**A Statistical and Geographical Analysis On the Demographics of a City**

**and Its Effects On Crime Rates On Campuses**

**Elaine Perrin, Travis Sellers, Lorena Hernandez Maxwell, Malachi Morgan, Sherod-Malik Davis, Cynthia Cook, Dr. Monica C. Jackson**

**Miami University**

**Department of Mathematics and Statistics**

**Oxford, OH, 45056, U.S.A.**

**Maryville College**

**Department of Mathematics**

**Maryville, TN, 37804, U.S.A**

**Rose-Hulman Institute of Technology**

**Department of Mathematics**

**Terre Haute, IN, 47803, U.S.A**

**Morehouse College**

**Department of Mathematics**

**Atlanta, Georgia, 30314, U.S.A**

**American University**

**Department of Mathematics and Statistics**

**Washington, DC, 20016, U.S.A.**

**Abstract**

Campus crime rates are generally lower than the national average; however thousands of crimes take place on college campuses daily. Cities that are notoriously dangerous would likely be undesirable locations for a college campus. This study examined the crime rates on campuses throughout the United States and whether or not they were significantly affected by surrounding cities. A multivariate regression analysis was performed to investigate which characteristics of a city, along with a few chosen demographics of a school, impacted the crime rate on a college campus. Our study suggested that the percentage of males of a population, and the diversity index of a city affected the number of crimes that occur on campuses. Whether a city was considered to be urban or rural played a role in determining crime rates as well. Our study also showed the higher percentage of students involved in Greek life, the more crimes were likely to occur. We also examined if a geographical pattern existed among the crime rates and among whether a city was urban or rural.

1. **Introduction**

Crime on college campuses began to receive special attention due to recent mass crime events occurring at major universities in the United States. Although campus crime rates are generally lower than the national average, they do pose a problem for academic communities which are often assumed to be safe. Crime can be classified into several categories such as theft, burglary, rape, murder, vandalism, etc. Unfortunately, these crimes can occur just as easily on campus. As a response to crime on campus, colleges and universities that receive Title IV funding are required to report their campus crime statistics to the public under the Jeanne Clery Disclosure of Campus Security Policy and Campus Crime Statistics Act. This act required schools to distribute procedures of criminal actions report, types and frequency of security policies, and crime prevention programs on an annual basis (Fisher, Hartman, Cullen, Turner, 2002/2003). The act was brought about after Jeanne Clery was found murdered in her dorm room. Her parents felt that if they had been aware of past violent-crime incidents at Lehigh University, they could have made a more informed decision about the location of Jeanne’s school (Scribner, Mason, Simonsen, Theall, Chotalia, Johnson, Schneider, DeJong).

Many studies have been explored to analyze which campuses have the most crime, but it is a strong possibility that some statistics may exude a misconception about the crime on certain campuses. Our exploration of this topic with an adequate amount of explanatory variables will further the study of crime on college campuses. Previous studies have shown with the use of some explanatory variables there have been results showing the correlation between these variables and crime. McPheters found that unemployment rates in a certain urban area affected the crime rates at colleges nearby. Robinson and Roh found that crime tended to cluster in areas. Fox and Hellman held a study analyzing the college setting including variables such as location, campus size, and scholastic quality. Their results showed that campus size and scholastic quality had a positive correlation to crime, while location had little to none.

We wish to further examine this by analyzing certain characteristics of the surrounding city and determine whether they have any correlation with crime on campus. The next sectionof our paper describes the data collected. Section III discusses our exploratory analysis and IV discusses our analysis method. Our paper is then concluded with conclusion and discussion sections.

1. **Data**

There are over 4,000 colleges and universities in the United States. We considered universities that the College Board considered large which resulted in 144 schools. All institutions were four year, non profit, private or public schools. We collected data on the cities that contained these 144 schools.

The crime rates that were used for each school were found using the Campus Safety and Security Data Analysis Cutting Tool, provided by the Office of Postsecondary Education [8-from website sources]. The crime rates were reported by each school's administrator at the end of the year 2011, which was the year with the most recent data. The number associated with each school was the number of criminal offenses which had occurred on campus divided by population per one thousand students.

