2015

18th Annual High School Mathematical Contest in Modeling (HiMCM) Summary Sheet

Team Control Number: 5967

Problem Chosen: B

Summary Sheet:

Crime is a problem shared by essentially all major metropolitan areas. It is often a priority of local and city governments to reduce the spread of criminal activity within their jurisdictions, as a city's safety is often an indicator of its economic success and the happiness of its citizens. As such, it is important to be able to gauge a city's safety with respect to crime in comparison to others, in order to determine if a city is effectively combatting the crime within its borders. This is the purpose behind our mathematical and computational model, which considers many different factors contributing to the safety of a city.

Our model generates a singular, easily interpretable 'crime rating' that can be compared to that of other major cities' in order to judge the overall safety of citizens. To do this, our model was constructed keeping in mind that we would need to be able to evaluate My City as accurately as possible utilizing the massive amount of information provided to us. At the same time, however, we would also need to be able to compare the safety of My City to that of other real-life cities which may not have as readily available data. As such, the model we designed is not only accurate, but flexible.

We constructed our model around the following parameters that would contribute to a city's safety: the number of crimes taking place over a given period of time, the number of people affected by these crimes, the severity of different crimes, the efficiency of police in arresting criminals, the proximity of different crimes to residential areas, and the relative distribution of crime throughout the city. In order to take into account all of the crimes reported and listed in the given spreadsheet, we utilized Microsoft Excel and Java programming to quickly analyze the data and yield results in a timely manner. We were also able to use far less comprehensive statistics from online sources in the same model to make comparisons between My City and real-life cities in order to determine safety.

The simplest version of our model generates a crime rating of 4.4 for the general metropolitan area of My City, and when additional factors are considered, the crime rating reaches 17.47. Downtown Detroit, an area historically known for its crime, received a basic score of 37.7, and received a score of 149.6 when additional factors were projected. Minneapolis, a city deemed the safest major city in the United States by Forbes magazine, received a basic score of 2.6, and received a score of 10.32 when additional factors were projected. The advanced scores required projections for the real-life cities because sufficient data was not available to use the full model, which requires data on the distribution of crime and the rate of arrests. For context, Seattle's greater metropolitan area received a basic score of 2.8 and Los Angeles county received a score of 3.1. This shows that My City is not nearly as unsafe as downtown Detroit but is still lagging behind some major cities in terms of safety.

Dear Mayor:

We understand you were interested in a professional evaluation of your city's safety. We understand the nature of your request, as such a measure would not only provide insight to you about future city planning, but would also be a compelling statistic to aid prospective citizens with their real estate decisions. In our model we took into account many of the factors contributing to public safety that we could derive from the data you provided to us in the Excel spreadsheet. Utilizing Excel and our own programming, we were able to create a model that would generate a 'crime rating' for your city that could be compared with other major cities' to determine the safety of your citizens.

We used numerous other cities as points of reference with which to compare yours. In order to do this we had to limit the types of data we could include in our model, because it was difficult to find detailed statistics on all crime for additional cities. Initially we considered only total population, the amount of each type of crime, the severity of different types of crimes, and the number of people victimized. Using these factors alone, our model yielded a crime rating of 4.4 for your city. To put this in context, Minneapolis, one of the safest cities in the country according to Forbes magazine, received a rating of 2.6, and downtown Detroit, a city historically known for its criminal activity, received a 37.7. We also conducted additional investigations on the Seattle metropolitan area and Los Angeles county, which received comparable scores of 2.8 and 3.1 respectively. Thus, while your city isn't ranked quite as safe as the safest in the nation, it is still far safer than downtown Detroit. This isn't to say that your city is without crime, however. In the process of calculating the overall crime rating we also found that over 10,000 crimes took place in your city in just a short two week period, with an average of 714 per day. By comparison, Minneapolis has an average of 64 crimes per day- if you account for the population difference, if Minneapolis had the same population as your city, there would be approximately 149 crimes per day. This shows the degree to which improvements could be made for a city the size of yours. While My City is clearly in good hands, we would still recommend that you pursue further measures to combat crime in your city.

Although these numbers alone can provide you a good understanding of the situation of crime in your city, we expanded upon our results with more factors to provide you with a more accurate rating. As a consequence of this consideration of additional factors, however, it becomes more difficult to make reliable comparisons to other cities. These new factors included the uniformity of crime distribution, the percent of people arrested for their crimes, and the locations of the crimes with respect to residential areas. The safety of your city, we figured, would also depend on these factors. With these new metrics we calculated a new crime rating of 17.4 for your city. This value cannot be directly compared to the aforementioned safety rankings of Minneapolis and Detroit because the number of crimes considered and the new metrics make it distinguishable. We did provide theoretical projections you could use for other cities, which include 10.3 for Minneapolis, and 149.6 for Detroit, but these reference points are somewhat inaccurate because they aren't based on real data from these cities. Ideally, you should contact mayors of the other cities in order to procure more accurate crime statistics, which we can then use to provide better and more accurate reference points with which we can compare your city. This will allow you to get a better sense of how safe your city is in the context of the safety of others.

We hope that you find the flexibility and efficacy of our model compelling, and adopt it for future application; we also suggest asking other cities for data on their crimes to help improve the accuracy of our model.

Thank you for contacting us to create this model for you,
-Team 5967

1 - Introduction

1.1 Background

Over a period of time, based on individual police reports and other methods of data entry, it is possible to collect and organize large amounts of information about the abundance and types of criminal offenses that take place throughout a city. These statistics are important in how they can be used to judge the overall safety of a city, which is an important factor that contributes to economic sustainability, population growth, and overall happiness, among other things. My City is a hypothetical city which has a downtown population of 2.8 million people and an additional 6 million people living in the surrounding metropolitan area, yielding a total population of 8.8 million people. The mayor of My City has provided a spreadsheet detailing all of the reported crimes within My City over a two-week period, in hopes that we would be able to quantify and evaluate the safety of My City as a whole.

1.2 Restatement of Problem

The inherent challenge in providing a concise analysis of My City lies in the massive amount of information that must be analyzed to do so. Our group decided that in order to create an applicable ‘safety rating,’ a mathematical model would be necessary. This model would have to take into account all of the factors that contribute to a city’s safety, which may or may not have been present in the given spreadsheet. The safety rating must be easily interpretable by the mayor of My City and must also be comparable to the safety rating of other well-known cities. The accuracy of our model as a whole can be gauged by its ability to accurately rate existing cities and by the number and relevance of factors that are considered in judging the city.

2 - Definitions, Assumptions, Justifications, and Variables

2.1 Definitions

Here we will define terms that will be used throughout the paper:

Crime rating: This is the singular value that we are attempting to calculate which will describe the impact of crime on a city. It will be used to determine which cities are safer than others. It is represented in our equations as C.

Beat: For all intents and purposes in this paper a beat is a region within a city within which crime takes place.

Victimization: Victimization occurs when one is profoundly affected by the consequences of a crime, even if one is not explicitly targeted. An example of this is a close relative of a murder victim, whose lifestyle and well-being is drastically changed as a result of the crime.

2.2 Assumptions and Justifications

Below we list some assumptions that were necessary to make in order to expediently and effectively construct our model, along with justifications for each assumption:

Assumption 1: Beats that are not referred to in the given spreadsheet do not exist.
Justification: The mean number of crimes in any given beat was calculated to be 40, and the lowest number of crimes for any given beat was 5. As a result, if beats did exist in the city where crime did not take place, there would likely be very few.

Assumption 2: The data given is for both the main city and the surrounding area described in the problem. All 8.8 million people are included.
Justification: There was not enough information to determine which crimes took place in or out of the city, and as such we assumed the crime report included both crimes within the city and in the surrounding area.

Assumption 3: The percentage of crimes taking place in residential areas contributes to the danger/safety rating of a city, but the percentage of crimes which are related to domestic/personal affairs does not.

Justification: If two cities have the same amount and severity of crime, the city in which less crime takes place in the home should be considered safer. However, the risk of victimization of a domestic crime is independent of one's location.

Assumption 4: One single person will not be the victim of multiple crimes during the two week period.
Justification: My City has a population of 8 million people, and according to our model, about 2,000 are victimized by crime in one day. As such, there is a 0.024% chance of being victimized any given day. Multiplying this number by 14 for the two week period and squaring the result yields .001%, which is the likelihood of being affected twice by crime in the same two week period. We considered this likelihood negligible because of its low value and because of the complications that would arise from needing to account for these people.

Assumption 5: A crime in somebody's home is worse than a crime in a public place.
Justification: If two cities have the same abundance and severity of crime, then the city in which more crime takes place outside the home is safer, because it is possible to make the conscious decision to stay at home and lower the likelihood of victimization.

Assumption 6: If a crime is listed with a "no" under the arrest column, the perpetrator escaped.
Justification: Had the perpetrator not escaped, an arrest would have been made.

