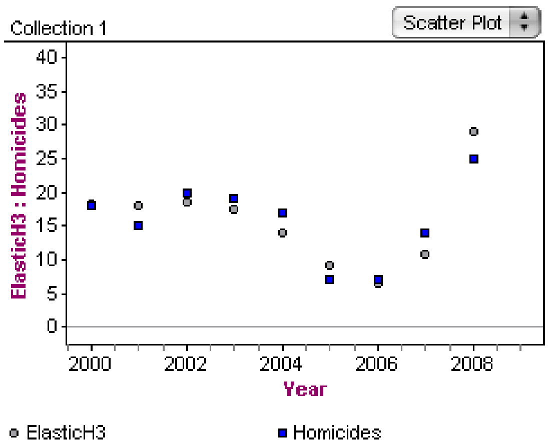
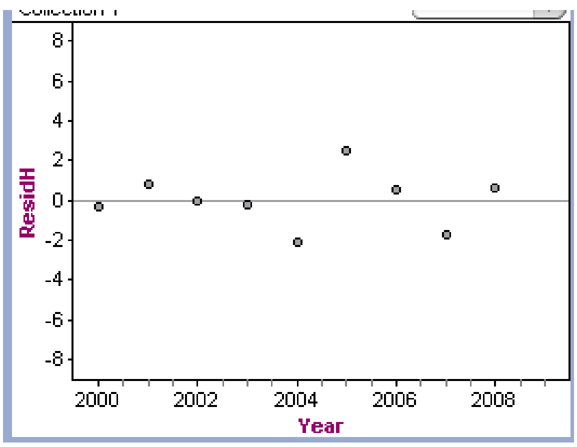


Figure 2.4.2: Residuals of New Multiple Regression Model of Homicides



We realized that these residuals were quite random, so we went onto the next step, which is to substitute each variable and simplify the multiple regression model to one final equation with time as the only variable. The resulting equation is as followed:

$$\text{ElasticH3} = \frac{1}{0.7273}\left(0.0564056248\,(\text{Year}-2000)^4 + \left(-0.639378778\,(\text{Year}-2000)^3\right) + 1.795595281\,(\text{Year}-2000)^2 - 1.357237959\,(\text{Year}-2000) + 13.27192998\right)$$

This fairly accurately models the homicide rates over the range of the data, as well as showing unemployment to be by far the largest contributor to homicide rates.

PART IC: Cyclical Model

As unemployment was assumed to be cyclical, and homicide is closely correlated to unemployment rates, we assumed that homicide was cyclical as well and attempted a cyclical change model to model this relationship.

We started with manual fit models for unemployment and homicide, which made it immediately clear that unemployment was more strictly cyclical than homicides, as the sin models were much closer fits.

Figure 3.1: Manual fits for Unemployment and Homicide

Figure 3.1.1: Manual fit for Unemployment over Time

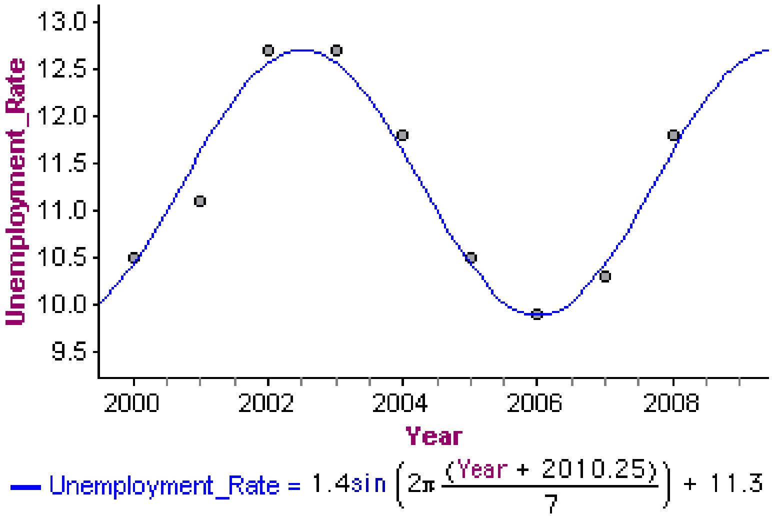
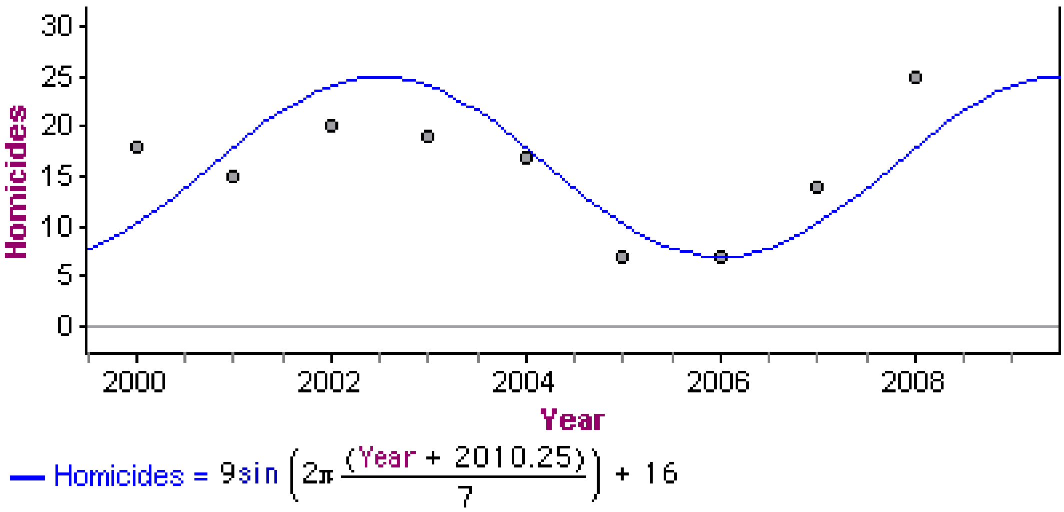


Figure 3.1.2: Manual fit for Homicide over Time



We then took the parent function for cyclical change:

