Team 6129 Page **1** of **45**

For office use only	For	office use only
T1	F1	
Τ2	F2	
Т3	F3	
Τ4	F4	

2015

18th Annual High School Mathematical Contest in Modeling (HiMCM) Summary Sheet (Please attach a copy of this page to your Solution Paper.)

Team Control Number: 6129 Problem Chosen: B

Please paste or type a summary of your results on this page. Please remember not to include the name of your school, advisor, or team members on this page.

Summary

Our task was to analyze the safety of My City given crime data of the city. In analyzing the city, it is important to be able to compare the city to other similar sized cities. Simply giving a value for the safety of a single city is meaningless without comparing it to other cities. In order to do this, we designed a simple model and extended model to process different specificities of crime data and return a bounded index of city safety.

Our model assigns a cumulative safety index from 0 to 100, in which an index of 100 indicates absolute safety and lower indices signify increased levels of danger. We categorized and weighted crimes based on their severity in the calculation, and, to calculate our safety index, we inputted the crime rate of a crime category into a logistic function.

We opted to use logistic functions to describe the effect of crime rate on the safety of a city. We believe that this approach yields an effective model because logistic functions are monotonic, meaning that they are either always increasing or always decreasing. This aligns with our assumption that the addition of a crime either negatively impacts the safety of a city or does not affect it. Furthermore, a logistic function is bounded above and below, yielding a bounded safety index. This allows one to make meaningful comparisons between the safety indices of different cities or other populaces.

We choose our solution set of logistic functions using a genetic algorithm, This process iterates over many generations, gradually shifting the set toward better fits for the predetermined data and eventually converging to a single solution if the conditions are selective enough.

Team 6129 Page **2** of **45**

Our simple model calculates the safety rating based on crime rate of the 7 major felonies. To create our model, we used a genetic algorithm to find the optimal set of logistic functions. Each logistic function correlates to a category of crime severity dependent on the crime rate.

When we ranked the cities using our simple model, the results were surprisingly similar to two published rankings of the 11 cities. My City had a relatively low safety index of 43.28 compared with other large U.S. cities, and was only .05 below that of Chicago using 2014 data.

The extended model expanded upon our simple model to use a much more comprehensive report of crime, such as the provided data for My City. This model is not applicable to most other cities just because they do not provide such a comprehensive and detailed list of offenses. Since we only have one city with such comprehensive data, we compared different districts within the city, determined by the thousands and hundreds places of the beat number location of the offense.

For both models, we determined the importance of different offenses when considering public safety. We determined that basing these rankings on the first punishment for the crime was almost always suitable in determining the safety hazard the crime had.

In addition to finding the safety of My City, our results from other cities hold mostly consistent to published safety ranking based on criminal activity in different cities.

Team 6129 Page **3** of **45**

No Safety, Know Pain Know Safety, No Pain

2015 HiMCM Problem B Team 6129

Team 6129 Page **4** of **45**

Introduction

In the world we live in today, many people live in urban or suburban areas. For this reason, residents of these metropolitan areas should be aware of the safety of the city they live in so they can take necessary precautions and educate youth about what actions are safe. In addition, judging the safety of a city allows the government of the city to analyze what factors and laws make other cities safer, and how they can replicate these conditions to make their city a safer place as well.

Problem Restatement/Interpretation

My City is a major international center for commerce, technology, finance and travel. Its population consists of 2.8 million people, as well as an additional 6 million people in the surrounding metropolitan area. Our modeling team has been given a data set of criminal activity in My City between July 5, 2014 and July 18, 2014. This data set includes additional details: primary and secondary crime description, crime location, whether an arrest was made, whether or not this was a domestic crime, and the beat number of the police route.

Our goal is to create a mathematical model to analyze the data and determine the safeness of My City.

Team 6129 Page **5** of **45**

The Model

Goals of Our Model

Using our mathematical model, we seek to:

- 1. Determine a safety rating for My City.
- 2. Compare the safety rating of individual districts in My City to each other and the city as a whole.
- 3. Compare My City's safety rating to that of a real city very similar to My City.
- 4. Compare My City's safety rating to that of real cities around the world, including similar sized cities and different sized cities, when given limited data.

Variables and Definitions

The relative *severity* of a crime indicates the degree to which the crime impacts the safety of a given region. For example, (for both reasons we provide later and intuitively) underage drinking is less severe than criminal sexual assault. Homicide is more severe than gambling.

Team 6129 Page **7** of **45**

Assumptions and Justifications

Our entire list of assumptions and justifications is listed below. Later on in our paper, the assumptions are provided again in the sections where they are relevant; please refer to the list below for their corresponding justifications.

Assumption 1: The relative severity of a crime can be directly determined from the severity of the potential maximum penalty for committing said crime, as put forth by the state of Illinois in 2000 (i.e. a crime with a more severe penalty is more severe than a crime with a less severe penalty). The severity of a penalty depends on monetary fine and number of years in jail or prison.

Justification: The government is fair and believes in justice. Therefore, offenders are given the punishments that they most deserve for committing their crimes.

Assumption 2: If the penalty for committing a crime becomes more severe after the first offense, only the penalty for the first offense is taken into consideration.

Justification: The primary purpose of increasingly severe punishments is to discourage offenders from repeating their crime, and not to give offenders the most deserved punishment.

Assumption 3: The penalties for certain crimes vary according to the "degree" of the crime (i.e. the penalty for property damage varies according to the monetary value of the property damaged; the penalty for possession of cannabis varies according to the amount of cannabis possessed, etc.). In instances such as these, the lower median penalty is the penalty taken into consideration for the building of our model. For example, if a crime results in penalties with four progressively increasing penalties depending on the "degree" of the crime, the more severe of the two least severe penalties is used for our model.

Justification: In many instances, our given data set does not distinguish between differing degrees of the same crime. We take the average or median punishment in order to build a practical model and account, as best as we can, for these different degrees. Differences in the degrees of the same crime are eventually negligible over time if the median punishment is taken.

Team 6129 Page **8** of **45**

Finally, because less severe crimes occur more often than severe crimes, we use the *lower* median.

Assumption 4: While locations may have an impact on how much of a safety hazard a crime creates, we will consider location data to be constant and not important.