2.3 Variables

The following variables will be used within our model:

n = number of instances of a particular crime
n_h = number of instances of a particular crime occurring in a home or private residence
n_a = number of instances of a particular crime in which the perpetrator was arrested
n_{total}= total number of crimes taking place in a city over a given period
S = the severity of a particular crime
P = the population of the city
P_v = the number of people victimized by a particular crime
d = the number of days over which the surveyed information was taken
C = Crime Rating
C_{ind} = Crime Rating for each individual type of crime
$C_{with\ deviation}$ = Crime Rating with variation between beats factored in

3 - Model Description

3.1 Utilizing Abundance of Crime Incidents and City Population

Before we began to analyze the massive amounts of information given in the spreadsheet for 'My City,' we figured it was important to first consider what factors would play into our determination of which cities are safer than others. The simplest of these and probably the most obvious is the number of crimes that takes place every day within the city. The easiest way to begin, as such, was to create a 'crime rating' value based on this number:

$$C = \frac{n_{total}}{d}$$

Where C is our hypothetical 'crime rating' value, n_{total} is the total number of crimes in the data given, and d is the number of days surveyed.

This equation is a good place to begin, because a city with more individual crimes every day will have a higher crime rating than a city with fewer crimes every day. Next, it is important to consider the size of the city in order to accurately compare cities of different sizes. According to the previous model, given two cities with the same number of crimes per day, the crime rating of both would be the same, even if one city was far larger. A smaller city with the same number of crimes as a larger city should instead have a larger crime rating, because crime affects a larger percentage of the city's population. Adjusting our model yields the following:

$$C = \frac{n_{crimes}}{dP}$$

Where P is the population of the city surveyed.

3.2 Incorporating Chance of Victimization

After this we had a good scale for judging the overall crime rating of a city, independent of its size. We then decided to utilize the percent likelihood of becoming a victim of a crime as a factor in determining our crime rating. To do this requires knowledge of how many people are victimized by crime as a fraction of the total population:

$$C = \frac{n_{crimes}}{d} \times \frac{E}{P}$$

Where E is the number of people victimized. To find E, it is necessary to find the number of people victimized by each crime that has taken place. E is thus the total result of the sum of each individual crime multiplied by the number of people it affected. This step is based on the idea that multiple people can be victimized by crime even if they are not explicitly targeted. As such, it is incorrect to assume that the number of people victimized by crime explicitly equals the number of crimes committed. To account for this, we assigned a hypothetical average number of people who are victimized to each type of crime:

Type of Crime	Number of People Victimized
Theft	2
Criminal Damage	2
Criminal Trespassing	3
Battery	4
Assault	4
Narcotics	1
Prostitution	2
Robbery	1
Weapons Violation	2
Motor Vehicle Theft	4
Homicide	6
Public Peace Violation	3
Interference with a Public Officer	4
Gambling	1
Criminal Sex Assult	3
Burglary	4
Sex Offense	3
Offense Involving Children	4
Arson	5

Deceptive Practice	1
Kidnapping	6
Liquor Law Violation	3
Other Offense	2
Stalking	3
Other Narcotics Violation	1
Concealed Carry Licence Violation	1
Intimidation	2

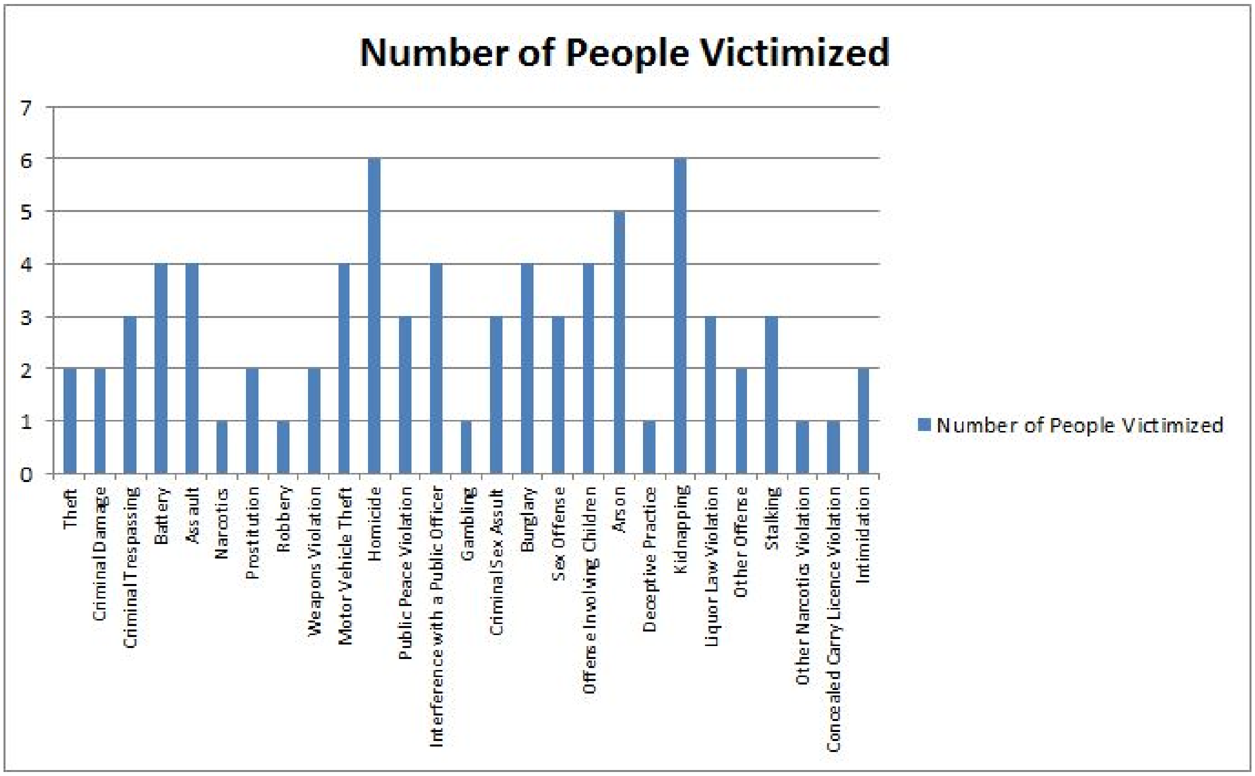

Figure 1a & b: This is the relation between each crime and our estimate on how many people are victimized.

These values can be modified to yield different results.

A mathematical expression for E can be written as such:

$$E = \sum nP_v$$

Where n is the number of instances of a specific type of crime and P_v is the corresponding number of people victimized. The summation will thus add together the products of n and P_v for each type of crime. Since there are 27 different types of crimes in the report from My City, writing a full expression would be impractical, and as such it is easier to use a program to perform this calculation.

3.3 Incorporating Severity of Crimes

After we had accounted for the chance of victimization, we decided that it would also be ideal to include the severity of different crimes in our calculation. If two cities have the same amount of crime and have the same chance of victimization, the city with more severe crimes should be considered less safe. To account for this, we performed research to determine how real-life city governments determine the severity of certain types of crime. Using this research we developed a 1 to 10 scale of severity with which each type of crime could be ranked. This was incorporated into our equation in a manner similar to the way in which we incorporated the number of people affected by a type of crime. It is also important to note that we opted to solve for the crime rating based on individual types of crime rather than the crime rating as a whole:

$$C_{ind} = \frac{(nSP_v)}{dP}$$

Where S is the severity of the crime, and C_{ind} is the crime rating for each individual type of crime; we will later add each individual crime rating to create one overall crime rating. Therefore the use of E is not necessary for calculating the overall crime rating of the city, but we later discovered that E has practical applications in determining the likelihood of being victimized by a crime.

The following table shows the severity we assigned to each type of crime:

Type of Crime	Severity
Theft	1
Criminal Damage	1
Criminal Trespassing	2
Battery	4
Assault	6

Narcotics	2
Prostitution	3
Robbery	7
Weapons Violation	3
Motor Vehicle Theft	4
Homicide	10
Public Peace Violation	2
Interference with a Public Officer	3
Gambling	1
Criminal Sex Assult	9
Burglary	5
Sex Offense	8
Offense Involving Children	8
Arson	6
Deceptive Practice	2
Kidnapping	10
Liquor Law Violation	2
Other Offense	4
Stalking	6
Other Narcotics Violation	2
Concealed Carry Licence Violation	3
Intimidation	5

Figure 2a: This is our relation of each crime to its severity, 10 being the most and 1 being the least severe.