h'=(au-1)h

u'=(1-ah)u

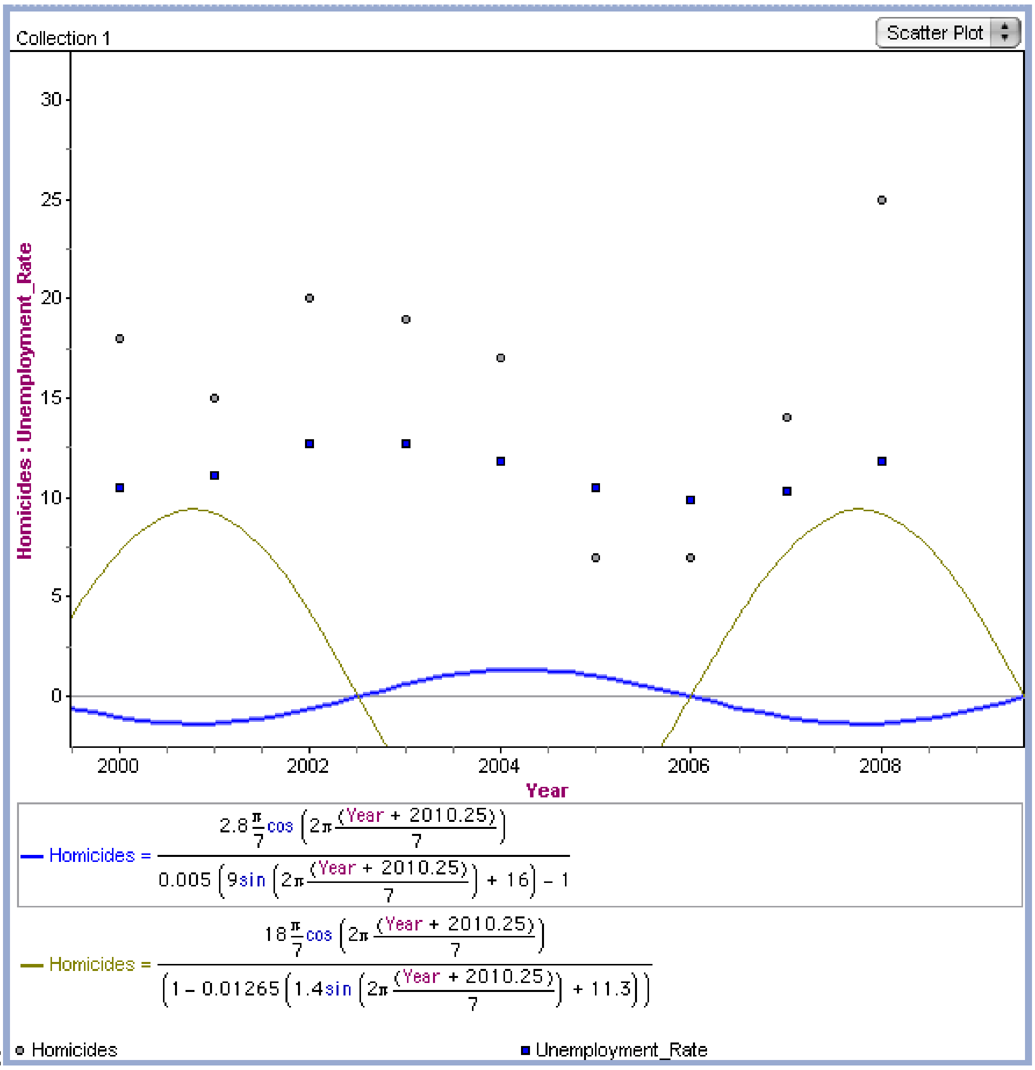
where h is homicides and u is unemployment, and a is the ratio of the linear regressions (which was 0.01265).

As this was a new function form that we were experimenting with, we then solved for the derivatives of each sin function so that we could check whether or not we correctly modeled the cyclical change

h'=18pi/7 cos(2pi(x+2010.25)/7)

u'=2.8cos(2pi(x+2010.25)/7)

We then modified the parent functions to u=(h'/h+1)/a and h=(u'/u-1)/a, and graphed



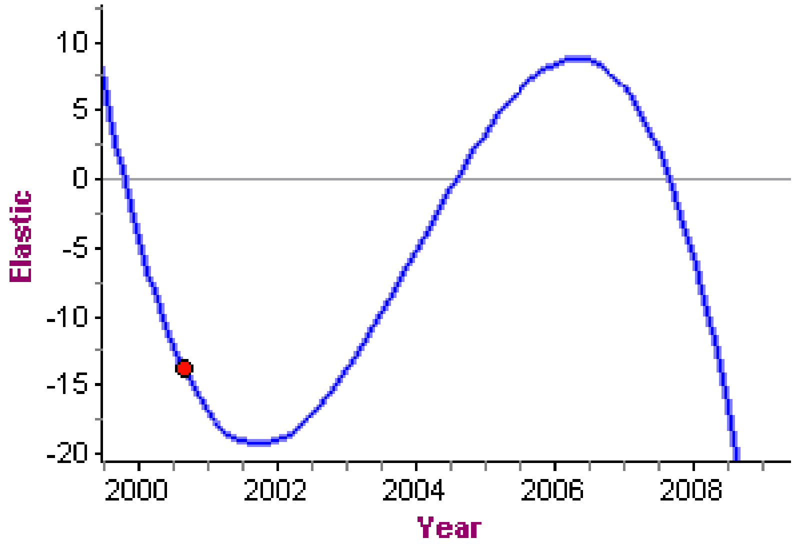
it:

As this function is clearly wrong, it shows that either there was some error in our information, our work (which is unlikely, as we had checked quite extensively) or that the data simply did not fit a cyclical model. Judging by mediocre fit of the sine function modelling, it seems most likely that homicides was simply not a cyclical function, but rather a cubic or other odd polynomial functions.

PART ID: Error Analysis

Figure 4.1:

The differential of the elasticity model for incidence of violence



A large part of the error for the multiple regression model compared to incidence of violence is likely caused by natural variation, as most real life statistics have so many variables that even if all of the major variables are considered, it is still only a rough estimation.

Our model was based solely on the given statistics, so it probably doesn't do that well, as it disregards statistics that were not given, such as inflation rates or the affects of ethnicity and gender. Moreover, from the given statistics, we have only considered the main contributors among the given variables, and have disregarded the other also important factors in the same group.

In addition, even the function of time, which is capable of modeling beyond the range of statistics, was designed only considering the statistics that were given (as we obviously could not consider information we did not have). As the derivative clearly shows, the function will simply continue to negative infinity as x approaches infinity, which may provide some sort of projection in the immediate short term (as the violent crime was falling in the last few points given), but will gradually become further and further from the actually and cannot be reliably used to estimate values beyond one or two years in the future.

Although this was important in avoiding overlapping variables, statistics from society are usually connected in complex and difficult to measure ways, so the variables are probably not entirely independent even if there is no obvious overlap.

This also brings up the question of lurking variables. For instance, both high school graduation and parole violations could be influenced by variables such as family income and drug abuse. In addition, the function assumes causation between education, population, parole violations and incidences of violent crime. However, in reality, these variables are clearly correlated, which does not imply causation.

While common sense dictates that population changes influence crime rates rather than the other way around, the relationship between graduation rates or parole violations and violent crime are less clear. It is very possible that parole violations are caused by widespread violent crime, as being surrounded by crime logically makes it easier for those on parole to lapse back into crime. Likewise, the relationship between crime and education most likely goes both ways, as high school graduates are less likely to commit crime. However, the graduates who are inclined to become criminals due to their social environment or personality are also less likely to graduate.

Similarly, for homicide rates, there is no real reason why homicide rates should cause unemployment. Thus, unemployment is most likely a causative factor. Factors such parole violation and graduation rate are likely share the common cause of prevalence of crime rather than directly causing increases in homicide rates. Because of this, while our model can be used to estimate homicides, or (with some adjustment) reporting rates for homicides, it is not much more reliable than that beyond that point.

Despite all of the uncertainties involved in the model, it is a reasonably good fit for the 9 years covered by the data, and thus should be useful for interpolation regardless of causation. It also clearly establishes that education is the factor most clearly linked to crime. While the assumption that education causes less crime is not proven, it is still a reasonable assumption, and correlation is useful information as well.

However, the first two incarnations of the model use given data as the explanatory variables, and thus cannot be used for any years outside of the given statistics. The third is reliant on imperfect models of each of the causative variables, and while it remains within a reasonable margin of accuracy for interpolation it would most likely be significantly off for any extrapolation beyond the immediate short term, as was clearly shown by the derivative. Long term statistics involve many more variables measured to much more degrees then we found or were given, and even then tend to be inaccurate. In addition, while the model is relatively accurate regardless of whether the variables in question are actually causative, it is fairly uninformative if they are not; if they are not causative variables it would most likely be simpler to directly use the statistics for violent crime.