Justification: While a crime in a residential area is more of a safety hazard than a crime in a deserted alleyway, most cities have similar percentages of crimes happening in similar locations, and because this data is similar, it is negligible, and would take a lot of extra time to create a model that is only marginally better.

Assumption 5: The generally accepted safety rankings of major cities in the United States, provided by the CQ Press, are correct and accurate. The generally accepted safety rankings of Chicago districts, provided by the Numbeo, are correct and accurate. Justification: Both are reputable and careful sources.

Assumption 6: Based on government data showing that crime rate increases in summer, we will assume that crime data per month followed a similar trend in 2014.

Justification: Government data has shown that the month of July has the most crime in the United States overall, and major cities such as New York and Chicago. While this government data was collected between 2010 and 2012, there is no major factor that would drastically change this statistic between then and 2014.¹

Assumption 7: The committing of a crime makes a city less safe.

Justification: Crimes affect safety of the average individual of a community either not at all, or negatively. Therefore, we can safely assume that more of any type of crime can only make a city less safe or keep it the same in terms of safety level.

¹ http://www.governing.com/gov-data/safety-justice/crime-data-statistics-for-cities-local-law-enforcem ent-by-month.html

Team 6129 Page **9** of **45**

Assumption 8: We determine districts by taking the thousands and hundreds digits of the beat number.

Justification: This is a common way of determining a district number based on beats, but even if this was not the actual way districts were recorded in My City, the data was just used to test the extended model with multiple data points, rather than just the one of My City.

Assumption 9: Crimes without arrests make a city more unsafe.

Justification: We can justify this for two reasons. The first is that the criminal may repeat crimes if not arrested, and this increases the danger for citizens of the city. The other reason is that a fewer number of arrests mean the police force of the city is less capable of protecting its residents.

Team 6129 Page **10** of **45**

Model Concept

Most current safety ratings for cities around the world are based on two factors: the number of crimes and the population, and the number of crimes per unit population (100,000 people) is generally accepted as a metric. Unfortunately, these safety ratings, while easy to calculate, are unbounded, ignore crime location or police action, and/or are unspecific to crimes of different nature. Our model seeks to rectify these issues.

Specifically, our model assigns a cumulative safety index from **0 to 100**, in which an index of **100** indicates absolute safety and lower indices signify increased levels of danger. We were provided only with the population and crime logs over two weeks of My City, and thus decided to base our safety model on crime rates, without regard to ease of living, income, etc. Because not all crimes equally endanger a city populace, as per Assumption 1, levels of severity are assigned to each category of crime, depending on the severity of the penalty. We use the state of Illinois's classification of misdemeanors and felonies, shown below in Table 1 and taken from <u>http://www.crimeandpunishment.net/IL/</u>.

Table 1. State of Illinois classification of misdemeanors and felonies, their corresponding penalties, and our assigned severity on a scale of 0 to 10.

Severity	Penalty Name	Penalty
0 (least severe)	N/A	Non-monetary and non-jail/prison penalty
1	Class C misdemeanor	Up to \$1500 fine and/or up to 30 days in jail
2	Class B misdemeanor	Up to \$1500 fine and/or up to 6 months in jail
3	Class A misdemeanor	Up to \$2500 fine and/or up to 1 year in jail
4	Class 4 felony	Up to \$25,000 fine and/or 1 to 3 years in pen.
5	Class 3 felony	Up to \$25,000 fine and/or 2 to 5 years in pen.
6	Class 2 felony	Up to \$25,000 fine and/or 3 to 7 years in pen.

Team 6129 Page **11** of **45**

7	Class 1 felony	Up to \$25,000 fine and/or 4 to 15 years in pen.
8	Class X felony	Up to \$25,000 fine and/or 6 to 30 years in pen.
9	Extended term Class X felony	Over \$25,000 fine and/or >6 to >30 years in pen.
10 (most severe)	Extended term Class X felony	Over \$25,000 fine and/or 20 to 60 years or natural life in pen. or death penalty

Based on the relative severities of crimes listed in My City's crime log, we grouped these crimes into certain categories. This process will be described in detail later on in this paper.

We wanted our model to take the severity of each crime category into account, so that more severe crimes would be given more weight than less severe crimes in the calculation of our safety index. Therefore, for each crime category, the crime rate (defined as crimes per 100,000 people per year) was inputted into a logistic function, which outputs a scaled value. These values are summed up to yield the final safety index.





Shown above in Figure 1 is an example of such a logistic function, which takes on a sigmoidal "S" shape, and is different from conventional linear models in that they are strictly bounded above and below. Essentially, we assign such a function to each crime category and input the crime rate, the total number of crimes in that category occurring over a 1 year span per

Team 6129 Page **12** of **45**

100,000 people (if less than 1 year of data is available, we scale it as per Assumption 6; the exact parameters are described later), to obtain a number, called the crime category value. We intentionally bound these functions below at 0 so that the sum of all minimum crime category values is 0, and we intentionally create these functions so that, once scaled, the sum of all maximum crime category values is 100. As such, the sum of the scaled crime category values is the safety index.

A logistic function was used to model the increase in danger relative to the crime rate because it outputs a bounded value for any input. Because of this property, the function value increases faster at lower crime rates and increases much slower at higher crime rates. This is ideal because from a statistical standpoint, a hypothetical city with an immensely high crime rate is not any safer than a city with a marginally higher crime rate. While each crime category is modeled by a distinct logistic function, each function follows several fixed conditions. Because a city with no crime should receive an index of 0, each logistic function must output a value of 0 for a 0 crime rate. Additionally, the maximum output for each logistic function is relative to the severity of its corresponding crime category. This is to ensure that a high rate of a less severe crime does not contribute more to the safety index than a high rate of a more severe crime.

A logistic function takes on the form:

$$y = \frac{L}{1 + e^{-k(x - x_0)}},$$

where L is the function's maximum value, k is the steepness of the curve, and x_0 is the value of the sigmoid's midpoint. In building our model, we must be careful in our manipulation of these variables because any change has important implications for the effect of crime on safety.

Team 6129 Page **13** of **45**



Figure 2. Manipulations on a logistic function.

The effect of various manipulations on the logistic function is shown in Figure 2. Specifically, they are a change in steepness resulting from a change in k, and a horizontal shift resulting from a change in x_0 (it is important to note here that, in constructing our model, all functions were vertically shifted to pass through the origin, at (0,0), but the graphs to the right do not reflect this).