Graph of Above Table:

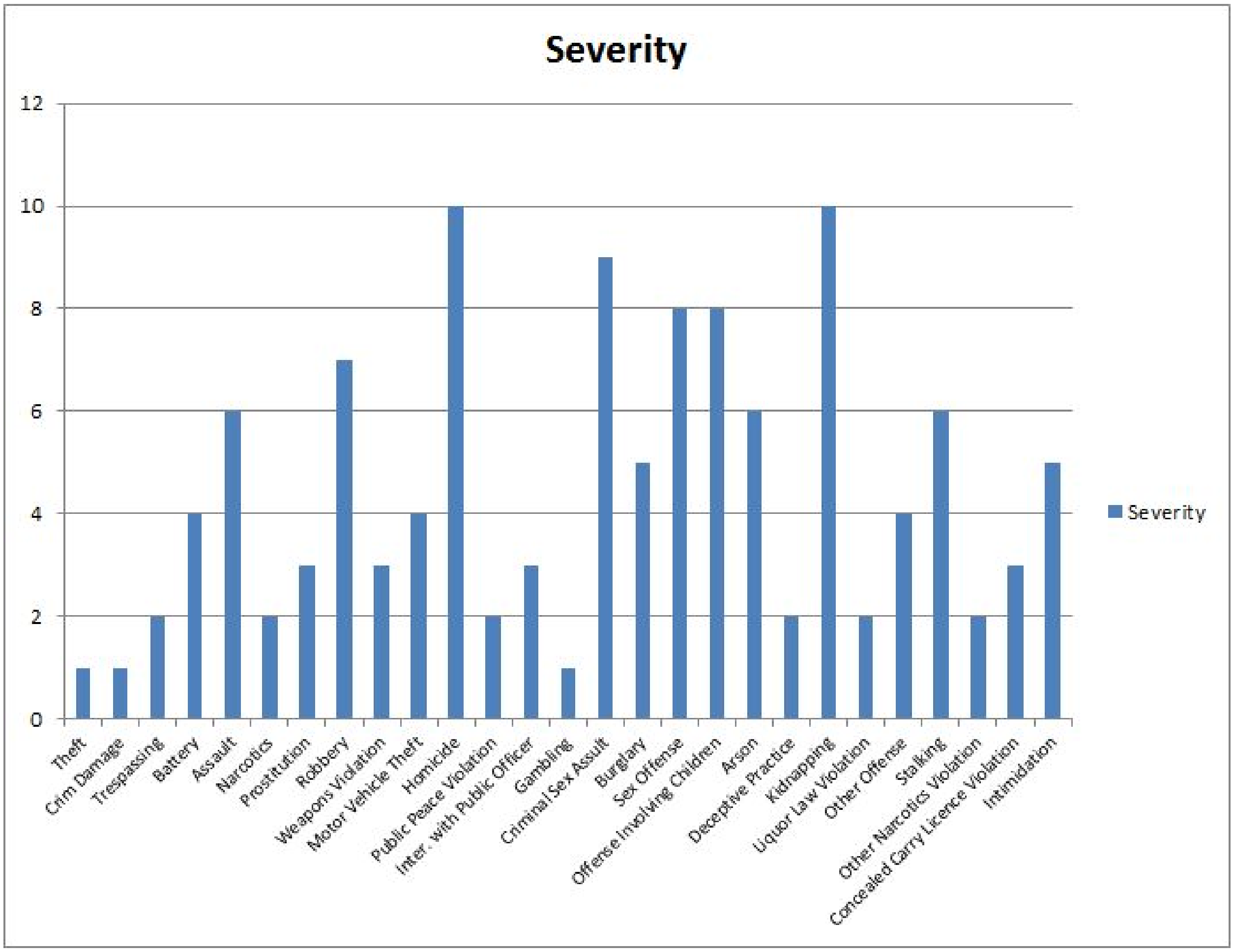

Figure 2b: This is our relation of each crime to its severity, 10 being the most and 1 being the least severe.

The severity values assigned are based upon extensive research of ranking systems utilized by real life city and local governments in determining the severity of different types of crime. Like the values we assigned in Figure 1, these values can be modified to yield different results based on how severe certain crimes must be considered in different situations.

3.4 Incorporating Police Efficiency and Residential Safety

At this point, our model was already very comprehensive, incorporating not only the number of crimes taking place in a given city, but also the severity of each crime, the number of people affected, and how large the city is.

In order to further improve our model's representation of overall safety in our city, we added a factor representative of how many of the perpetrators were arrested. This factor will cause the crime rating to increase if a city's police force is not as effective at apprehending criminals. This is based on the idea that given two cities with equal crime abundance and severity, the city where more perpetrators are arrested should be considered safer. The following equation will give us a number between one and two; if all criminals are arrested after committing a crime, the value will be .75, decreasing the original crime rating. If no criminals are arrested, the value will be 1.25, causing the crime rate to increase.

$$A = 0.75 + \frac{n - n_a}{2n}$$

Where n_a is the number of people arrested for each crime. Hence:

$$C_{ind} = \frac{(A \times n \times S \times P_v)}{dP}$$

To make our model even more accurate we created a new factor that would incorporate the percentage of crimes that took place in residential areas. This is based on the idea that given two cities with equal abundance and severity of crime, the city in which less crime takes place within or near the home should be considered safer. In a similar manner to how *A* was incorporated, we created a new factor *H* whose value will also lie between .75 and 1.25. If no crimes take place in the home, then the value of *H* will be .75, lessening the crime rating. If all of the crimes in a city take place in the home, the value of *H* will be 1.25, increasing the crime rate:

$$H = 0.75 + \frac{n_h}{2n}$$

Where n_h is the number of each crime that took place in somebody's home. Hence:

$$C_{ind} = \frac{(H \times A \times n \times S \times P_v)}{dP}$$

When we evaluate the equation above for each type of crime, and summate them, we find the crime rating for our city:

$$C = \sum C_{ind} = \sum \frac{(H \times A \times n \times S \times P_v)}{dP}$$

When writing the full equation we find:

$$C = \sum \frac{(0.75+\frac{n_h}{2n})\times(0.75+\frac{n-na}{2n})\times n\times S\times P_v)}{dP}$$

For quick reference,

n = number of instances of a particular crime
n_h = number of instances of a particular crime occurring in a home or private residence
n_a = number of instances of a particular crime in which the perpetrator was arrested
n_{total}= total number of crimes taking place in a city over a given period
S = the severity of a particular crime
P = the population of the city
P_v = the number of people victimized by a particular crime
d = the number of days over which the surveyed information was taken
C = Crime Rating
C_{ind} = Crime Rating for each individual type of crime
$C_{with\ deviation}$ = Crime Rating with variation between beats factored in

Since our current model gives us answers all less than 0.0001 we introduced a scaling factor of 10,000 to our equation, simply for ease of understanding for the viewer:

$$C = 10^4 \times \sum \frac{(0.75+\frac{n_h}{2n})\times(0.75+\frac{n-na}{2n})\times n\times S\times P_v)}{dP}$$

This constant of 10^4 causes the value of C to generally lie between 0 and 50 for most cities, as will be explained in our analysis.

3.5 Extension: Incorporating Uniformity of City

When looking at the beat numbers, we realized that it was not possible to draw conclusions about the locations of different beats relative to one another. Initially this impeded our ability to utilize the relative dispersal of crime throughout the city within our model- however, we later realized that using individual statistics for each beat could allow us to draw conclusions about the distribution of criminal activity. To do this, we decided to utilize the standard deviation of the sums of the products of severity, the number of instances of crimes, and number of people affected for each crime in a beat. A larger standard deviation would indicate that, on average, the abundance and severity of crime differed by greater amounts from area to area within a city. A smaller standard deviation, on the other hand, would indicate that the majority of areas in the city had the same amount of crime and severity.

$$x_b = \sum \frac{nSP_v}{d}$$

The above equation expresses a simple way to represent the impact of crime on a certain beat. Ideally, it would be best to observe the standard deviation of crime rate *C* for every beat, but the spreadsheet does not indicate the population of each beat. We also could potentially include the aforementioned factors *A* and *H* for each beat to have a more accurate representation of the impact of crime in each beat, but we ruled that the simplified version of our model without these factors would be adequate for determining the standard deviation of the impact of crime in each area of the city.

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \ldots}{n_b}}$$

The above equation takes the standard deviation for all of the beats in a given city, where x_1, x_2, x_3, etc. represent the value for x_b evaluated for each beat in the city. The symbol n_b represents the number of beats in the city.

We discussed the degree to which the standard deviation should affect the overall crime rating, and decided that when factoring the standard deviation into our formula, we should divide it by the mean of the crime ratings. This is called the coefficient of variation. This is very likely to be a value less than 1, and will still remain relatively small even in extreme circumstances with high variability. It is also unitless, which allows it more versatility in being incorporated into our model.