PART II. Historical Model

Next, we decided it would be appropriate to derive a model using data from historical records in the United States of America, and compare them to the data provided.

We decided to take records from the USA's national records, in addition to four states: California, Illinois, New York and Texas. These four states were chosen upon deciding that they were the stereotypical representations of four major regions of the USA—East, South, West and Mideast. California is known for being the most influential state in the West. Illinois is known for its crime rate because of the history of its gangster capital, Chicago. Texas is known for its stereotypical "dangerous" laws that it enforces to stop crime. New York has New York City, which is known for being the business center of the world, which would naturally draw many criminals. In addition, New York City pioneered many of the changes that became popular throughout the USA's cities.

Following the groupings that were decided in Part I of this paper, we found the data for US national and the four states for the sections of unemployment, education and incarceration. In Part II of the paper however, we relabeled them to be socioeconomic conditions, education and law enforcement respectively. It was also decided that we disregard the group "population" because there is nothing much we can do about controlling population growth. The only widespread measure of controlling population growth thus far is China's One-Child Policy, which goes against both U.S. rights and ideals, thus putting it out of the question. By analyzing the trends of the data, it was possible to find some possible solutions to these factors of violent crime.

PART IIA. Law Enforcement

Table 2.1: Crime Rate Over Time for various regions in the USA

Year	USA Crime Rate	NY Crime Rate	CA Crime Rate	IL Crime Rate	TX Crime Rate
1960	160.9	N/A	239	365.1	161
1964	190.6	N/A	265.6	351.9	190.1
1968	298.4	543.9	422.9	408	270.2
1972	401	754.3	540.7	508.1	354.4
1976	467.8	868.1	669.3	625.8	355.7
1980	596.6	1029.5	893.6	808	550.3
1984	539.9	1069.6	763.4	724.9	505
1988	640.6	1097.3	929.8	810.4	652.6
1992	757.7	1122.1	1,119.70	977.3	806.3
1996	636.6	727	862.7	890.4	644.4
2000	506.5	553.9	621.2	653.8	545.1
2004	463.2	440.4	527.8	545.7	540.9
2008	457.5	398.3	506.2	528.2	508.5

Figures 5.1: Crime Rate over Time for various regions

Figure 5.1.1: USA National Crime Rate over Time

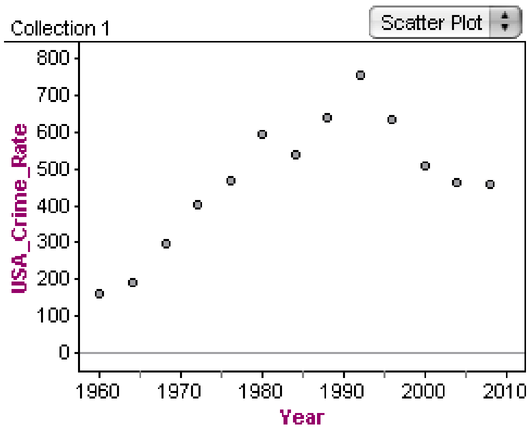


Figure 5.1.2: NY Crime Rate over Time

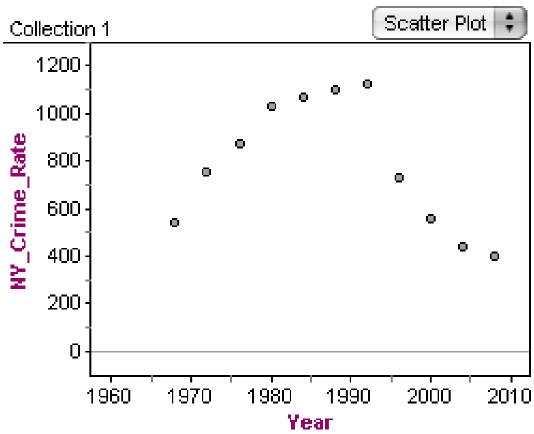


Figure 5.1.3: California Crime Rate over Time

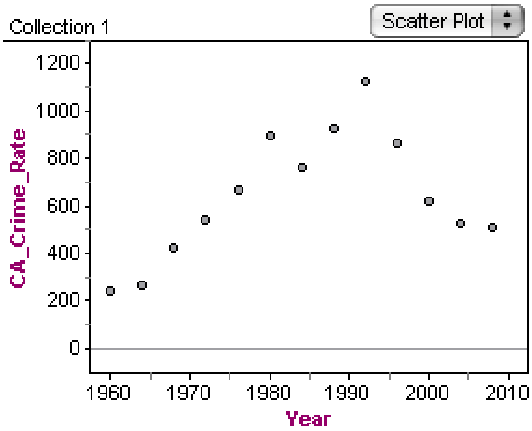


Figure 5.1.4: Illinois Crime Rate over Time

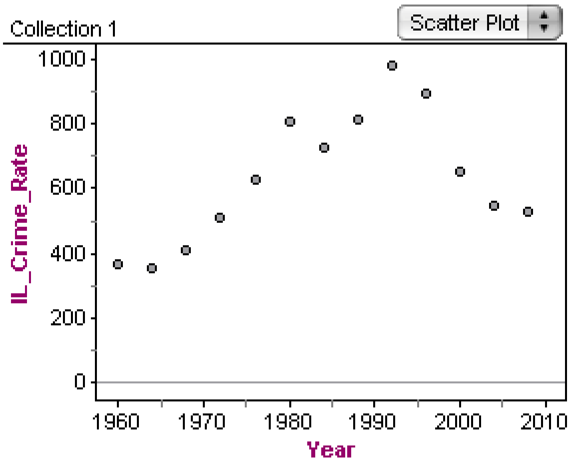
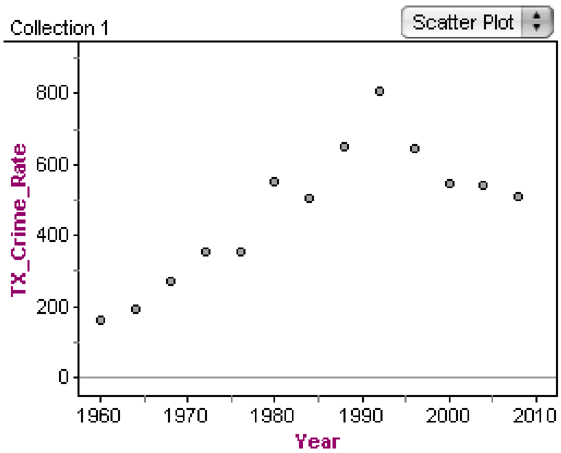


Figure 5.1.5: Texas Crime Rate over Time



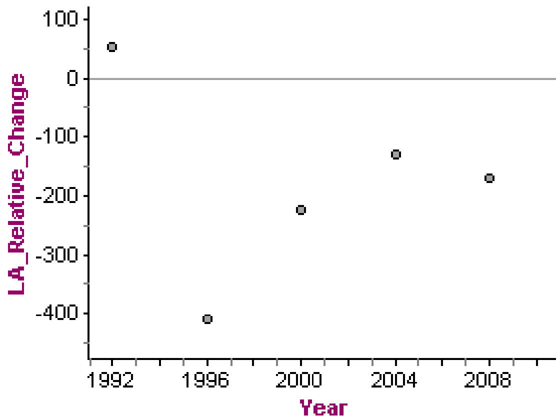
Observations based on the data and graphs:

- Nation experienced spike in crime in late 80s and early 90s
- Most regions followed national trend, however New York had a higher and more sustained peak but a more dramatic fall afterwards.