The American Fact Finder available on the United States Census Bureau's website provided us with an abundance of information on American cities and its citizens. We collected data from the US Census Bureau [American Fact Finder] site and obtained the data for several of our variables including population, unemployment rate, male percentage, percentage of women in the labor force, whether the city was considered rural or urban, different race percentages, different ethnicity percentages (i.e. Hispanic or not Hispanic), median income, median age, and percentages of the different households (i.e. married couples, single parent households, or nonfamily households). The Fact Finder also provided the number of liquor stores in each county, and we mapped the city of the university to the corresponding county for this variable. The site did not provide data for Lexington, Kentucky, Athens, Georgia, and University, Mississippi. The universities in these cities were University of Kentucky, University of Georgia, and University of Mississippi, respectively, and therefore we excluded them from this analysis.

We used Simpson's Diversity Index [http://www.countrysideinfo.co.uk/simpsons.htm] as a way to measure diversity of the different races present in each city. The Simpson Diversity Index is the probability of two individuals randomly selected from a population being the same race. The numeric value of this index is found by using the following equation

,

where is the percentage of a race and is the population of the city. In our project, the index measured the variety of races in the cities. To calculate this index value, the total number of races and the percentage of each race in the cities were needed. We considered the following seven race categories: White, Black, Asian, Native American, Native Hawaiian, “other”, and those of two or more races. The index value could range from one to seven, but no city in our data set had a number that exceeded four. A higher index value represented a more diverse population. Since our values did not exceed four, this suggested the cities used were not highly diverse. We decided to use Simpson’s Diversity Index instead of just the percentage of each race for a few reasons. For one, the values were easier to compare. The index incorporated seven different numbers into one, which minimized the number of parameters we had to consider. We also wanted to avoid singling out a specific race. The Simpson’s Diversity Index created a more level ground for the comparison of the races.

Finding the percentage of students involved with Greek life was difficult to obtain. We found all of the Greek life percentages on College Board, with the exception of twenty-four schools. From there we visited the websites of these schools to see the most recent number of students involved in a sorority or fraternity and divided it by the student population. We did this for seven schools. Thirteen of the schools did not have these numbers available so we used a phone number listed to contact a member of the school’s faculty to obtain this information. Unfortunately, this process of research did not lead to Greek life percentages for the following schools: Weber State University, University of Wisconsin-Milwaukee, Utah Valley University, and City University of New York: Hunter College. College Board was also used to find the tuition rates for each university/college.

The Dru Sjodin National Sex Offender Public Website (NSOPW) is a public safety resource that provided us with access to sex offender data nationwide. NSOPW is a partnership between the U.S. Department of Justice and state, territorial, and tribal governments. We obtained the majority of the sex offenders per city using this site. After searching for the city, the website transferred us to the city government's website which contained the number of sex offenders in each city. For thirty of the 144 cities, city-data.com was used. The data were obtained from public records. After gathering these numbers, they were changed into percentages as a number of sex offenders per 1,000 people.

1. **Exploratory Analysis**

Preliminary analysis of our data were performed. The response variable, crime rates per 1,000 people, was checked for normality. Based on the histogram and Q-Q plot, a square root transformation was needed for normality of the data. We then performed the Shapiro-Wilk normality test on the data and found it to be approximately normal (*p*-value = 0.09319). It is also important to note that three cities, Cincinnati, Ohio, Adelphi, Maryland, and Trenton, New Jersey, had reported crime rates of zero. We suspected these cities would be outliers and was likely a result of an error made by someone reporting the data.

We continued by looking at each of the explanatory variables and their corresponding ranges and median. Figure (1) shows the minimum, 25% quartile, median, 75% quartile, and maximum of each of our variables. It is important to notice the ranges of some of the variables. For example, the number of liquor stores in each county ranged from zero to 1,135. Also, population size ranged from 4,005 people to 8,175,133 people. These wide ranges will later help explain our residuals plots.