We also discussed whether or not the uniformity of a city would contribute to its safety or danger. Ultimately, we concluded that for safer cities, uniformity should contribute to the city's overall safety, and thus lower the crime rating. This is because cities trying to optimize their safety should avoid harboring individual pockets of crime and relegating it to certain areas. However, for more dangerous cities, we concluded, uniformity should contribute to a city's crime rating. This is because prospective visitors or residents of relatively unsafe cities would value the presence of areas with higher safety. In order to model this, we used the average of the simplified crime ratings for Detroit and Minneapolis, two cities that represent typical perceptions of unsafe and safe cities, respectively. If the crime rating, when calculated without factoring in standard deviation, yields a value less than 25.30, then the standard deviation factor should increase the crime rating. If a value less than 25.30 results from calculating crime rating without factoring in standard deviation, then the standard deviation factor should decrease the crime rating. Keeping this in mind, and adjusting constants in order to keep the uniformity factor from drastically changing the value, the following rules were devised:

$$C_{\textit{with deviation}} = C\left(\frac{1+\frac{\sigma}{mean}}{4}+1\right) \text{ if } C < 25.30$$

$$C_{\textit{with deviation}} = C/\left(\frac{1+\frac{\sigma}{mean}}{4}+1\right) \text{ if } C > 25.30$$

This factor requires a detailed knowledge of the number and type of crimes taking place in each beat within a city, and thus can rarely be applied without a comprehensive report from the city.

3.6 Final Model and Application Guidelines

The final, completed model, which assumes a comprehensive knowledge akin to that provided to us within the spreadsheet for My City, can be written as follows:

$$C = 10^4 \times \sum \frac{(0.75+\frac{n_h}{2n})\times(0.75+\frac{n-na}{2n})\times n\times S\times P_v)}{dP}$$

To factor in the uniformity of crime distribution throughout the city, utilize the following rules as dictated in Section 3.5:

$$C_{\textit{with deviation}} = C\left(\frac{1+\frac{\sigma}{mean}}{4}+1\right) \text{ if } C < 25.30$$

$$C_{\textit{with deviation}} = C/\left(\frac{1+\frac{\sigma}{mean}}{4}+1\right) \text{ if } C > 25.30$$

It is important to keep in mind that *in order* to compare cities using our model, the same information must be utilized in the model for both calculations. As such, to compare crime ratings between two cities where n_a and/or n_h are unknown requires that one utilize the abbreviated version of the model, shown here:

$$C = 10^4 \times \sum \frac{nSP_v}{dP}$$

Because finding detailed information about the arrest rate for different types of crimes and the frequency of crimes occurring in residential areas is exceedingly difficult without a comprehensive report, using the above model is far more convenient for comparing different real-life cities.

Additionally, when using the above model, it is important to note that accurate comparisons between cities can only be made if the *same* types of crimes are used to evaluate for C. For example, if data are not present for prostitution violations in one city, but are present in another, then summing $\frac{nSP_v}{dP}$ for prostitution violations when calculating C for both would cause the second city to appear more dangerous than the first simply because the second city had more data on reported incidents. As such, when comparing two cities' C values, it is important to only summate the terms for each crime that are represented by data for both cities.

3.7 Strengths and Weaknesses of our Model

Strengths:
-Our model accounts for a large number of factors that could potentially contribute to the impact of crime on a city, including the number of incidents, the severity of the crimes, the size of the city, the number of perpetrators that escape arrest, the proximity of crimes to residential areas, and even the distribution of crime throughout the city
-Our model is flexible- if certain information is not available one can simply remove factors and create an abbreviated version of our model that still produces viable crime ratings
-One can effectively compare different cities even if the amount of data available is inconsistent for the two cities
-The model, by default, assigns reasonable and researched weights to the severity of crimes and the importance of arrests that reasonably scales real-life cities based on their data
-Our model was devised around the performance of key cities such as Minneapolis, Detroit, and Los Angeles in order to allow a reasonable placement of other cities and to provide an accurate scale for our results

Weaknesses:
-The model requires at a minimum the knowledge of a city's population, the number of crimes taking place over a certain period of time, and what different types of crime took place and how many of each took place
-Our model assumes a fixed weight and number of persons victimized per crime; to be incredibly accurate, it would be possible to assign a more accurate weight and number of persons to each individual infraction, but this requires careful analysis of very large amounts of data, which could be possible with more time

-The weights assigned to factors A, H, and the standard deviation, as well as the weights of each type of crime were assigned arbitrarily based on extensive research and calculations, and were designed to make reasonable changes in the C value based on our interpretations. Ultimately, the severity of crimes relative to one another and the importance of other factors are entirely subjective and thus cannot be reliably quantified by any standard
-To gather many of the values necessary for computing C, computer programs are necessary to avoid painstakingly large amounts of manual computation and data entry

4 - Analysis

4.1 Spreadsheet Analysis

Our model was designed to rate a city based on how safe it is, given crime reports over a two week period. By taking all of the relevant information for a given city, we could then calculate its crime rating. To gather all of the necessary information and to perform the necessary summation, we decided to use Excel. We believe it is best to utilize multiple models to affirm our results, so in addition to our Excel analysis, we also created a computer program which would parse the individual rows in the original spreadsheet and thus scan through all of the reported crimes over the entire two-week period.

In our Excel analysis, we assigned rows to each type of crime, since our model is essentially based on a sum of calculations for each individual classification of crime. Within each row we used COUNTIF() functions to determine the number of crimes that took place by having Excel search through the original spreadsheet to find all of the crimes that were labeled a certain type and count the number of each. We used similar Excel functions to determine how many arrests were made for each separate type of crime and how many of them took place in locations that would imply they were in residential areas. Once we had these values, we could easily perform the calculation for C in our model, though this wouldn't include the standard deviation factor.

Excel was similarly used to calculate the crime rating using the abbreviated model for other cities using data from online sources about the number and types of crimes.

4.2 Programming Analysis

To calculate the standard deviation we needed to iterate through all of the data and for each beat create a data entry which represented the sum of the severity of each crime times the number of times that crime happened times the number of people that crime affected. These values were taken from the tables that we attached earlier. To store the data that represents the net crime of a beat we used an array where the dimension represents the beat. To add the excel data to our IDE we exported the excel data in a tab delimited format. We looped through each line of the text document such that at the location in the array where the beats where the value equaled the sum of all of the crime data that occurred in that beat. From that point we calculated the standard deviation.

To make sure that the program was effective we used a small subsection of the data set that was provided to test the model for standard deviation/mean. Because the data set was quite small we were able to manually affirm that the program worked.

Below is our flow chart for the program we wrote:

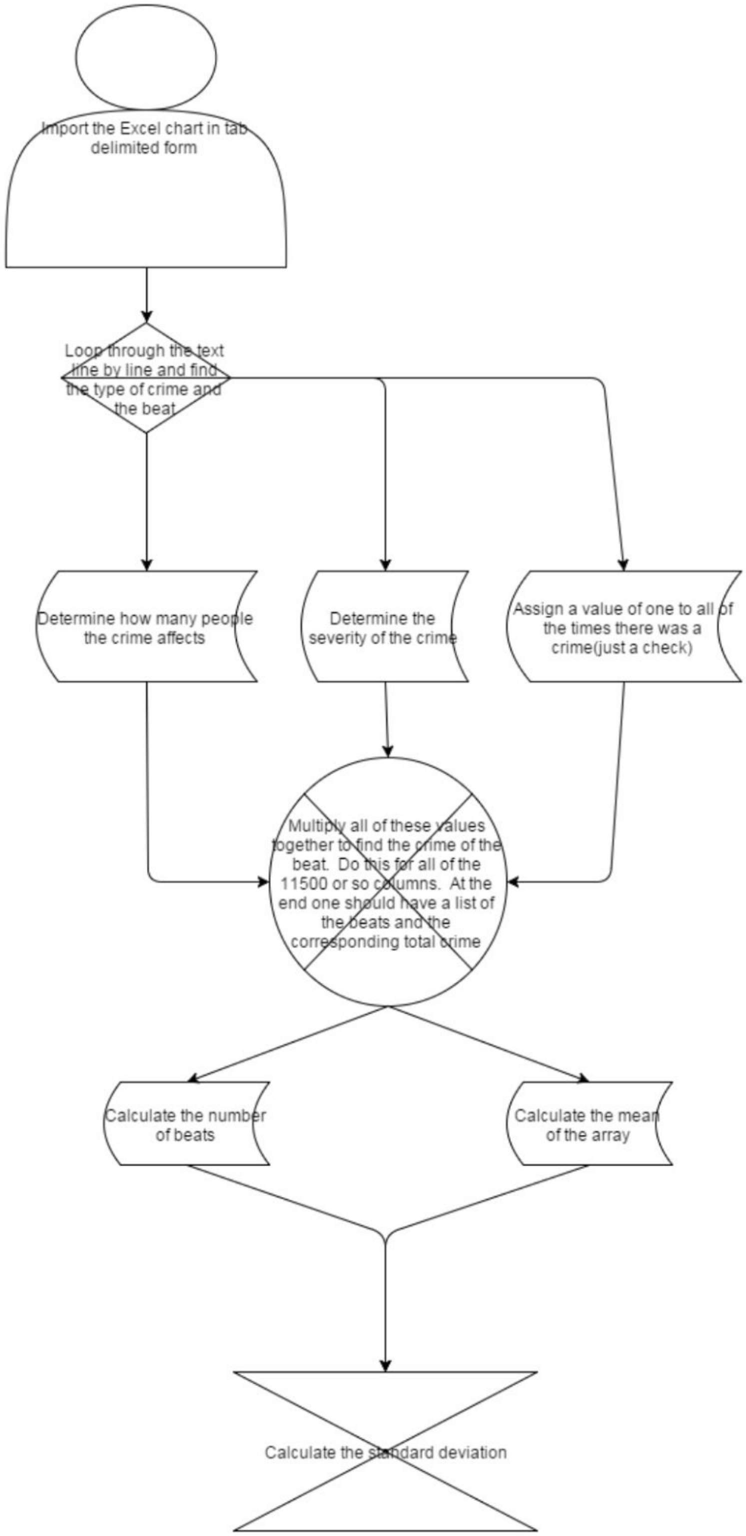

Figure 3: A flow-chart outlining the central strategies we used in our program to calculate the standard deviation.