The fall of New York's crime rate is often attributed to Mayor Rudy Giuliani's policy of CompStat and the Broken Windows theory. CompStat stands for Computer Statistics, a system in which police keep more detailed statistics to ensure that police enforce the law to a higher degree. The Broken Windows theory is a theory that crime is encouraged in certain environments. Therefore, the mayor ordered the police to clean up the city to lower crime rates starting in 1996. As seen in the graph, this is the roughly the same time as the sharp drop in New York's crime rate.

However, CompStat was used in other cities such as Austin, San Francisco, Los Angeles without comparable change in state statistics. This is evident because Los Angeles' change in crime rates varies randomly around the national one, as seen in Figure 2.2. If CompStat was statistically significant, crime rate in Los Angeles should outperform the National Average after the adoption of CompStat in 2002. However, it already began to outperform the National Average in 1996. This may be because CompStat was ineffective, or because of lurking variables that we are unaware of. This evidence suggests that CompStat was not a major contributor to the national decline in crime rate.

Figure 5.2.1: Los Angeles Change in Crime Rate over Time

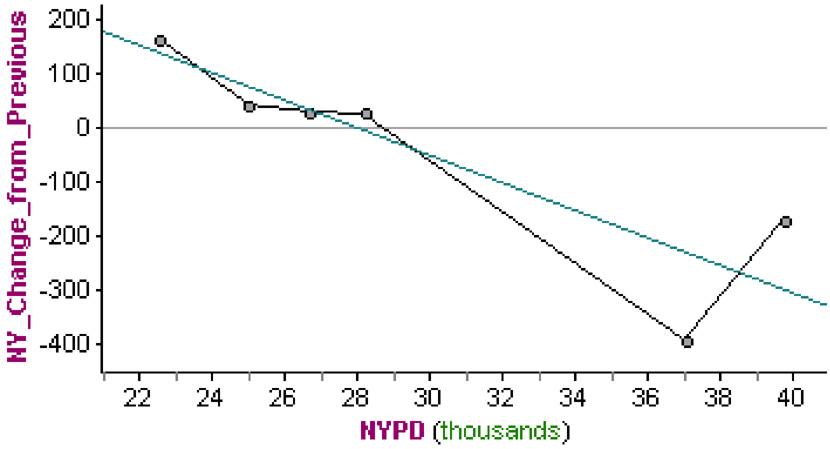


Likewise, the Broken Windows Policy is also highly disputed. There is no real proof anywhere it was implemented whether or not it assisted in lowering violent crimes. This may be due to the fact that enforcing this policy would require a larger police force, which may have been what actually lowered violent crime rates.

An example of this is the increase in police force in New York, which increased 33% between 1992 and 1996. This process began under Mayor Dinkins, and continued under Mayor Giuliani. This has a fairly strong correlation, indicating that the increase in police force is a major impact on the decrease in crime rate.

Figure 5.3.1:

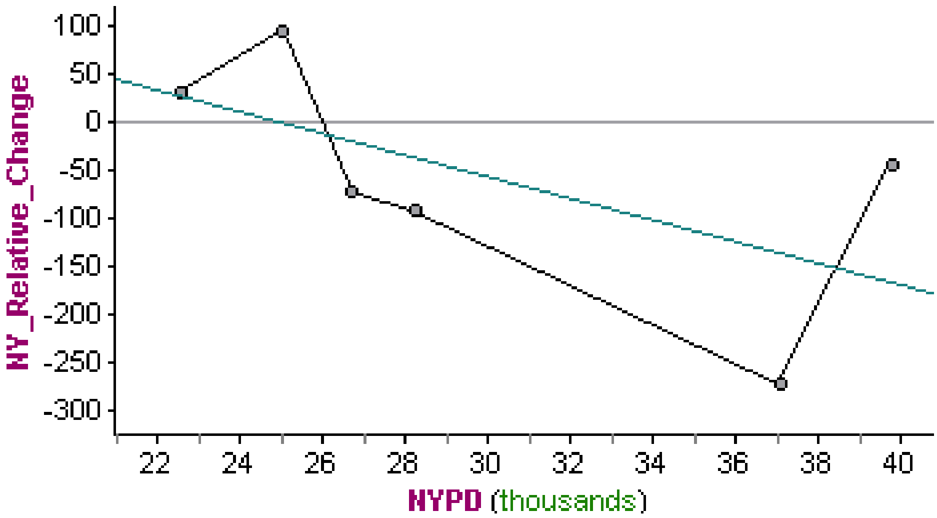
New York's Change in Crime rates as a function of the size of NYPD



Equation: Change in crime rates = -.0253NYPD + 704

r^2: 0.77

Figure 5.3.2: New York's Relative change in crime rates as a function of the size of NYPD



Equation: NY Relative change = -.0114NYPD +280

r^2= 0.38

When you compare Figure 2.3 to Figure 2.4, a graph showing the difference between New York and national growth, the correlation is smaller. Because the correlation is smaller, it indicates that the decrease in crime rate is partially because of other factors, though the change in police force size was still a significant factor. Therefore, the most effective law enforcement change would be to simply hire more policemen, as most of the other factors are out of the mayor's control.

PART IIB. Education

Figure 6.1: Juvenile Delinquency over Time

Figure 6.1.1: US National Juvenile Delinquency over Time

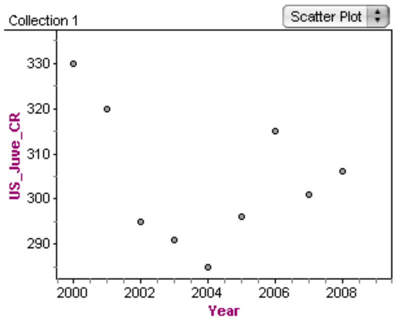


Figure 6.1.2: California Juvenile Delinquency over Time

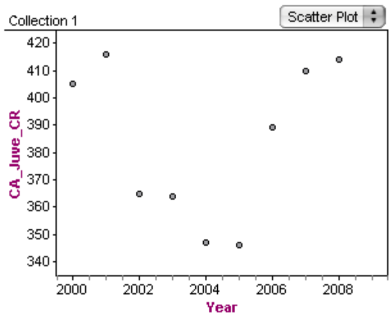


Figure 6.1.3: Illinois Juvenile Delinquency over Time

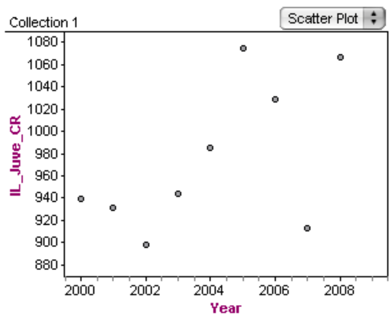


Figure 6.1.4: Texas Juvenile Delinquency over Time

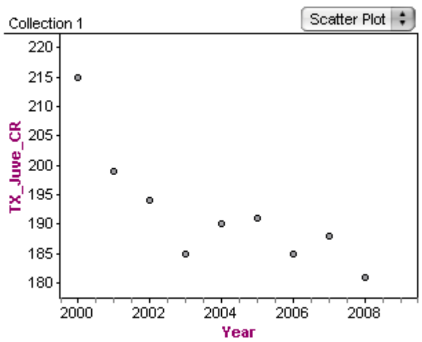
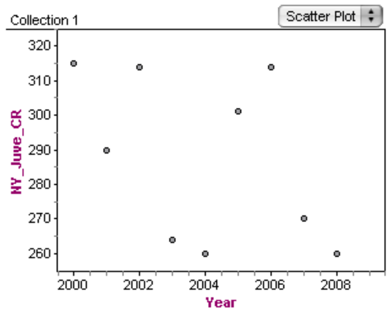