As mentioned previously, we chose a logistic function because it increases faster at lower crime rates and increases much slower at higher crime rate, which is statistically favorable. However, the question remains as to what the best values for L, k, and x_0 are--where should the function increase rapidly, how rapidly, and up to what extent? In the context of the problem, these questions play out as: how much safer is City A if it has 1 murder instead of 2? 80 instead of 100? At what point does it become extremely unsafe?

To answer these questions, we could choose to arbitrarily assign values to L, k, and x_0 . However, such a method is extremely *arbitrary*, and there is really no justification for choosing three values to represent each crime category. Therefore, instead of doing this, we provide a set of conditions that our set of logistic functions must satisfy, and then allow a **genetic algorithm** to determine the most mathematically optimal values for L, k, and x_0 .

To clarify, we choose our solution set of logistic functions using a genetic algorithm, which, as the name implies, was inspired by the recombinative nature of genetics in biological evolution.

Team 6129 Page **14** of **45**

This algorithm takes a population of randomly generated "chromosomes," which in this case are sets of logistic functions, to form the first generation of solutions. At each generation, the fitness of each solution to predetermined input data is evaluated (i.e. known data/characteristics is plugged into the solution set, and then the solutions are ranked based on fitness, where fitness is defined as how close the result of a solution is to our desired outcome). The top ranking solutions are retained in the population, and crossover and mutation processes act on these solutions to yield variants that are then added to the population until the population reaches its original size. This new population is the next generation. This process iterates over many generations, gradually shifting the set toward better fits for the predetermined data and eventually converging to a single solution if the conditions are selective enough.

Process Overview and Description

We seek to generate a model to calculate a safety rating. Because safety ratings are only meaningful if relative, we first create a simple model based off of the major crime data and known safety ratings of the eleven largest cities (since My City is large) in the United States. In particular:

- The cities we considered were Chicago, New York, Los Angeles, Houston, Philadelphia, Las Vegas, Phoenix, San Antonio, San Diego, Dallas, and San Jose.
- All of the major crime data for these cities is taken from <u>https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2014/crime-in-the-u.s-2014/</u>.
- The safety ratings of each of these cities, which we call *expected ratings*, is courtesy of <u>http://www.numbeo.com/crime/region_rankings.jsp?title=2015-mid®ion=021</u>.
- Our simple model considers five crime categories in decreasing order of severity: murder, rape, robbery, aggravated assault, and theft. These were chosen based on the wide availability of data.

We put only one restriction on the genetic algorithm, concerning the last bullet point above: if crime category A is more severe than crime category B, then the maximum value of crime category A's logistic function must be greater than that of crime category B ($L_A > L_B$). Beyond this, we rely solely on the genetic algorithm to produce the most mathematically optimal solution.

Team 6129 Page **15** of **45**

We let the genetic algorithm's "input data" be the crime rate for each of the five crime categories of each major city. At each generation, 10,000 different solution sets of 5 logistic functions are generated, and the data for each city is plugged into these functions to obtain a safety rating for each city. We define the fitness of each solution to be how closely it matches the *expected ratings*; in particular, if we let the fitness be *F* (sum of squared differences), the set of cities be *C*, the calculated safety rating of city $i \in C$ be R_i , and the expected rating be E_i , then

$$F = \sum_{i \in C} \quad (R_i - E_i)^2$$

The smaller F is, the more "fit" a solution set is (we seek to minimize F). From each generation, we keep the 50 most fit solutions, and mutations take place to create variant solutions that are then added to the next generation until it reaches 10,000 total solutions. This process was repeated for 100 generations. After 100 generations, the solutions generally converged to one set of logistic functions that satisfied our given conditions.

It is important to note here that we defined fitness relative to another safety index, courtesy of <u>http://www.numbeo.com/crime/region_rankings.jsp?title=2015-mid®ion=021</u>. These safety ratings consider variables and use methods completely different from our model. While it is not best practice to build one model off of another one, we believe it is important emphasize that we neither desire nor expect for *F* to ever reach 0. That is, we use the expected ratings as a general guideline to help our model evolve.

The solution we attain is our simple model. We can then calculate the safety rating of My City using this model, calculate the safety ratings of the major cities and compare our calculated rankings with generally accept rankings, and compare My City to these major cities. If the results of our simple model are comparable to existing rankings and indexes, then we consider our model to be successful and work to extend it so that it considers the more comprehensive My City crime logs data available to us.

Team 6129 Page **16** of **45**

Thus, after building and analyzing our Simple Model, we created our Extended Model. This second model considers ten crime categories especially tailored to the level of detail given by My City crime logs, and does not use the genetic algorithm. Instead, we use the logistic functions produced from the Simple Model and their average values for a given set of data to calculate a crime category value for each of the 10 crime categories, while considering the role of arrests (this process is described in detail in the section **Extended Model**).

We inputted crime rates from My City to calculate an alternative safety rating and compared it to the one obtained from the Simple Model, and then we input the crime rates from each of 22 districts in My City to examine the safety of different regions (a *district* is defined in Assumption 8).

Team 6129 Page **17** of **45**

Simple Model

We sought to create a model that would provide a safety rating for a given city, My City in particular, given crime data. We constructed our model using data from other large cities in the United States, and used a genetic algorithm guided by existing models and rankings to find the mathematically optimal function. Our crime categories revolve around simple data, such as general crime rates and crimes rates of violent crimes, since these are the most accessible for many large cities. The details of this process are described in the previous section.

First, we obtained the crime rate data for the seven major felonies of the eleven largest cities in the United States, by population (all have populations over 1 million). These felonies were, in order of severity from most to least: homicide, rape, robbery, aggressive assault, burglary, larceny, and larceny of a motor vehicle. We grouped together the last three into our theft category, as they are very similar crimes and punishments are dependent on the amount stolen, leaving us with five final categories for our simple model.

We then used preexisting safety ratings of these eleven cities as the targets of our genetic algorithm. Our model is shown below:

 $S(C) = 100 - K * (M(m_0) + R(r_0) + B(b_0) + A(a_0) + T(t_0))$

where

C is a city m₀ is the murder crime rate associated with C r₀ is the rape crime rate associated with C b₀ is the robbery crime rate associated with C a₀ is the aggravated assault crime rate associated with C t₀ is the theft crime rate associated with C K is a scaling factor equal to $\frac{100}{lim(M(x)+R(x)+B(x)+A(x)+T(x))} =$

77.0870253322

By multiplying each function by K, we ensure that S outputs between 0 and 100.