To test the efficacy of our model we used two real world cities to provide a basis of comparison for the results we got for My City. In comparing My City with other real world cities the essential principle is using the same amount of data from both cities. Often times it is hard to find data concerning a city that is of the same resolution as the data that we were provided. Thus using only data that we have a point of comparison to makes the most sense, in essence this ensures that the d value is the same for the both cities.

4.3 Sensitivity Analysis:

Varying the severity of the crimes:
By doubling the severity of each crime, we also double the Crime Rating. They are directly proportional.
Varying the population of the city:
When all else is constant the higher the population; the lower the Crime Rating.
Varying the efficacy of the police:
When the police department is one hundred percent efficient the result of the other calculations is multiplied by .75; when it is zero percent efficient the result is multiplied by 1.25.
Varying the number of crimes which take place in a residential area:
When the rate of crimes in homes is one hundred percent the result of the other calculations is multiplied by 1.25; when it is zero percent the result is multiplied by .75.
Varying the amount of variation between beats:
When there is no variation the result is multiplied by 0.75; when the standard error is double the mean the result is multiplied by 1.25.

5 - Results and Conclusion

By using a simplified version of the our model and only considering the types of crimes that were documented by all cities:

$$C = 10^4 \times \sum \frac{(n \times S \times P_v)}{dP}$$

We concluded that ‘My City’ had an overall Crime Rating of 4.4. To put this in context, Minneapolis, the safest city, received a score of 2.6, and Detroit, the most dangerous city, received a score of 37.7.

By using the equation for the percent chance of being victimized we were able to conclude that on a given day there is a 0.023% chance of being victimized, and in a given year there is an 8.83% chance of being victimized.

The following table on factors in the types of crime that all of them shared

	My City (Greater metro area)	Minneapolis (greater metro area)	Detroit (city)	Seattle (Greater metro area)	Los Angeles (county)
Total Grief Per Day	3910	1021	2352	1005	3119
Crime Rating	4.4	2.6	37.7	2.8	3.1

Figure 4: The relation between major cities in the US and My City and their crime ratings.

Given information about the arrest rate, location, and distribution of crime throughout each city, we would be able to add resolution to our results by using our full model- however, since such statistics were only available for My City, the following comparisons were made by projecting the crime ratings for Detroit and Minneapolis linearly with respect to the growth of My City’s crime rating. The following table shows how the crime rating for My City changes when more and more of these factors are considered, and how it would compare with that of Minneapolis’ and Detroits’ crime ratings:

	Not factoring in whether or not the crime was in a home. Not factoring in whether or not the criminal escaped.	Not factoring in whether or not the crime was in a home.	Factoring in both homes and escapes.	Factoring in variance in crime per beat..
Crime Rating For My City	8.606	9.763	13.976	17.47
Projected Rating for Minneapolis	5.085	5.768	8.257	10.321
Projected Rating of Detroit	73.737	83.650	119.747	149.683

Figure 5b: A more in depth relationship between My City, Minneapolis, and Detroit.

Projections are purely hypothetical, but provide some context for understanding our results. Here is a graph of our function and projections.

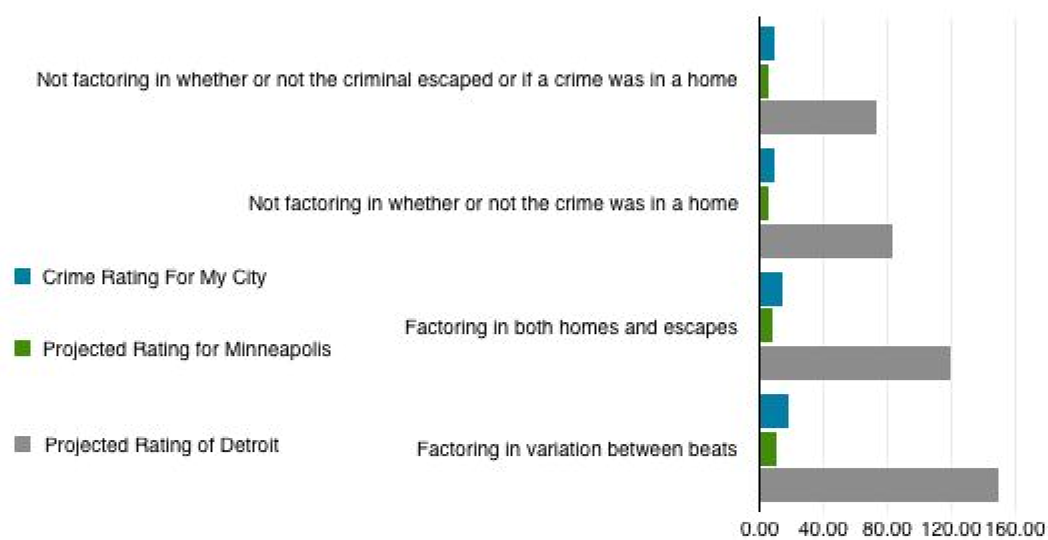

Figure 5b: This is the graph for the table above in Fig. 5.

Our model shows that the city provided is a more dangerous place to live than Minneapolis by a small factor; and is a safer place to live that Detroit by a large factor. This trend continues when we use our full model and the complementary projections. Ultimately, our model shows that My City is safe by comparison to downtown Detroit, and thus is not in danger of being historically remembered for its crime, but steps must be taken in order for the overall safety of the city to match that of most other major cities, such as Minneapolis, Seattle, and Los Angeles.

Earlier in the process we calculated that the overall percent chance that a person will be affected by *any* crime to be 8.8%. This number closely resembles the 12% provided by the Florida Correctional Department, therefore we can conclude that our model is, yet again, accurate.

6 - Appendix

6.1 Explanation of Spreadsheet

Explanation of Our Excel Document: (Organized by Column)

1. In order to show and calculate what we needed to, we opted to create an excel document. In this document we first laid out each of the types of crime in the far left column.
2. In the second column we put our severity factor.
3. In the third column we put the estimated number of people affected by each instance of the crime.
4. In the fourth column we put the number of each crime that took place.
5. Next we divided the number of each crime by the number of days surveyed, fourteen, and stored each in the fifth column.
6. In the sixth column we took the data we gathered from the original spreadsheet regarding the number of crimes taking place in a private residence or home, and typed it into the cell corresponding to the type of crime.
7. In the seventh column we divided the number that took place at home by the number total for each crime, giving us a percentage of crimes that were at home.
8. In the eighth column we typed data from the original spreadsheet that pertained to the number of arrests made for each type of crime, and typed it into the corresponding cell on the new excel spreadsheet.
9. We then subtracted this value from the number of each instance of a crime that took place, divided it by the number of each instance of a crime that took place, and added one to the result. This gave us a number between one and two, which we would later multiply our overall crime rating by. This would double the crime rating if no arrests were made, but do nothing if every criminal was arrested.
10. Was going to be number domestic, but we decided against going down that route.
11. Was going to be percent domestic, but we decided against going down that route.
12. This column we simply put the total population in the highest cell so that we could change the value for population easily later on.
13. In the thirteenth column we multiplied the number of each type of crime that took place in a day by its corresponding 'people affected' from column two to give us the total number of people affected by this crime per day.
14. In the fourteenth column we divided the number of people affected by the total population and multiplied by one hundred to find the percent chance of being affected by the corresponding type of crime in a day.