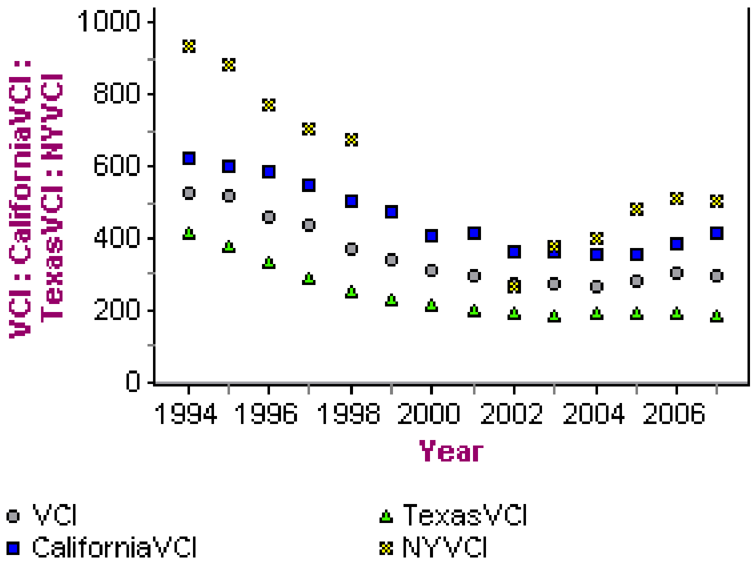
Figure 6.1.5: New York Juvenile Delinquency over Time



Observations based on graphs:

- Texas seems to have a negative correlation between juvenile delinquency and time
- The US National and California seem to follow parabolic curve as juvenile delinquency decreases until 2004 where it begins to rise again
- New York has seemingly random patterns
- Illinois seems to have a positive correlation between juvenile delinquency and time

Figure 6.2: Juvenile Violence Crime Index over states of the USA



Observations of this table:

- The violent crime index of Texas and California seem to model that of the national, except that Texas did significantly better than California.

Figure 6.3: Graduation Rates of the selected states, Texas and California

Figure 6.3.1: Graduation Rates in Texas

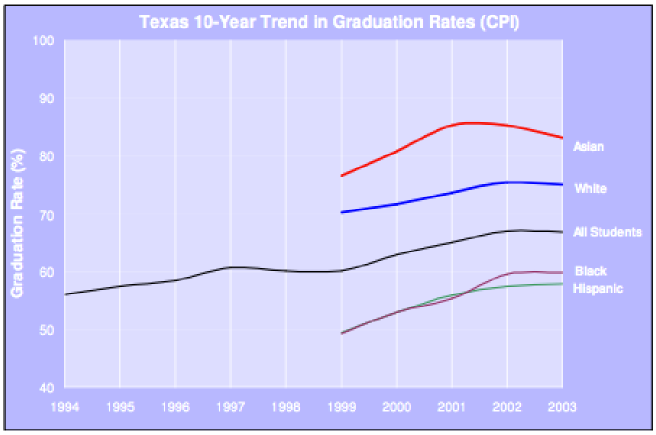
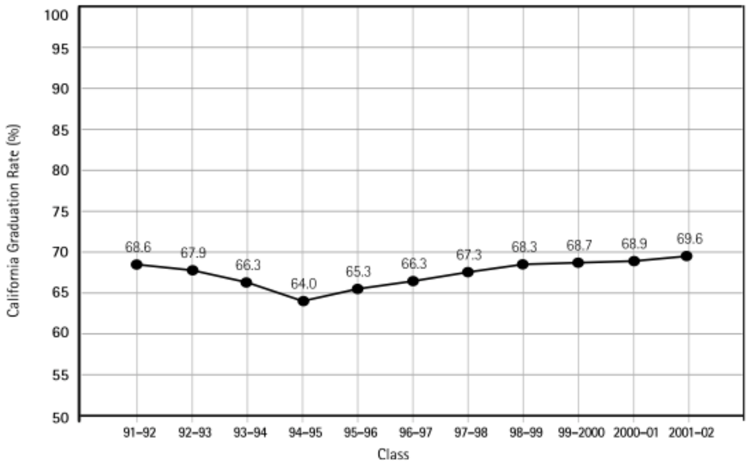


Figure 6.3.2: Graduation Rates in California



Then, we looked into what educational policies the state governments implemented into Texas and California societies. Our discoveries led us to believe that in order for education to be successful, the state government needs to be proactive and not complacent.

Because our data is a representation between the time frame of 1994-2007, we looked at educational policies in that time. Texas state government established the state's first accountability system in 1990 to keep tabs on the public education system based on school districts and camps ratings. Their accountability system is based on the Texas Assessment of Knowledge and Skills (TAKS) test scores.

After keeping these tabs and rating the school districts as“Exemplary,” “Recognized,” “Academically Acceptable” or “Academically Unacceptable” for eight years, the Texan state government realized that core academic courses was essential to the future of students. Therefore, in 1998, they implemented the Texas Essential Knowledge and Skills (TEKS) to ensure the success of their students. This is recognized by the gradual increase of graduation rate between 1994 to 2003 in Figure 6.3.1. The government then changed their TAKS to cover these new standards in the spring of 2003.

For three decades in California, students took the CAP (California Assessment Program) as a standardized test. In 1993, the state government of California implemented CLAS (California Learning Asessment System), a poorly designed educational policy that based heavily on free responses rather than multiple choice. However, this policy caused failure rates to skyrocket so high that which caused graduation rates to decrease until the policy was lifted in 1995, where the graduation rates began to increase again, seen by Figure 6.3.2. It was then boosted in 1998 by the government’s act of implementing STAR (Standardized Testing and Reporting), so that the government could keep track of individual student’s processes.

This analysis of Texas and California, two much larger populations than our hypothetical city, suggest that our assumption of decrease in graduation rates does not positively correlate with increase in crime rate is false. This makes it clear that there are many other lurking variables, which are the main causes of education’s influence on violent crimes.