Functions M(i), R(j), B(x), A(y), and T(z) are all logistic functions of the crime rate of a certain category of crime within city C. For the eleven cities, we determined an expected value

Team 6129 Page **18** of **45**

for S(C). In assessing a potential solution, we wished to take into account our assumption that severe crime on a large scale negatively impacts the safety of a society more than less severe crime on a large scale. To do so, we included the condition $L_M > L_R > L_B > L_A > L_T$, where L_F is the upper limit of function F(x) as x takes on arbitrarily large values. While this ordering may not necessarily hold when x is very small, because our model is designed to model cities with substantial crime rates.

Resulting Model

Shown below are the logistic functions we obtained for each of the crime categories.



Murder Logistic Function





Team 6129 Page **21** of **45**

Application of Model: Safety Ratings

After we got our simple model, we were able to run the data given to us for My City. To align the specific My City data with the broader crime categories, we counted the number of crimes related to each category and produced corresponding crime rates. Also, after some research, we found that July has the highest crime rate in major U.S. cities. Because all the crimes rates we used were crimes per 100,000 people per year we needed to scale our data, since we were only given data for two weeks during July 2014. The average crime rate during July 2014 was 1,282.3, while the average crime rate per month during 2014 was 1,181.4. To account for this, we calculated a proportional set of crime rates representing an average two week period during the year, and then we scaled the rates by time to obtain yearly crime rates. Our mathematical process is shown below:

My City Crime Rates
(100,000 people)(2 weeks) *Average Crime Rate Per Month in 2014
Crime Rate of July 201452 weeks
1 year

Using the newly found crime rate (crimes per 100,000 people per year), we were able to determine a safety rating for My City using our simplistic model: 43.28. This value is between the scores of Philadelphia and Chicago, making it a fairly dangerous city to live in.

Team 6129 Page **22** of **45**

Table 2. Safety indices as calculated from our simple model, with higher numbers indicating greater safety. The last 3 columns contain the rankings of the 11 cities compared to each other: the first based on our calculated indices and the latter two based on outside sources. The highest rankings indicate the least safe cities.

City	Calculated Index	Our Ranking	Numbeo Ranking ²	CQ Press Ranking ³	
My City	43.28	11	N/A	N/A	
Chicago	43.33	10	7	N/A	
New York	53.08	4	2	1	
Los Angeles	53.45	3	1	3	
Houston	45.66	8	8	9	
Philadelphia	42.15	12	11	10	
Las Vegas	45.41	9	10	6	
Phoenix	48.09	5	6	7	
San Antonio	47.57	7	5	5	
San Diego	55.39	2	3	2	
Dallas	47.95	6	4	8	
San Jose	55.61	1	9	4	

According to **Table 2**, the results of our simple model had similar results, in terms of rankings, to preexisting models. However, there were some key differences. The biggest was the ranking of San Jose. While San Jose is ranked 9th and 4th in other models, we found that San

² http://www.numbeo.com/crime/region_rankings.jsp?title=2015-mid®ion=021

³ http://os.cqpress.com/citycrime/2013/2014_CityCrimeRankings(LowtoHigh).pdf

Team 6129 Page **23** of **45**

Jose is the safest city out of the eleven. After looking through the raw data of crime rates for the major felonies, we found that this was appropriate.

Rather than forcing our ranking to match the rankings of others, or genetic algorithm favored the importance of our parameters in this certain case. This is important because it shows that our algorithm grew to create a better and more accurate model.

Team 6129 Page **24** of **45**

Extended Model

Our extended model used the results of our simple model design while accommodating the additional details in the crime logs of My City. On a 0 to 10 scale (see Appendix A), our murder logistic covered level 10 crimes, our rape logistic covered level 8 crimes, our robbery logistic covered level 6 crimes, our aggravated assault logistic covered level 4 crimes, and our theft logistic covered level 2 crimes. While theft is actually categorized as level 3, the theft logistic produces such a small value compared to the others that this difference is negligible. Furthermore, crimes in these low categories are relatively inconsequential with regard to the overall safety of the city. Using this setup, we calculate scores for crime levels 1,3,5,7,9 by taking the average of the values of the functions adjacent to them (e.g. crime level 7 is calculated by averaging the robbery and rape logistic functions, corresponding to levels 6 and 8 respectively, at the crime rate for level 7). After summing these scores, we scale this sum to an index from 0 to 100,

We categorized the severity of the various crimes on a scale of 0 to 10 (see Appendix A) and computed a crime rate for each classification. In computing each crime rate, we took into account whether perpetrators were arrested for their crimes, giving crimes without arrests greater weight. With these arrest-augmented crime rates, we calculated a safety index of 19.970 for My City. In addition, we calculated separate crime rates for each district, as defined by Assumption 8, in My City, and computed individual safety indices (see **Table 3**). These values are indicative of the relative safety of different districts within My City. In particular, districts 20, 17, and 12 rank as the safest, while districts 7, 22, and 15 rank as the most dangerous.

Team 6129 Page **25** of **45**

City/District	Calculated Index	Ranking
My City	19.97	N/A
District 1	42.09	6
District 2	26.09	12
District 3	29.09	10
District 4	21.45	16
District 5	23.82	14
District 6	21.53	15
District 7	7.31	22
District 8	29.23	9
District 9	26.95	11
District 10	18.30	18
District 11	17.77	19
District 12	51.59	3
District 14	44.16	5
District 15	17.40	20
District 16	48.80	4
District 17	55.91	2
District 18	41.03	7

Table 4. Safety indices as calculated from our extended model, with higher numbers indicating greater safety. The ranking column ranks the districts in order of greater to lesser safety.