15. In the fifteenth column we multiplied the percent chance of being affected by a type of crime in a day by three hundred sixty five and a quarter to find the percent chance that someone would be affected by the corresponding type of crime in a year.
16. In the sixteenth column we added up all of the percent chances that a crime would affect somebody to create a percent chance of any crime affecting a person over the course of a year.
17. In the seventeenth column we added up all of the 'People affected per day', from column thirteen, to give us the total people affected per day by any crime.
18. In the eighteenth column we divided the total people affected per day and divided it by the total population, then we multiplied this number by one hundred to give us a percent chance of being affected by any crime per day.
19. We multiplied the result from the eighteenth column by three hundred sixty five and a quarter, which gave us the same results as in column sixteen. Because… redundancy.
20. In column twenty we multiplied the number of each type of crime per day by column seven by our severity value by our people affected value to give us a 'Grief' rating per day for each type of crime.
21. Columns 21 - 27 are simply the same as the previous columns,
22. But only factoring in the types of crime 'My City' shared with Detroit.
23. Doing this allowed us to accurately and fairly compare 'My City'
24. To the City of Detroit.
25. The numbers in bold tells the 'Crime Rating' for each set of circumstances illustrated
26. in the top cells of each column.
27. It is clear that the number increases drastically when we include whether or not the crime was at a private residence and whether or not the perpetrator was arrested.
28. In column twenty eight we took the 'Grief' rating for each type of crime, divided it by the total population, and multiplied the result by the result in column nine to give us a 'Crime Rating' for each type of crime.
29. In column twenty nine we summed up the 'Crime Ratings' for each type of crime to give us an overall 'Crime rating' for the city. We then multiplied by 10000 to give us a better and more comprehensive number, which is our overall 'Crime Rating' for the city.

Our Excel Spreadsheet: (In this file we named 'Crime Rating' 'Nastiness')

Type	Severity	People Effected	Number	Number Per Day	Number At Home	% At Home (<1)	Number Arrested	% Got Away (<1)	Number Domestic	% Domestic (<1)	Total Population	People Effected Per Day	% People Effected Per Day For Each Crime	% Effected Per Year	Total Chance Per Year
Theft	1	2	2658	189.857142857143	619	0.23288186606471	289	1.19563581640331		0	**8800000**	379.714285714286	0.004314935064935	1.57603003246753	8.83567025162338
Criminal Damage	1	2	1093	78.0714285714286	430	0.393412625800549	86	1.21065873741995		0	Niceness Number	156.142857142857	0.001774350649350	0.648081574675325	
Criminal Trespassing	2	3	310	22.1428571428571	117	0.37741935483871	227	0.883870967741936		0	**10000**	66.4285714285714	0.000754870129870	0.275716314935065	
Battery	4	4	2040	145.714285714286	1003	0.491666666666667	496	1.12843137254902		0		582.857142857143	0.006623376623376	2.41918831168831	
Assault	6	4	712	50.8571428571429	276	0.387640449438202	187	1.1186797752809		0		203.428571428571	0.00231168831168	0.844344155844156	
Narcotics	2	1	1179	84.2142857142857	147	0.124681933842239	1178	0.750424088210348		0		84.2142857142857	0.000956980519480	0.34953713474026	
Prostitution	3	2	92	6.57142857142857	0	0	92	0.75		0		13.1428571428571	0.000149350649350	0.054550324675324	
Robbery	7	4	403	28.7857142857143	29	0.071960297766749	37	1.20409429280397		0		115.142857142857	0.001308441558441	0.477908279220779	
Weapons Violation	3	2	142	10.1428571428571	31	0.21830985915493	116	0.841549295774648		0		20.2857142857143	0.000230519480519	0.084197240259740	
GTA	4	4	382	27.2857142857143	32	0.083769633507853	25	1.21727748691099		0		109.142857142857	0.001240259740259	0.45300487012987	
Homicide	10	6	20	1.42857142857143	0	0	7	1.075		0		8.57142857142857	0.000097402597402	0.035576298701298	
Public Police Violation	2	3	119	8.5	14	0.117647058823529	101	0.82563025210084		0		25.5	0.000289772727272	0.105839488636364	
Interfering With Public Officer	3	4	65	4.64285714285714	5	0.076923076923076	65	0.75		0		18.5714285714286	0.000211038961038	0.077081980519480	
Gambling	1	1	24	1.71428571428571	2	0.083333333333333	24	0.75		0		1.71428571428571	0.000019480519480	0.007115259740259	
Criminal Sex Offence	9	3	43	3.07142857142857	27	0.627906976744186	3	1.21511627906977		0		9.21428571428572	0.000104707792207	0.038244521103896	
Burglary	5	4	566	40.4285714285714	458	0.809187279151943	42	1.21289752650177		0		161.714285714286	0.001837662337662	0.671206168831169	
Sex Offence	8	3	23	1.64285714285714	6	0.260869565217391	9	1.05434782608696		0		4.92857142857143	0.000056006493506	0.020456371753246	
Offence With Children	8	4	69	4.92857142857143	52	0.753623188405797	10	1.17753623188406		0		19.7142857142857	0.000224025974025	0.081825487012987	
Arson	6	5	24	1.71428571428571	3234	134.75	3	1.1875		0		8.57142857142857	0.000097402597402	0.035576298701298	
Deceptive Practise	2	1	509	36.3571428571429	219	0.430255402750491	49	1.20186640471513		0		36.3571428571429	0.000413149350649	0.150902800324675	
Kidnapping	10	6	11	0.785714285714286	3	0.272727272727273	1	1.20454545454545		0		4.71428571428571	0.000053571428571	0.019566964285714	
Liquor Law Violation	2	3	21	1.5	1	0.047619047619047	21	0.75		0		4.5	0.000051136363636	0.018677556818181	
Other Offence	4	2	643	45.9285714285714	335	0.52099533437014	148	1.13491446345257		0		91.8571428571429	0.001043831168831	0.381259334415584	
Stalking	6	3	7	0.5	4	0.571428571428571	3	1.03571428571429		0		1.5	0.000017045454545	0.006225852272727	
Narcotics (Other)	2	1	1	0.071428571428571	0	0	0	1.25		0		0.071428571428571	0.000000811688311	0.000296469155844	
Concealed Carry Violation	3	1	1	0.071428571428571	0	0	1	0.75		0		0.071428571428571	0.000000811688311	0.000296469155844	
Intimidation	2	2	5	0.357142857142857	2	0.4	1	1.15		0		0.714285714285714	0.000008116883116	0.002964691558441	

Total Chance Per Year	Toal People Effected Per Day For Any Crime	Total % Chance Per Day Of Any Crime	Total Chance Per Year	Grief Per Crime Per Day	Grief Per Crime Per Day, Considering Only Those With Detroit	Nastiness Only Considering Those We Share With Detroit	Grief Per Crime (No Home)	Nastiness Per Crime (No Home; No Escapes)	Total Nastiness (No Home; No Escapes)	Nastiness Per Crime (No Home)	Total Nastiness (No Home)	Nastiness Per Crime	Total Nastiness For Every Crime In The Entire City
8.83567025162338	2128.78571428571	0.024190746753246	8.83567025162338	329	3910.92857142857	0.000444423701298	379.714285714286	0.000043149350649	0.000860600649350	0.000051590909090	0.000976392045454	0.000044700475408	0.001397661543621
				147.821428571429		Nicer Number	156.142857142857	0.000017743506493	Nicer Number	0.000021481331168	Nicer Number	0.000020336511827	Nicer Number
				124.714285714286		**4.44423701298701**	132.857142857143	0.000015097402597	**8.60600649350649**	0.000013344155844	**9.76392045454546**	0.000012526288227	**13.9766154362122**
				2321.71428571429		Detroit (Calculated by hand)	2331.42857142857	0.000264935064935		0.000298961038961		0.000297715367965	
				1152		**37.7**	1220.57142857143	0.000138701298701		0.000155162337662		0.000146445352400	
				136.821428571429		Minneapolis (Calculated by hand)	168.428571428571	0.000019139610389		0.000014362824675		0.000011667510884	
				29.5714285714286		**2.6**	39.4285714285714	0.000004480519480		0.000003360389610		0.000002520292207	
				633.5			806	0.000091590909090		0.000110284090909		0.000086681106192	
				52.2857142857143			60.8571428571429	0.000006915584415		0.000005819805194		0.000005000114322	
				345.714285714286			436.571428571429	0.000049610389610		0.000060389610389		0.000047821615557	
				64.2857142857143			85.7142857142857	0.000009740259740		0.000010470779220		0.000007853084415	
				41.25			51	0.000005795454545		0.000004784902597		0.000003870141806	
				43.9285714285714			55.7142857142857	0.000006331168831		0.000004748376623		0.000003743912337	
				1.35714285714286			1.71428571428571	0.000000194805194		0.000000146103896		0.000000115665584	
				88.2321428571429			82.9285714285714	0.000009423701298		0.000011450892857		0.000012183217400	
				933.571428571429			808.571428571429	0.000091883116883		0.000111444805194		0.000128673463241	
				34.7142857142857			39.4285714285714	0.000004480519480		0.000004724025974		0.000004159196781	
				177.714285714286			157.714285714286	0.000017922077922		0.000021103896103		0.000023780114812	
				3503.57142857143			51.4285714285714	0.000005844155844		0.000006939935064		0.000472783076298	
				70.1785714285714			72.7142857142857	0.000008262987012		0.000009931006493		0.000009584689469	
				41.7857142857143			47.1428571428571	0.000005357142857		0.000006452922077		0.000005719635478	
				6.96428571428572			9	0.000001022727272		0.000000767045454		0.000000593547077	
				371.285714285714			367.428571428571	0.000041753246753		0.000047386363636		0.000047883809910	
				9.32142857142857			9	0.000001022727272		0.000001059253246		0.000001097083719	
				0.107142857142857			0.142857142857143	0.000000016233766		0.000000020292207		0.000000015219155	
				0.160714285714286			0.214285714285714	0.000000024350649		0.000000018262987		0.000000013697240	
				1.35714285714286			1.42857142857143	0.000000162337662		0.000000186688311		0.000000177353896	

Figure 6: The excel sheet we used for our calculations.