PART IIC. Socioeconomic Situations

Figures 7.1: Graphs of Annual Household Income vs Number of violent crimes per 1,000 persons age 12 or older

Figure 7.1.1: 1993 Annual Household Income vs Number of violent crimes

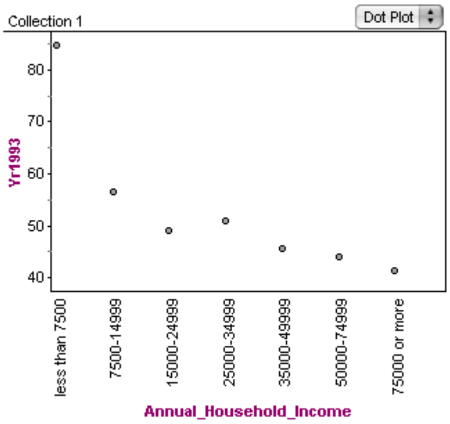


Figure 7.1.2: 1997 Annual Household Income vs Number of violent crimes

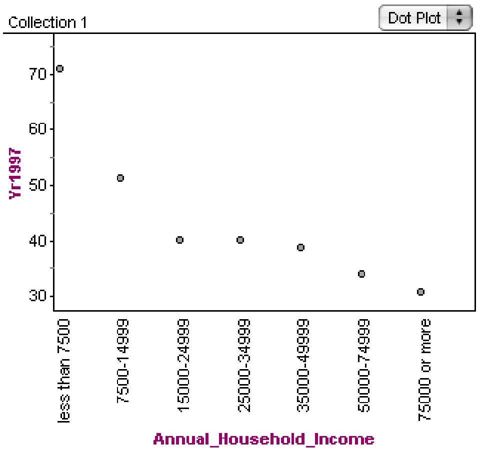
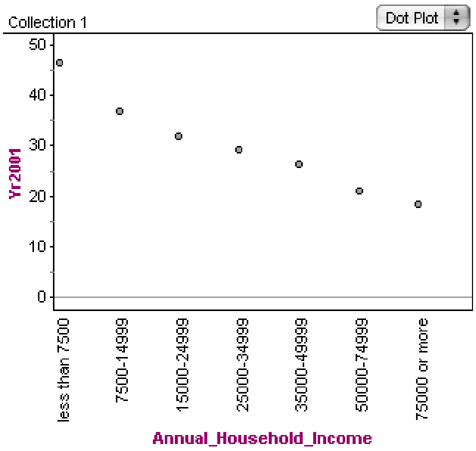


Figure 7.1.3: 2001 Annual Household Income vs Number of violent crimes



Observations from and graphs:

- All graphs show a negative correlation between annual household income and number of violent crimes.

Through these graphs, we came to the conclusion that the more money a household makes, the less likely that people from that household are to commit violent crimes. This then implies that people who are unemployed, who have zero income, would then be the most likely to commit crime. From this analysis, we think that the state government needs to emphasize and help the poorer people, and help them rise faster, instead of letting the "Trickle Down Theory" take its toll.

Therefore, we browsed several populations that have the same problem to see what they did throughout history. A quick glance at these populations showed that they implement a short-term and long-term strategy.

One appropriate and similar model would be in the early 1980s. In 1982, there was a relative peak in crime rates. This was followed by 3 years of reduced crime rate growth. Characteristically, out of the variables we tested, unemployment rates provided the highest correlation of r = 0.59.

Table 3.1: Incidence of crime over Time

Incidences of Crime/ Years after 1982	Incidences of Crime	dy/dx
0	526200	~
1	499390	-26810
2	493960	-5430
3	497560	3600

Let x = Years After 1982

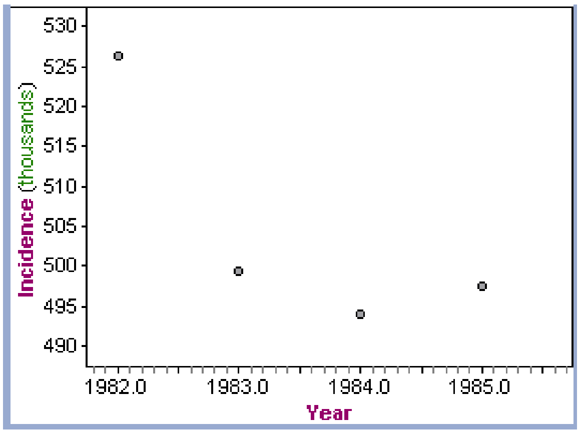
Let y = Incidences of Crime

Performing a regression:

dy/dx = 7.493x^2-6.723x+2003205.951

y^ = 2.498x^3-3.3615x^2+2003205.951x+C

Figure 7.2: Incidence of violence compared to time



In the short run, we may assume that fluctuations in unemployment rate may lead to fluctuations in incidences of crimes. This would lead us to the conclusion that we want to minimize unemployment to minimize number of incidences in crimes. From an economical standpoint, there are several ways to lower unemployment rate.

First, we can achieve this by the Phillips curve, by increasing inflation rates. This would lead to a decrease in unemployment, but is highly inefficient and is thus not preferable.

Second, at a national level, we could change the monetary policy and minimum wage, which would help unemployment rates. If we have expansionary monetary policy, then the monetary authority of a country would control the main supply of the money. This would cause inflation, which increases the value of things you're trying to buy. This would then effectively lower minimum wage, which is set by the sticky wages theory. Then, by lowering the minimum wage, it means that we can have more people working for the same price. This will allow people who are unemployed to have low-paying jobs.

Last, the state government can increase the county's spending at the cost of budget deficit. First, by giving out transfer payments to the unemployed people of low social economic classes, you may decrease the incentive of larceny and theft. Second, by providing unskilled jobs to build on the community infrastructure, you may both advance the infrastructure and curb crime rates.

PART IID: Error Analysis

Both the historical crime rates and juvenile delinquency showed clear signs of lurking variables, however there was insufficient data to pinpoint these variables, which made our models unreliable and our conclusions estimations. In both the nationwide juvenile delinquency rates and many of the individual states, juvenile delinquency began to increase after 2004, however we could find no significance of the year besides the fact it is the year after No Child Left Behind was instituted. As such, it is most likely the culmination of hidden demographic and/or criminal justice trends. Furthermore, California had consistently higher graduation rates than Texas, however Texas had much lower rates of juvenile delinquency. Both states implemented similar forms of education reform, outcome based education, however it was more successful in raising standards in Texas, while it was abolished after just 2 years in California. This suggests that education rates do not directly cause crime, and that the correlation is caused by lurking variables such as quality and content of education, family environments, etc.

In terms of all enforcement, despite all of the efforts of New York and Los Angeles to fight crime, the vast majority of the drop in crime occurred in a national level (including many cities and states which took no action against the crime surge) as well, and is likely the result of nationwide trends such as demography rather than the efforts of local authorities. While there was a significant correlation between the size of the police force and crime rates in New York, we lacked statistics for the police forces of any other city for comparison, and so this connection is somewhat nebulous. In addition, comparisons between cities are already not particularly reliable, as cities are generally too unique to share the same trends even when the same policies are pursued. Also, such trends are only likely to be observed in a large population in the long run. Therefore, any conclusion that we make on the national level may deviate significantly from the county data. Rather, we should have modeled our base line on data that is collected with counties similar to the one outlined by this problem (eg. Salinas, Monterrey County, Ca).

We also found a strong positive correlation between poverty and crime, however most antipoverty measures (such as fiscal policies and job creation) occur at a state or even national level, and so there is little that we can advise the mayor to do.

This study of past policies showed a number of empirically proven but obvious trends, such as that the most effective way to fight crime is to strengthen the police force or that the poor are more likely to commit crimes. However, there were no clear historically successful methods to reform education to reduce dropout rates, nor were there simple ways to fight poverty.

PART III: Letter to the Mayor

Dear Mayor,

After being given the assignment of determining causes of violent crime and the data of your city, and we separated the given variables into four categories: population, unemployment, education and incarceration. We disregarded the category of population because we figured that there wouldn't be many solutions to population growth. We then investigated further into the other three topics, and analyzed the data. Between mathematical reasoning and extensive foreign research, we were able to determine the importance of those three variables, and possible solutions for them.