Team 6129 Page **26** of **45**

District 19	25.92	13
District 20	57.42	1
District 22	14.64	21
District 24	37.45	8
District 25	20.90	17

Team 6129 Page **27** of **45**

Comparison and Sensitivity Analysis

Our Simple Model yielded a safety rating of 43.28 for My City, while our Extended Model yielded a safety rating of 19.97 for My City. Even though the second model was merely an extension of the first, these two safety ratings are extremely different and not comparable. This is not necessarily a bad thing, and there is a logical explanation for this result. Our extended model considers a greater number and variety of crimes, which lowers the safety rating by a considerable amount. It also considers arrest as something that increases the safety of a city, which marginally increases the safety rating. These two changes, along with the abandonment of the concept of using the sums of purely logistic functions, causes a net decrease in safety rating.

As such, it is difficult to make any meaningful comparison between the two ratings. With more input data, we could perhaps compare My City's ranking relative to other cities using both models to realize their merit.

It is most productive to analyze the mechanics of our Simple Model, as our Extended Model is merely an application of the Simple Model. As mentioned previously in the **Simple Model** section, the safety rankings we obtained for other major cities, though different than the expected ratings, are extremely comparable to those obtained by other reputable sources and models. Furthermore, our safety rating matches almost exactly with that of Chicago (less than 0.5 difference), supporting the trend we have consistently observed while building our model: My City is extremely similar to Chicago.

The sensitivity of our simple model to different inputs can easily be determined from each of the logistic functions for each crime category. The maximum value of each logistic function, shown in Table 5, corresponds to the weight of the input in that category.

Team 6129 Page **28** of **45**

Crime Category	Maximum Value
Murder	35.67834444
Rape	26.20413996
Robbery	19.1660764
Aggravated Assault	18.73171744
Theft	0.2197217639

Table 5. Crime categories and their corresponding maximum values.

As can be seen from the table, theft is almost inconsequential to the final safety rating--its contribution is barely over 0.2%. Mathematically speaking, this is because theft and other petty crimes occur very often and thus need to be weighted less in order to maintain balance in our model. As such, even large variations in the number of thievery crimes would barely affect our safety rating. In the context of real world safety, however, the results of our model seem to imply that theft, which does not result in any physical harm has less to do with safety than any of the other categories. This conclusion matches general perception--even in a very safe city, few people would be willing to leave of their belongings in public for fear of them being stolen. Theft contributes to the overall moral character of a city, not necessarily the safety; theft and safety are only loosely associated, or at least to a lesser degree than the rest of the crime categories we have considered. This may be inferred from our model.

For the rest of the four categories, the sensitivity is not as clear, as it depends both on the slope, shift, and maximum value of the logistic function.

The rape and robbery logistic functions have extremely steep slopes followed by a plateau, whereas aggravated assault and murder both increase over the entirety of the crime rates that we have considered (seen from the graphs shown in the **Simple Model** section). This means that, after x axis value where the sigmoid starts to plateau, both rape and robbery rates are deemed high enough and equally horrible in terms of a city's safety, and thus become close the

Team 6129 Page **29** of **45**

Maximum Possible Value very quickly after this one point. Robbery is a common occurrence, and thus, when a high crime rate is inputted into the logistic function, a value very near the maximum value will be outputted. As such, the safety rating is not very sensitive to robbery rates. Rape, however, is less common than robbery, and as such the safety rating is still reasonably sensitive to changes in rape rate.

The murder and aggravated assault logistic functions both increase for a considerable period before plateauing, in comparison to rape and robbery. Thus, the safety rating is sensitive to both, but more so to murder because the maximum value is larger for murder than for aggravated assault. Note that the discussion above considers sensitivity above a certain point (once the logistic function begins to plateau), and thus, our model is still very sensitive to extremely decreased rates of murder, aggravated assault, robbery, and rape.

Our Extended Model is extremely sensitive to the arrest rate, and the sheer largeness of the amount of crimes means that the scores will all be at the low end. With more time, we would liked to scale or adjust this in some manner, but we recognize that this model is a little too sensitive to changes. The primary purpose of this model was to yield insight into the safety of the districts. Unfortunately, we did not have time to conduct extremely thorough sensitivity analysis for our two models. Given more time, we would have liked to look at the role of nonconsequential inputs in our two models and vary conditions or input data for our genetic algorithm.

Team 6129 Page **30** of **45**

Strengths and Weaknesses

Strengths of Simple Model

- 1. The simple model is applicable to many cities, because the 7 major felony data is published for many cities in the United States and around the world.
- 2. Our model is bounded by 0 and 100. Even without other cities' scores to compare, one can get a general idea of the safety of a city.
- 3. Our model uses logistic functions to accommodate for a potentially infinite amount of a certain crime, while still keeping the function bounded.

Weaknesses of Simple Model

- 1. While the indices for city safety are bounded between 1 and 100, the actual values we see for our 11 cities are contained between 7 units from 50. In order for more comprehensive analyses, we would want the indices to occupy a larger percentage of the full range of values.
- 2. Only using the 7 major felony data, we excluded some of the potentially violent crimes. While the 7 major felony data is quite representative of a city's crime, it is not necessarily a perfect representation

Strengths of Extended Model

- 1. The extended model appear to give a much wider range of ratings. This makes it easier for individuals to interpret the ratings and what they mean.
- 2. This model takes into account many more details than the simple model. More crimes were accounted for, as well specific categories of crimes rather than broad categories. Also this looks at whether an arrest was made or not for each crime.

Weaknesses of Extended Model

- 1. The scores are skewed towards the lower end of the range.
- 2. Our extended model was not optimized using a genetic algorithm. There may have been more optimal models.

Team 6129 Page **31** of **45**

Conclusion

In order to determine a safety rating for My City, we had to first look towards other cities and pre-existing models and rankings. Only by comparing the rating of My City to other cities using our model can we get an idea of how safe My city is.

Our model is based on the idea that more severe crimes hold more weight in determining the safeness of a city. Each crime is put into five or ten categories based on the severity of a crime. Each city gets a score in each of these categories dependent on the crime rate. The sum of these scores is the final safety rating of a city. The scores are calculated using logistic functions obtained by using a genetic algorithm. Logistic functions are favorable because they have an upper and lower bound. Also, an increase in crime rate does not linearly correlate to the safeness of a city.