6.2 Equations

Archive of all equations used & derived:

$C = \frac{n_{total}}{d}$ $C = \frac{n_{crimes}}{dP}$ $C = \frac{n_{crimes}}{d} \times \frac{E}{P}$ $E = \sum nP_v$ $C_{ind} = \frac{(nSP_v)}{dP}$

$A = 0.75 + \frac{n-n_a}{2n}$ $C_{ind} = \frac{(A \times n \times S \times P_v)}{dP}$ $H = 0.75 + \frac{n_h}{2n}$ $C_{ind} = \frac{(H \times A \times n \times S \times P_v)}{dP}$

$C = \sum \frac{(0.75+\frac{n_h}{2n}) \times (0.75+\frac{n-n_a}{2n}) \times n \times S \times P_v)}{dP}$ $C = 10^4 \times \sum \frac{(0.75+\frac{n_h}{2n}) \times (0.75+\frac{n-n_a}{2n}) \times n \times S \times P_v)}{dP}$ $x_b = \sum \frac{nSP_v}{d}$

$$\sigma = \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+...(x_n-\bar{x})^2}{n_b}}$$

$$C_{with\ deviation} = C(\frac{1+\frac{\sigma}{mean}}{4}+1) \text{ if } C < 25.30$$

$$C_{with\ deviation} = C/(\frac{1+\frac{\sigma}{mean}}{4}+1) \text{ if } C > 25.30$$

$$C = 10^4 \times \sum \frac{nSP_v}{dP}$$

6.3 Code Explanation

Code to calculate standard deviation in the amount and severity of crime per beat:

```
import java.util.Scanner;
import java.io.File;
import java.io.IOException;
public class CountBeat
{
    //function to find the nth instance of a string
    public static int nth(String source, String pattern, int n) {

        int i = 0, pos = 0, tpos = 0;

        while (i < n) {

            pos = source.indexOf(pattern);
            if (pos > -1) {
                source = source.substring(pos+1);
                tpos += pos+1;
                i++;
            } else {
                return -1;
            }
        }

    return tpos - 1;
}
    public static void main(String[] args) throws IOException
    {
        File filename = new File("My_City_Crime_Data.txt");
        Scanner file = new Scanner(filename);
        //defining ints for the number
        int otherOffense = 0;
        int theft = 0;
        //condensing for the sake of space
        int concealedCarry = 0;
        int intimidation = 0;
        int n = 0;
        int beatNumber = 2535;
        //defining arrays
        int [] beatArray = new int[2536];
        int numberOfBeats = 0;
        int [] finalArray = new int[2536];
        boolean isArrest = false;
        int arrested [] = new int[28];
        for(int Count = 0; Count < (beatNumber+1); Count++) {
            beatArray[Count] = 0;
            finalArray[Count] = 0;
        }
        for(int Count = 0; Count < 28;Count++)
        {
            arrested[Count] = 0;
        }
```

```
            severity +=2;
            peopleAffected = 2;
        }
        //assigning values in an array where index is beat number and
the value is the total crime(for lack of a better word)
        beatArray[beatNumero] +=severity*peopleAffected;
        }
        file.close();
        //calculate mean
        int meanBeforeDivide = 0;
        for(int count = 0;count < beatNumber+1;count++)
        {
            if(beatArray[count]!=0)
            System.out.println(count + ": " + beatArray[count]);
            numberOfBeats++;
            meanBeforeDivide += beatArray[count];
        }
        double mean=meanBeforeDivide/numberOfBeats;
        //calculate average
        double standardDeviationBeforeAverage = 0;
        for (int count = 0;count<beatNumber+1;count++)
        {
            if (beatArray[count]!=0)
            {
                standardDeviationBeforeAverage =
standardDeviationBeforeAverage+((beatArray[count]-
mean))*(beatArray[count]-mean);
            }
        }
        double standardDeviationBeforeRoot =
standardDeviationBeforeAverage/numberOfBeats;
        double standardDeviation =
Math.sqrt(standardDeviationBeforeRoot);
        System.out.println(standardDeviation/14);
        System.out.println(mean/14);

}
}
```

6.4 Data from Original Spreadsheet

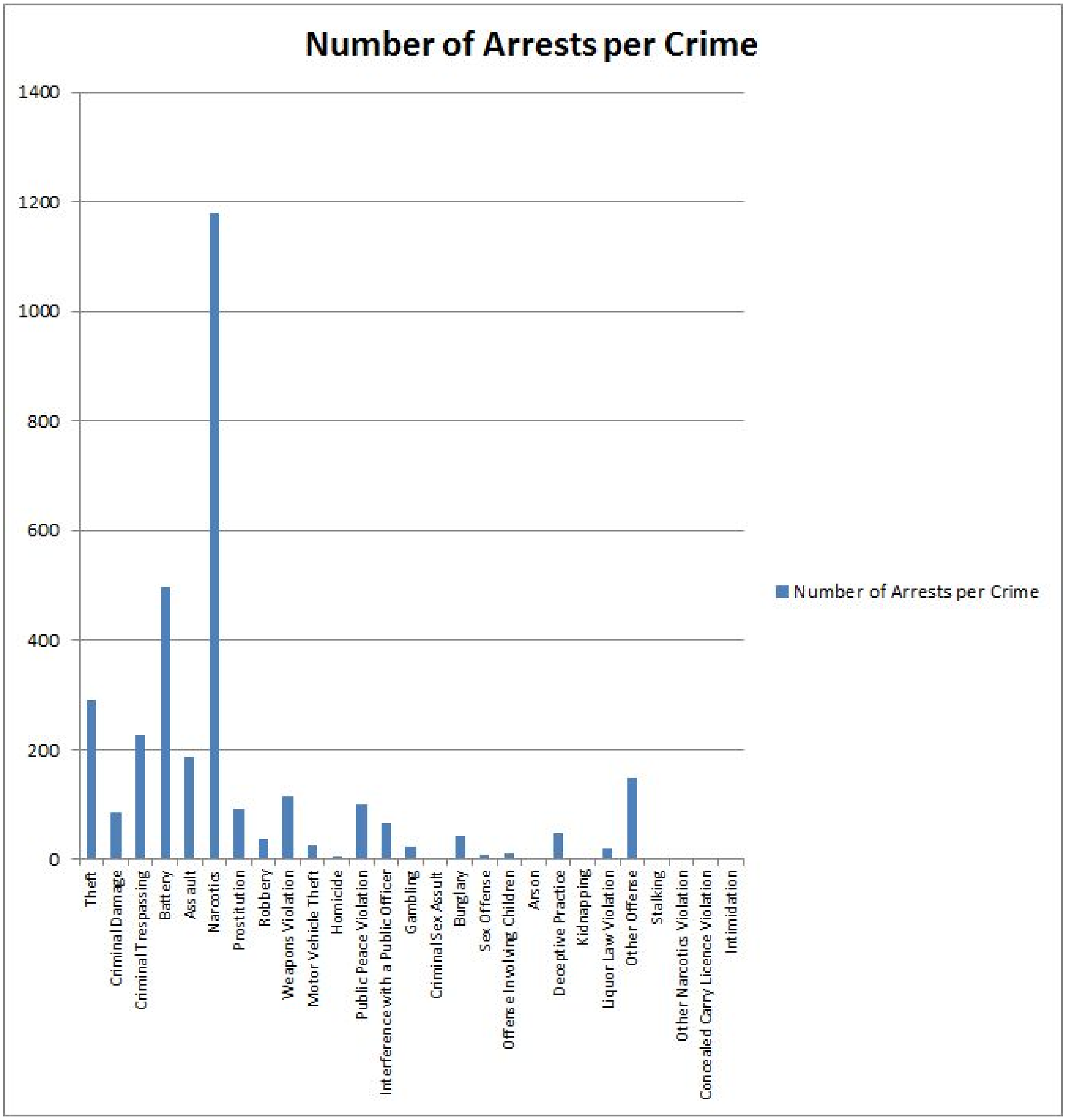

Figure 7: The amount of arrests per crime given in the excel spread.