Firstly, we would like to present the case of unemployment or socioeconomic situations and how it affects violent crime. By comparing data that was related to the relative income of households and the rate of violent crime, it was determined that the less money a family makes, the more likely a person will commit a violent crime. From there, we then concluded that a person who is unemployed and makes no money at all would be even more likely to commit violent crime. Therefore, our solution involves methods of lowering unemployment rates.

However, the statistics we researched counter-intuitively show that there is no significant correlation between unemployment and violent crime rates. This is most likely due to the fact that most crimes committed by the unemployed are property crimes, rather than violent crimes. Because we were assigned to curb the violence of cities, the rate of unemployment is relatively unimportant to the question at hand.

Secondly, we would like to present the case of incarceration or law enforcement and how it affects violent crime. We created a statistical analysis of New York and Los Angeles compared to national standards over the last thirty years. The final result implied that increasing the size of the police force was the most effective measure implemented by state governments during that time, and that implementing changes in the methods used by police such as CompStat and the Broken Windows Policy were generally ineffective.

Lastly, we would like to present the case of education and how it affects violent crime. The first step was to mathematically analyze the data you gave us by determining the correlation coefficient between incidents of violent crime and all the education-related variables. The result was that high school graduation rate was 0.55, high school dropouts were 0.39 and high school enrollment was 0.34. These three education-related constituted the three largest correlations for all the variables. Implying that lack of education is the leading factor of violent crime, these numbers further support our argument by showing us the correlation coefficient between high school enrollment and juvenile inmates is 0.87. This means that on average, for every 100 students that are enrolled in high school, 87 of them who normally would have become juvenile delinquents would have stayed “normal”.

However, efforts to change educational policies in order to curb crime have generally been unsuccessful. In the Three State Recidivism study, education in prisons dramatically lowered the rate of property crimes on release. However, violent crimes such as assaults actually increased amongst some of the criminals.

Furthermore, a comparative study of education and juvenile delinquency between California and Texas showed that even if graduation rates decrease, juvenile delinquencies decreases as well. This suggests that changes in graduation rates were not significant enough to overcome changes in other variables, and the correlation between graduation rates is likely either due to other lurking variables, such as family structure or environment. It is also possible that crime is actually the causative factor of the decrease in graduation rate, and that children who dropped out of school did so because of increase in crime in the area, rather than becoming criminals because they dropped out of school.

Therefore, based on these three studies, the only definite way our statics discovered to curb violence is to raise the size of the police force, though other measures such as education reform and lowering unemployment rates are still worth deliberation.

Sincerely,
Team #2561

Appendix I: Bibliography

1. Bartollas, Clemens. *Juvenile Delinquency*. Boston: Pearson/A & B, 2006. Print.

2. *Bureau of Justice Statistics*. Office of Justice Programs, 16 Nov. 2010. Web. 20 Nov. 2010. <http://bjs.ojp.usdoj.gov/index.cfm>.

3. Combs, Susan. "Public Education." *Texas in Focus: A Statewide View of Opportunities*. Window on State Government. Web. 20 Nov. 2010. <http://www.window.state.tx.us/specialrpt/tif/public.html>.

4. Confrey, Jere; Carrejo, David *A Content Analysis of Exit Level Mathematics on the Texas Assessment of Academic Skills: Addressing the Issue of Instructional Decision-Making in Texas*, Retrieved on 2008-03-26

5. "Easy Access to FBI Arrest Statistics." *Office of Juvenile Justice and Delinquency Prevention*. U.S. Department of Justice. 31 Oct 2009. Web. 20 Nov. 2010. <http://www.ojjdp.gov/ojstatbb/ezaucr/>.

6. Glick, Leonard, and J. Mitchell. Miller. *Criminology*. Boston: Allyn & Bacon, 2008. Print.

7. Gottfredson, Denise C. *Schools and Delinquency*. Cambridge, UK: Cambridge UP, 2001. Print.

8. Hirschi, Travis. *Causes of Delinquency.* Berkeley: University of California, 1969. Print.

9. "Juvenile Arrests." *Office of Juvenile Justice and Delinquency Prevention*. Web. 20 Nov. 2010. <http://www.ojjdp.gov/ojstatbb/crime/qa05103.asp?qaDate=2008>.

10. "Juvenile Crime in California--Introduction." *California Legislative Analyst's Office*. Web. 20 Nov. 2010. <http://www.lao.ca.gov/1995/050195_juv_crime/kkpart1.aspx#howmuch>.

11. "Keeping Illinois Competitive." *Illinois Status Report: Science, Technology, Engineering & Math Education* (2006). Northern Illinois University, June 2006. Web. 21 Nov. 2010.

12. L., Priemaza. "Multiple Regression Analysis." Lecture. *University of Alberta*. University of Alberta, 15 Apr. 2008. Web. 20 Nov. 2010. <http://www.ualberta.ca/~priemaza/.../Econ%20399%20Chapter6a.ppt>.

13. "Labor Force Statistics from the Current Population Survey." *U.S. Bureau of Labor Statistics*. United States Department of Labor, 2010. Web. 19 Nov. 2010. <http://stats.bls.gov/>.

14. Langan, Patrick A., and Matthew R. Durose. "Langan Rel." *Scribd*. US Department of Justice, 21 Oct. 2004. Web. 20 Nov. 2010. <http://www.scribd.com/doc/322928/Langan-rel>.

15. "Population Estimates." *Census Bureau Home Page*. Web. 19 Nov. 2010. <http://www.census.gov/popest/states/NST-ann-est2008.html>.

16. "Prisoners in 1994." *DRCNet Online Library Main Page - Frames*. Web. 20 Nov. 2010. <http://druglibrary.net/schaffer/GovPubs/prs94as.htm>.
Bureau of Justice Statistics (BJS). Web. 20 Nov. 2010. <http://bjs.ojp.usdoj.gov/content/data/corpop12.csv>.

17. Puzzanchera, C. "Trends." *Office of Juvenile Justice and Delinquency Prevention*. Office of Juvenile Justice and Delinquency, 31 Oct. 2009. Web. 20 Nov. 2010. <http://www.ojjdp.gov/ojstatbb/crime/JAR_Display.asp?ID=qa05201>.

18. "School Enrollment." *Census Bureau Home Page*. Web. 20 Nov. 2010. <http://www.census.gov/population/www/socdemo/school.html>.

19. Swanson, Christopher B. "High School Graduation in Texas." *Editorial Projects in Education Research Center* (2006). Oct. 2006. Web. 21 Nov. 2010.

20. United States of America. Department of Education. Office of Correctional Education. *Three State Recidivism Study*. By Stephen J. Steurer, Linda Smith, and Alic Tracy. Correctional Education Association, 30 Sept. 2001. Web. 21 Nov. 2010. <www.ceanational.org/PDFs/3StateFinal.pdf>.

21. "United States Unemployment Data." *North/Northwest Indiana Data Plus - NI DataPlus*. Web. 20 Nov. 2010. <http://www.nidataplus.com/lfeus1.htm#annl>.

22. "United States Unemployment Rate." *Infoplease: Encyclopedia, Almanac, Atlas, Biographies, Dictionary, Thesaurus. Free Online Reference, Research & Homework Help. — Infoplease.com*. Family Education Network, 2005. Web. 19 Nov. 2010. <http://www.infoplease.com/ipa/A0104719.html>.