In order to find the logistic functions, we first found raw data that gave the crime rates for the eleven biggest cities in the United States, as well as its safety index and ranking based on other models. These data points were then used to initialize our first model, the Simple Model. Using a genetic algorithm, we mutated our model until it converged towards a model with similar results to pre-existing models. This was necessary in order to ensure that values given to each category were not arbitrary and held weight. Using this model, we calculated the safety rating for My City, and compared it to the rating of other cities, where the higher the rating, the safer the city. My City received an index of 43.28, relatively low compared to other large cities in the United States. It was only 0.05 lower than that of Chicago, a city with a similar population size to My City.

Furthermore, we used our simple model to create an extended model. We were given much more data for My City than we could find for other cities, allowing us to go into a more indepth analysis of My City's safety. In order to do this, we used the functions from our simple model, to create more specific categories. In our extended model, we also took into account whether someone was arrested for a crime, putting less weight on crimes where someone was arrested.

Team 6129 Page **32** of **45**

Using our extended model, we calculated the safety rating for the 23 districts within My City in order to test our model. This allows us to get a much more specific insight on My City and pinpoint dangerous and safe areas.

Team 6129 Page **33** of **45**

Extensions

Given more time, we would have been able to more rigorously refine our model using the genetic algorithm. Using more generations, a different mutation rate, and a larger population size, we could have more accurately synthesized our model. Increasing these significantly would leave the algorithm running past the given time, but would give us a more accurate model.

In addition, with more time we could have tested the accuracy and stability of our model by testing it against more safety indices from a variety of sources and tried rerunning the algorithm provided with less or more data and seen if our results were similar, or changed drastically. If the results were similar we would be able to ensure the stability, and in extension, accuracy of our model.

Given more data, we could factor the actual punishment of each crime into our model. Currently, we used the maximum possible punishment to determine the severity of a crime, but depending on many factors, these punishments vary widely. This wide variance of punishments for the same crime indicates that the maximum possible punishment is not always indicative of the average punishment.

Also given more time, we could have included the location in our model. While we assumed that location wasn't very important to consider, it would still be something to include when refining the model to increase the accuracy and comprehensiveness of the model.

We also would focus on making our simple model less concentrated near the center of the range of values. Our goal would be for real cities to be spread throughout the range of indices so we could more accurately analyze the differences between cities. We would have collected district data using both the simple and extended models to find a better comparison between our two models.

Team 6129 Page **34** of **45**

Appendix A

The categories each crime are put into.

Group	Crimes				
0	Kidnapping: Unlawful Interfere/Visitation				
	Liquor Law Violation: Illegal Possession By Minor, Illegal Consumption By Minor				
	Other Offense: Probation Violation, Gun Offender: Duty To Register, Parole Violation, License Violation, Violate Order Of Protection, False/Stolen/Altered Trp, Gun Offender: Duty To Report Change Of Information, Sex Offender: Fail To Register, Vehicle Title/Reg Offense, Violent Offender: Fail To Register New Address, Gun Offender: Annual Registration, Sex Offender: Fail Reg New Add, Sex Offender: Prohibited Zone, Violent Offender: Annual Registration, Violent Offender: Duty To Register				
1	Assault: Simple				
	Deceptive Practice: Bogus Check, Impersonation				
	Other Offense: Obscene Telephone Calls				
	Public Peace Violation: Mob Action				
2	Criminal Trespass: To Land				
	Interference With Public Officer: Obstructing Service				
	Narcotics: Poss: Cannabis 30gms Or Less				
	Other Offense: Harassment By Telephone, Harassment By Electronic Means, Other Crime Against Person, Animal Abuse/Neglect				
3	Assault: Agg Po Hands No/Min Injury, Aggravated: Handgun, Aggravated: Knife/Cutting Instr, Aggravated: Other Dang Weapon, Pro Emp Hands No/Min Injury, Aggravated Po:Knife/Cut Instr, Aggravated: Other Firearm, Aggravated Po: Other Dang Weap, Agg Pro.Emp:Knife/Cutting Inst, Agg Pro.Emp: Other Dang Weapon				
	Battery: Simple, Domestic Battery Simple, Pro Emp Hands No/Min Injury				

Team 6129 Page **35** of **45**

	Concealed Carry License Violation: Other				
	Criminal Damage: To Vehicle, Criminal Defacement				
	Criminal Trespass: To Residence, To State Sup Land, To Vehicle				
	Deceptive Practice: Counterfeiting Document, Counterfeit Check, Unlawful Use Of Recorded Sound, Financial Identity Theft \$300 And Under, Attempt - Financial Identity Theft, Deceptive Collection Practices, Unidentifiable Recording Sound, Illegal Possession Cash Card				
	Gambling: Game/Dice, Other, Game/Cards				
	Interference With Public Officer: Resist/Obstruct/Disarm Officer, Obstructing Identification				
	Liquor Law Violation: Sell/Give/Del Liquor To Minor, Liquor License Violation				
	Narcotics: Possession Of Drug Equipment, Pos: Hypodermic Needle				
	Offense Involving Children: Endangering Life/Health Child				
	Other Offense: Telephone Threat, Other Crime Involving Property, Other Weapons Violation, Possession Of Burglary Tools				
	Prostitution: Solicit On Public Way, Solicit For Business, Solicit For Prostitute				
	Public Peace Violation: Reckless Conduct, Peeping Tom				
	Sex Offense: Public Indecency				
	Theft: \$500 And Under, From Coin-Op Machine/Device				
	Weapons Violation: Poss Firearm/Ammo:No Foid Card, Unlawful Poss Other Firearm, Unlawful Use Other Dang Weapon, Unlawful Poss Ammunition				
4	Assault: Aggravated Po: Handgun, Aggravated Po: Other Firearm				
	Criminal Damage: To Property				
	Deceptive Practice: Credit Card Fraud, Financial Identity Theft Over \$ 300, Illegal Use Cash Card, Fraud Or Confidence Game, Computer Fraud				
	Stolen Prop: Buy/Receive/Pos., Finan Exploit-Elderly/Disabled, Unlawful Use Of A Computer				
	Interference With Public Officer: Obstructing Justice, Bribery				
	Narcotics: Manu/Del:Cannabis 10gm Or Less, Poss: Cannabis More Than 30gms,				