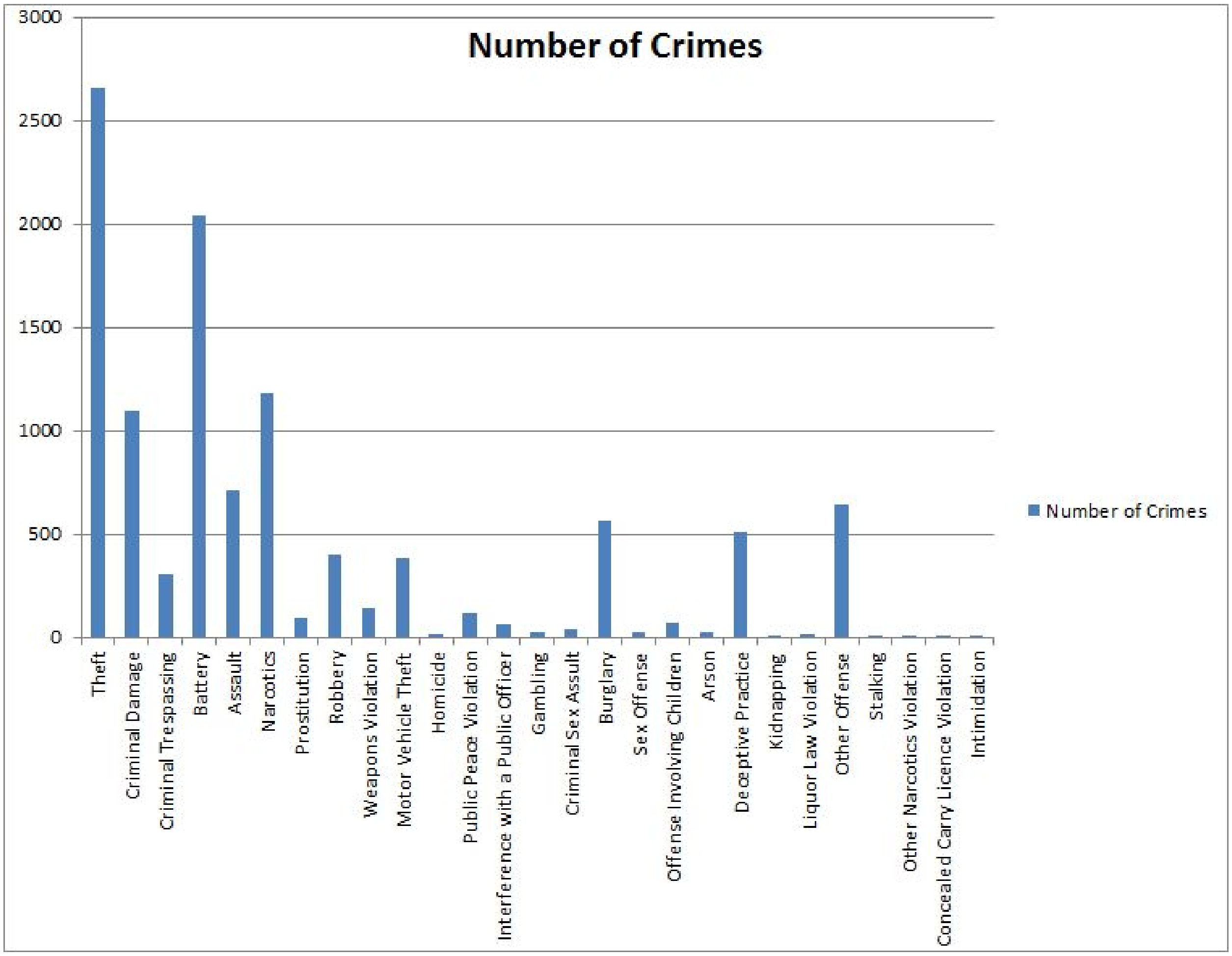

Figure 8: The amount of crimes per criminal act given in the excel spread.

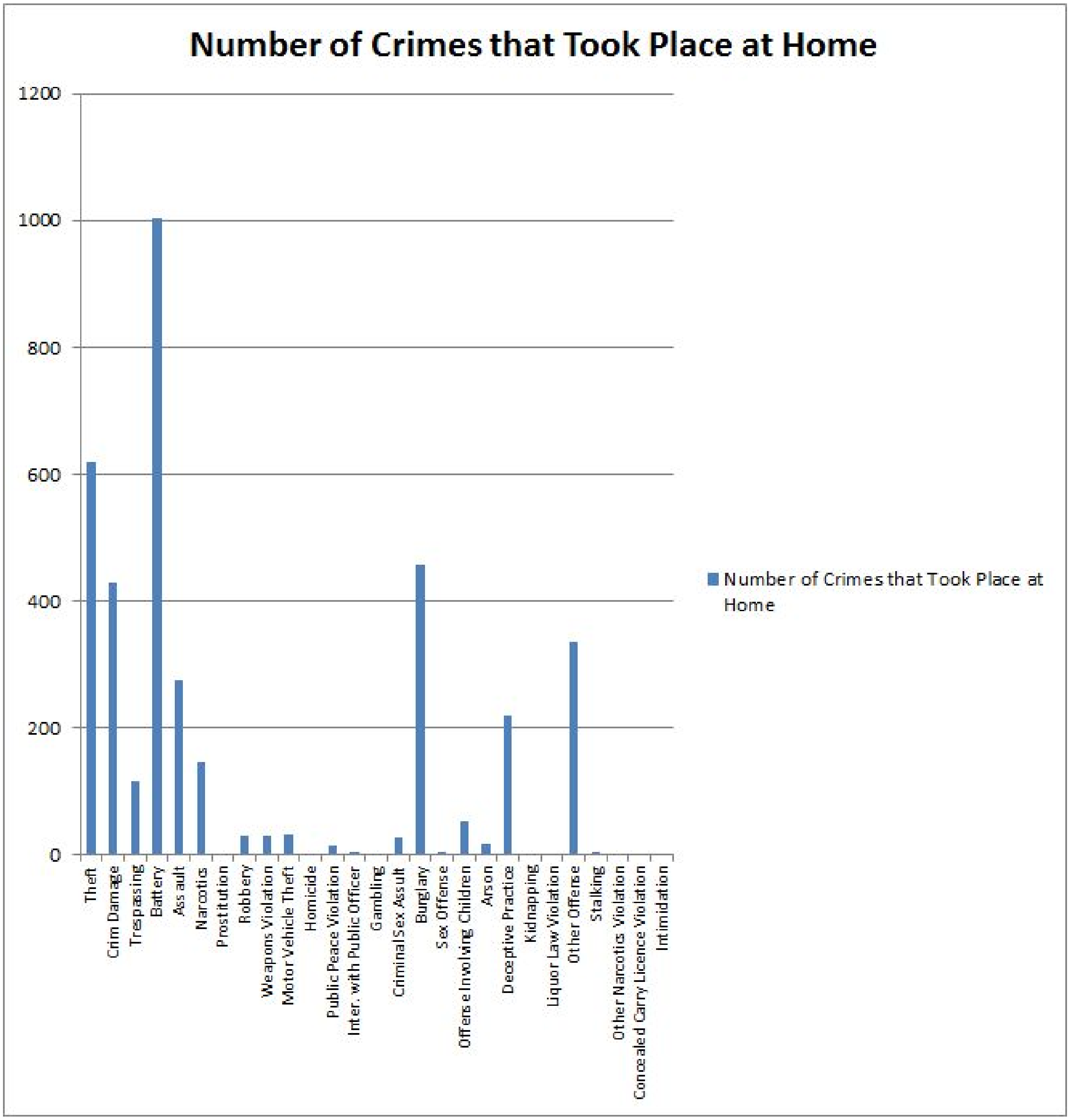

Figure 9: The amount of crimes that occurred in a Residence per type of crime given in the excel spread.

Works Cited

"Crime Dashboard." Seattle Police Department. *City of Seattle.* 2014. Web. 13 November 2015.

"Crime Rate in Detroit, Michigan (MI): murders, rapes, robberies, assaults, burglaries, thefts, auto thefts, arson, law enforcement employees, police officers, crime map." City-Data. *Advameg.* 2015. Web. 13 November 2015.

"Crime Rate in Los Angeles, California (Ca): murders, rapes, robberies, assaults, burglaries, thefts, auto thefts, arson, law enforcement employees, police officers, crime map." City-Data. *Advameg.* 2015. Web. 13 November 2015.

"Florida Criminal Punishment Code: Offense Severity Ranking Chart Section 921.0022, Florida Statutes." Florida's Criminal Punishment Code. *Florida Department of Corrections.* 2003-2004. Web. 13 November 2015.

"Full List: America's Safest Cities." Forbes. *Forbes.com LLC*. 2015. Web. 13 November 2015.

"How to Find nth Occurrence of Character in a String?" Stack Overflow. *Stack Exchange*. 2015. Web. 13 November 2015.

"Minneapolis Police Department Uniform Crime Report Summary." Minneapolis Police Department. *Minneapolis Police Department.* 8 August 2014. Web. 13 November 2015.