Appendix II: Glossary

- Causation: when one variable affects another. This implies correlation.
- Causative variable: independent variable
- CompStat: Abbreviation for computer statistics, a system where policy keep statistics in order to monitor the effectiveness of law enforcement within subareas in their jurisdiction. This is intended to improve incentives by tying police pay to the success of law enforcements in their area.
- Correlation: when two variables are linked, but do not necessarily mean causation
- Correlation coefficient (r-squared): The amount of change in y that is explained by the change in x
- Crime: A human who violates the criminal law of a state, federal government or a local jurisdiction that has the power to make laws.
- Explanatory variable: independent variable
- Homicide: the act of one human being killing another human being
- Multiple Regression Model: A model in which the estimation of y is determined by an average regressions, weighted by correlation coefficient.
- Phillips curve: inflation is inversely proportional to umemployment

Appendix III: Solving the Multiple Regression Equation

The following steps were used to determine the first multiple regression equation, plotting the incidences of violence against time.

Step 1: Determine the three cubic or quartic equations for the three variables

A. County Pop =

$$87.7331649813\,2\,(Year-2000)^3 - 1358.1572871569\,(Year-2000)^2 + 8612.3483645972\,(Year-2000) + 400859.04040404040$$

(Cubic Regression of Time vs County Population)

$r^2 = 0.984$

B. Graduation Rate =

$$0.0001611305\,4\,(Year-2000)^4 - 0.0027270785\,(Year-2000)^3 + 0.0121395688\,(Year-2000)^2 + 0.0014252137\,(Year-2000) + 0.8365641026$$

(Quartic Regression of Time vs Graduation Rate)

$r^2 = 0.842$

C. Parole Violation =

$$-111.7622378\,(Year-2000)^4 + 1735.939912\,(Year-2000)^3 - 7480.797009\,(Year-2000)^2 + 6341.596024\,(Year-2000) + 89241.48407$$

(Quartic Regression of Time vs Parole Violation)

$r^2 = 0.977$

Step 2: Combining the three regressions into one final equation

Distributing the values from the three functions:

$$\frac{1}{0.97}\left(0.29(-0.002894\,CountyPop + 1932) + 0.55(-1234\,GraduationRate + 1813) + 0.13(0.00266\,ParoleViolation + 495)\right)$$

(0.29 * 1932) + (0.55 * 1813) + (0.13 * 495) = 1621.78

$$\frac{1}{0.97}\left(0.29(-0.002894 CountyPop+1932)+0.55(-1234 GraduationRate+1813)\right.$$
$$\left.+0.13(0.00266 ParoleViolation+495)\right)$$

Let A = 0.29*-0.002894 = -8.3926 * 10^-4

Let B = 0.55*-1234 = -678.7

Let C = 0.13*0.00266 = 3.458 * 10^-4

Step 3: Combining the distributions from above:

Incidence of violent crimes=

(1/0.97)(A*County Population Equation + B*Graduation Equation + C* Parole Violation Equation + 1621.78)

Step 4: Simplifying the equation:

Incidence of violent crimes=

(1/0.97) (ax^4+bx^3+cx^2+dx+e)

a = 0.00016113054B – 111.7622378C

= -0.1480066793

b = 87.3316498132A – 0.0027270785B + 1735.939912C

= 2.377862239

c = -1358.1572871569A + 0.0121395688B – 7480.797009C

= -9.686137865

d = 8612.3483645972A – 0.0014252137B + 6341.596024C

= -4.067783045

e = 400859.040404040404A + 0.8365641026B + 89241.48407C + 1621.78

= 748.4386905

The following steps were used to calculate the second multiple regression model, plotting homicides against time:

We once again used cubic and quartic regressions to find each variable as a function of time:

A. Homicide vs City

Population

$$EqCityPop = 53.23863636\,(Year-2000)^4 - 915.9823232\,(Year-2000)^3 + 5100.503788\,(Year-2000)^2 - 9395.366162\,(Year-2000) + 150772.5758$$

r^2 = 0.932

B. Homicide vs Graduation

Rate

$$EqGraduation\_Rate = 0.00016113054\,(Year-2000)^4 - 0.0027270785\,(Year-2000)^3 + 0.0121395688\,(Year-2000)^2 + 0.0014252137 + 0.8365641026$$

r^2 = 0.842

C. Homicide vs % of Parole Violation

$$EqParoleViolation = (-0.0004227505\,(Year-2000)^4 + 0.0071595136\,(Year-2000)^3 - 0.0369538917\,(Year-2000)^2 + 0.0497507008\,(Year-2000) + 0.7042580389)$$

r^2 = 0.753

D. Homicide vs Unemployment

$$EqUnemployment = 21.39405682\,(Year-2000)^4 - 249.2840126\,(Year-2000)^3 + 537.4040947\,(Year-2000)^2 + 1005.598558\,(Year-2000) + 15583.62773$$

r^2 = 0.897

From there, we substituted those functions for each of the variables to find homicides as a function of time:

(1/0.7273)(A*City Population Equation + B*Graduation Equation + C* Parole Violation Equation + D*unemployment – Resid -15.97749261) = (1/0.97)(ax^4+bx^3+cx^2+dx+e)

a = 53.23863636A + 0.00016113054B – 0.0004227505C +21.39405682D + 0.0234438721 = 0.0564056248
b = -915.9823232A – 0.0027270785B + 0.0071595136C – 249.2840126D – 0.2678457404 = -0.639378778
c = 5100.703788A + 0.0121395688B – 0.0369538917C + 537.4040947D +1.101750172 = 1.795595281
d = -9395.366162A + 0.0014252137B + 0.0497507008C + 1005.598558D – 3.153249553 = -1.357237959
e = 150772.5758A + 0.8365641026B + 0.7042580389C+15583.62773D + 3.264802607 + (0.0083(-23.3)+0.1(75.8)+0.059(-22.7)+0.56(-33.5)) = 13.27192998

Appendix IV: Miscellaneous Figures

Table 4.1: Juveniles per 100,000 people arrested for violent crime

	US	CA	IL	TX	NY
2000	330	405	939	215	315
2001	320	416	931	199	290
2002	295	365	898	194	314
2003	291	364	944	185	264
2004	285	347	985	190	260
2005	296	346	1075	191	301
2006	315	389	1029	185	314
2007	301	410	913	188	270
2008	306	414	1066	181	260

Table 4.2: Annual Household Income vs Number of violent crimes per 1,000 persons age 12 or older

Annual Income	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
less than 7500	84.7	86	77.8	65.3	71	63.8	57.5	60.3	46.6	45.5	49.9
7500-14999	56.4	60.7	49.8	52.1	51.2	49.3	44.5	37.8	36.9	31.5	30.8
15000-24999	49	50.7	48.9	44.1	40.1	39.4	35.3	31.8	31.8	30	26.3
25000-34999	51	47.3	47.1	43	40.2	42	37.9	29.8	29.1	27	24.9
35000-49999	45.6	47	45.8	43	38.7	31.7	30.3	28.5	26.3	25.6	21.4
50000-74999	44	48	44.6	37.5	33.9	32	33.3	23.7	21	18.7	22.9
75000 or more	41.3	39.5	37.3	30.5	30.7	33.1	22.9	22.3	18.5	19	17.5