Team 6129 Page **36** of **45**

	Alter/Forge Prescription
	Offense Involving Children: Other Offense, Child Abduction
	Public Peace Violation: Other Violation, False Police Report, Public Demonstration
	Sex Offense: Criminal Sexual Abuse
	Stalking: Simple, Cyberstalking
	Theft: From Building, Retail Theft, Attempt Theft, Delivery Container Theft
	Weapons Violation: Unlawful Poss Of Handgun, Unlawful Use/Sale Air Rifle, Unlawful Use Handgun, Unlawful Sale Handgun, Unlawful Use Other Firearm
5	Battery Aggravated: Other Dang Weapon, Aggravated: Handgun, Aggravated:Knife/Cutting Instr, Agg Po Hands No/Min Injury, Aggravated Po: Other Dang Weap, Agg: Hands/Fist/Feet No/Minor Injury, Aggravated: Other Firearm, Aggravated Po: Knife/Cut Instr, Agg Pro.Emp: Other Dang Weapon
	Crim Sexual Assault: Attempt Non-Aggravated, Non-Aggravated
	Criminal Damage: To City Of Chicago Property, To State Sup Prop
	Deceptive Practice: Theft Of Labor/Services, Theft Of Lost/Mislaid Prop, Aggravated Financial Identity Theft, Forgery
	Interference With Public Officer: Escape
	Intimidation: Extortion, Intimidation
	Motor Vehicle Theft: Cycle, Scooter, Bike W-Vin, Attempt: Cycle, Scooter, Bike W-Vin
	Offense Involving Children: Contribute Delinquency Of A Child
	Other Offense: Other Vehicle Offense
	Public Peace Violation: Arson Threat, Bomb Threat
	Sex Offense: Indecent Solicitation/Child
	Stalking: Violation Of Stalking No Contact Order, Aggravated
	Theft: Over \$500, Pocket-Picking, Purse-Snatching
	Weapons Violation: Reckless Firearm Discharge

Team 6129 Page 37 of 45 6 Arson: By Fire **Battery:** Aggravated Domestic Battery: Knife/Cutting Inst, Aggravated Domestic Battery: Other Dang Weapon, Aggravated Of A Senior Citizen **Burglary:** Forcible Entry, Unlawful Entry, Attempt Forcible Entry Deceptive Practice: Theft By Lessee, Motor Veh, Embezzlement, Theft By Lessee, Non-Veh Kidnapping: Kidnapping Motor Vehicle Theft: Automobile, Theft/Recovery: Automobile, Theft/Recovery: Truck, Bus, Home, Att: Automobile, Truck, Bus, Motor Home Other Offense: Hazardous Materials Violation Robbery: Strongarm - No Weapon, Armed: Knife/Cutting Instrument, Armed: Handgun, Attempt: Strongarm-No Weapon, Armed: Other Dangerous Weapon, Attempt: Armed-Handgun, Vehicular Hijacking, Attempt: Armed-Knife/Cut Instr, Attempt: Armed-Other Dang Weap, Attempt: Aggravated, Armed: Other Firearm Sex Offense: Agg Criminal Sexual Abuse, Att Agg Criminal Sexual Abuse, Att Crim Sexual Abuse 7 **Burglary:** Home Invasion Narcotics: Poss: Crack, Poss: Heroin(White), Poss: Barbituates, Poss: Cocaine, Found Suspect Narcotics, Poss: Heroin(Brn/Tan), Poss: Pcp, Poss: Synthetic Drugs, Forfeit Property, Poss: Hallucinogens, Poss: Look-Alike Drugs. Poss: Methamphetamines, Poss: Amphetamines, Attempt Possession Narcotics Offense Involving Children: Child Pornography **Other Narcotic Violation:** Intoxicating Compounds **Robbery:** Aggravated, Aggravated Vehicular Hijacking 8 Arson: Aggravated Battery: Aggravated Of A Child Crim Sexual Assault: Aggravated: Other, Predatory, Aggravated: Knife/Cut Instr, Aggravated: Handgun, Aggravated: Other Dang Weapon, Attempt Agg: Other Kidnapping: Unlawful Restraint, Child Abduction/Stranger, Aggravated Narcotics: Solicit Narcotics On Publicway Offense Involving Children: Child Abuse, Agg Crim Sex Abuse Fam Member,

Team 6129 Page **38** of **45**

	Crim Sex Abuse By Fam Member, Sex Asslt Of Child By Fam Mbr	
9	Battery: Aggravated Domestic, Hands/Fist/Feet Serious Injury, Agg Pro Emp Hands Serious Inj, Hands/Fist/Feet Serious Injury, Po Hands Etc Serious Inj Narcotics: Manu/Deliver:Crack, Manu/Del:Cannabis Over 10 Gms, Manu/Deliver: Heroin (White), Manu/Deliver: Hallucinogen, Manu/Deliver:Pcp, Manu/Deliver:Synthetic Drugs, Manu/Deliver:Cocaine, Manu/Deliver:Barbituates, Del Cont Subs To Person <18, Criminal Drug Conspiracy	
10	Homicide: First Degree Murder	
Raw Data of Cities		

Raw Data of Cities

City	Populati on	Murder and Nonnegligent Manslaughter	Rape	Robber y	Aggravated Assault	Theft (total)
	2,800,00		36.786	344.76		
MY CITY	0	17.1101	72	86	510.7366	2758.149
	2,724,12					
Chicago	1	15.1	49.3	359.9	460	3126.1
	8,473,93					
New York	8	3.9	25.8	195.7	371.3	1602
Los	3,906,77					
Angeles	2	6.7	28.8	203.5	251.8	2128
	2,219,93					
Houston	3	10.9	36.6	458.8	485.1	4693.7
Philadelp	1,559,06					
hia	2	15.9	77.4	447.1	481.1	3387.7

Team 6129 Page **39** of **45**

Las Vegas	1,530,89 9	8	51	319.1	463.1	2923.4
Phoenix	1,529,85 2	7.5	65.8	193	305.7	3724.3
San Antonio	1,428,46 5	7.2	75.4	124.4	332.3	5417.8
San Diego	1,368,69 0	2.3	27.1	96.3	255.2	1959
Dallas	1,272,39 6	9.1	61.4	303.1	291.1	3589.2
San Jose	1,009,67 9	3.2	30.3	106.2	181.4	2434.1

Team 6129 Page **40** of **45**

Appendix B

The

package that facilitates the genetic algorithm

library(genalg)

maximum scaled index

SCALE_VAL <- 100

Team 6129 Page **41** of **45**

number of crime categories for simple model

NUM_CAT <- 5

city and district crime rates according to 1-10 scale (very large data set)
district <- c(...)</pre>

 $\ensuremath{\texttt{\#}}$ main function that initiates the genetic algorithm and returns the results

```
rungenalg <- function() {</pre>
```

rbga(stringMin=rep(c(0.001, 0.001, -10), NUM_CAT), stringMax=rep(c(1, 0.5, 10), NUM_CAT),

popSize=10000, iters=100, elitism=50, evalFunc=evaluate, verbose=TRUE)

}

```
# function invoked by genetic algorithm to evaluate fitness of solution
evaluate <- function(string=c()) {
  mat <- matrix(string, nrow=3)
  zeroes <- apply(mat, 2, function(a) {
    return(a[1] / (1 + exp(a[2] * a[3])))
  })
  limits <- mat[1,] - zeroes</pre>
```

if (is.unsorted(rev(limits))) {

return(100000000)

}

Team 6129 Page **42** of **45**

```
matInput <- matrix(input, nrow=NUM_CAT+1)</pre>
  indices <- apply(matInput, 2, function(caseData) {</pre>
    index <- caseData[1]</pre>
    stats <- tail(caseData, -1)</pre>
    total <- sum(apply(rbind(mat, stats, zeroes), 2, function(a) {</pre>
      return(a[1] / (1 + \exp(a[2] * (a[3] - a[4]))) - a[5])
    }))
    return(((1 - total / sum(mat[1,], -zeroes)) * SCALE_VAL - index))
  })
 return(sum(indices^2))
}
# evaluated safety index of city and districts according to extended model
evalDistricts <- function(a) {</pre>
  f10 <- function(x) {</pre>
      return(a[1] / (1 + exp(a[2] * (a[3] - x))) - a[1] / (1 + exp(a[2] * a[3])))
  }
  f8 <- function(x) {</pre>
    return(a[4] / (1 + exp(a[5] * (a[6] - x))) - a[4] / (1 + exp(a[5] * a[6])))
  }
  f6 <- function(x) {</pre>
   return(a[7] / (1 + exp(a[8] * (a[9] - x))) - a[7] / (1 + exp(a[8] * a[9])))
  }
```

Team 6129 Page **43** of **45**

```
f4 <- function(x) {</pre>
   return(a[10] / (1 + exp(a[11] * (a[12] - x))) - a[10] / (1 + exp(a[11] * a[12])))
  }
  f2 <- function(x) {
   return(a[13] / (1 + exp(a[14] * (a[15] - x))) - a[13] / (1 + exp(a[14] * a[15])))
  }
 mat <- matrix(a, nrow=3)</pre>
  zeroes <- apply(mat, 2, function(a) {</pre>
  return(a[1] / (1 + exp(a[2] * a[3])))
 })
  limit <- 2 * sum((mat[1,] - zeroes)) - (mat[1,1] - zeroes[1]) / 2</pre>
  distRaw <- districts
  distRaw[is.na(distRaw)] <- 0</pre>
 distMat <- matrix(distRaw, nrow=10)</pre>
  results <- apply(distMat, 2, function(b) {</pre>
    return(f2(b[1])/2 + f2(b[2]) + (f2(b[3])+f4(b[3]))/2 + f4(b[4]) +
              (f4(b[5])+f6(b[5]))/2 + f6(b[6]) + (f6(b[7])+f8(b[7]))/2 +
             f8(b[8]) + (f8(b[9])+f10(b[9]))/2 + f10(b[10]))
  })
 return((1 - results / limit) * 100)
}
```

Works Cited and Bibligraphy

- <u>http://www.nyc.gov/html/nypd/downloads/pdf/analysis_and_planning/seven_major_felon</u> y_offenses_2000_2014.pdf
- <u>https://portal.chicagopolice.org/portal/page/portal/ClearPath/News/Crime%20Statistics/1</u> _pdfsam_CompStat%20Public%202015%20Week%2043.pdf
- <u>http://www.numbeo.com/crime/region_rankings.jsp?title=2015-mid®ion=021</u>
- <u>https://safecities.economist.com/wp-</u> <u>content/uploads/2015/01/EIU_Safe_Cities_Index_2015_white_paper-1.pdf</u>
- <u>http://www.governing.com/gov-data/safety-justice/crime-data-statistics-for-cities-local-law-enforcement-by-month.html</u>
- <u>https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2</u>
- <u>http://www.sandiego.gov/police/pdf/2014/201412cumneighbor.pdf</u>
- https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2014/crime-in-the-u.s.-2014/
- <u>http://os.cqpress.com/citycrime/2013/2014_CityCrimeRankings(LowtoHigh).pdf</u>
- <u>http://www.crimeandpunishment.net/IL</u>

Team 6129 Page **45** of **45**

Mr. Mayor of My City,

It is very important that you pay close attention to the safeness of your city, in respect to crime rates. The safeness of your city can greatly impact the economics and living conditions within My City. Using the data available to us, we attempted to create a model that would give us a safety rating of My City so that we could compare it to other cities. We found that My City scored very poorly, compared to other major U.S. cities. Using our model, we ranked My City 11th out of the 12 cities with over a million citizens: San Jose, San Diego, Los Angeles, New York City, Phoenix, Dallas, San Antonio, Houston, Las Vegas, Chicago, My City, and Philadelphia. According to our analysis, the low ranking of My City can be largely attributed to is very high murder rate. This may be a good starting place in order to make My City a much safer place.

Additionally, we used the data to analyze the safety rating of each district within My City. We defined districts as the thousands and hundreds digits of the beat numbers. In this analysis and model, we took into account much more data by looking at a wider range of crimes. We also accounted for whether someone was arrested for a crime. The results of this model should give you a good idea of which districts need more focus.

In conclusion, I would recommend that you immediately begin focusing on making My City a safer place. We found that the worst districts in the city are 7, 15, and 22. In addition, our model has shown that My City and Chicago have very similar indices, and further analysis shows that individual types of crimes have similar rates in the cities. This similarity, coupled with a similar population size means that we should definitely work with Chicago to create smarter solutions to the issues we are faced with. While My City is a large international hub of commerce, technology, finance, and travel, we want to see our city become a safer place for residents and visitors alike.

Best regards,

Team 6129