**Table 1.** The following table contains the five number summary of the perspective variables used in our project.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Min | 1st Quart | Median | 3rd Quart | Max |
|  | 0 | .9869 | 1.2100 | 1.5040 | 2.2650 |
| Married House | 20.20 | 31.00 | 36.50 | 42.30 | 70.00 |
| Male House | 0 | 3.300 | 4.300 | 5.300 | 11.200 |
| Female House | 0 | 8.70 | 11.80 | 16.30 | 29.70 |
| Nonfamily | 19.50 | 39.50 | 45.00 | 53.20 | 74.00 |
| White | 10.60 | 52.23 | 69.55 | 80.60 | 93.80 |
| Black | 0.60 | 4.30 | 9.60 | 25.15 | 82.70 |
| Indian | 0.1000 | 0.3000 | 0.5000 | 0.8000 | 11.700 |
| Hawaiian | 0 | 0 | 0.1000 | 0.2000 | 2.000 |
| Asian | 1.000 | 2.925 | 4.400 | 8.400 | 39.200 |
| Other | 0.5000 | 1.700 | 3.850 | 9.800 | 41.800 |
| Two or more races | 1.100 | 2.500 | 3.150 | 4.175 | 7.100 |
| Hispanic | 2.000 | 4.875 | 10.300 | 27.280 | 92.500 |
| Non-Hispanic | 7.500 | 72.72 | 89.70 | 95.12 | 98.00 |
| Tuition | 4025 | 6617 | 8961 | 10390 | 46360 |
| Median Income | 18670 | 35970 | 44120 | 53750 | 123200 |
| Population | 4005 | 52340 | 113000 | 451700 | 8175000 |
| % Male | 0.4250 | 0.4822 | 0.4925 | 0.5018 | 0.5910 |
| % Female | 0.4090 | 0.4982 | 0.5075 | 0.5178 | 0.5750 |
| Unemployed | 0.0200 | 0.0450 | 0.0580 | 0.0720 | 0.1520 |
| % Women Working | 0.1900 | 0.4600 | 0.5100 | 0.5600 | 0.7100 |
| Median Age | 19.70 | 25.25 | 30.10 | 33.20 | 45.40 |
| Liquor stores in county | 0 | 15.0 | 41.0 | 126.8 | 1135.0 |
| Sex Offenders (Per 1000) | 0 | 0.5147 | 1.3440 | 2.2990 | 18.21 |
| % Greek Life | 0 | 0.03613 | 0.0725 | 0.1391 | 0.3420 |
| Simpson Index | 1.135 | 1.505 | 1.940 | 2.415 | 3.770 |

1. **Results**
2. **Preliminary Analysis**

Pearson’s correlation (Table 3) and scatterplots were used to assess bivariate relationships. Several variables were found to be highly correlated with one another, such as the Simpson’s Diversity Index and female households. Due to multicolinearity problems when trying to find a best fit, highly correlated variables were not included in the regression model. As stated before, the Simpson’s Diversity Index was highly correlated with female households. This led us to use married households in our model. Married households was not highly correlated with any of the other final variables in the equation.

1. **Regression**

Our first step in our multivariable regression analysis was to perform a backward elimination model using the statistical programming software R. After performing this, we found income and the percent of campus that participated in Greek life to be significant (*p*-value =0.00812 and *p*-value = 0.01719, respectively). From there, other variables were added to form equations. The final model included the following variables: percent of Greek life on campus (%Greek), married households (MarHouse), percent of population that was male (%MPop), the Simpson’s Diversity Index (SimpsDiv), whether a city was rural or urban (this was a qualitative variable, where urban was denoted by a 1 and rural was represented by a 0), and the number of liquor stores per county (LiqStor). Estimates and *p*-values for these variables can be found on Table 2. Our model is given by the following equation:

Sqrt(Crime) = β0 + β1\*(%Greek) + β2\*MarHouse + β3\*(%MPop) + β4\*Simps + β5\*Rural + β6\*Liquor + β7\*(Rural\*LiqStor) + β8\*(Rural\*Simps) + β9\*(%Greek)2 + β10\*(%MPop)2 + β11\*(Simps)2. (1)

The associated *p­*-values, standard errors, and βcoefficients for the model can be found in the table below. One asterisk represents the *p*-value was significant at the hundredth place, two asterisks represents significance at the thousandths place, and three asterisks represent significance passed the thousandths place.

**Table 2.** Coefficient estimates and associated standard error and *p*-values for equation (1).

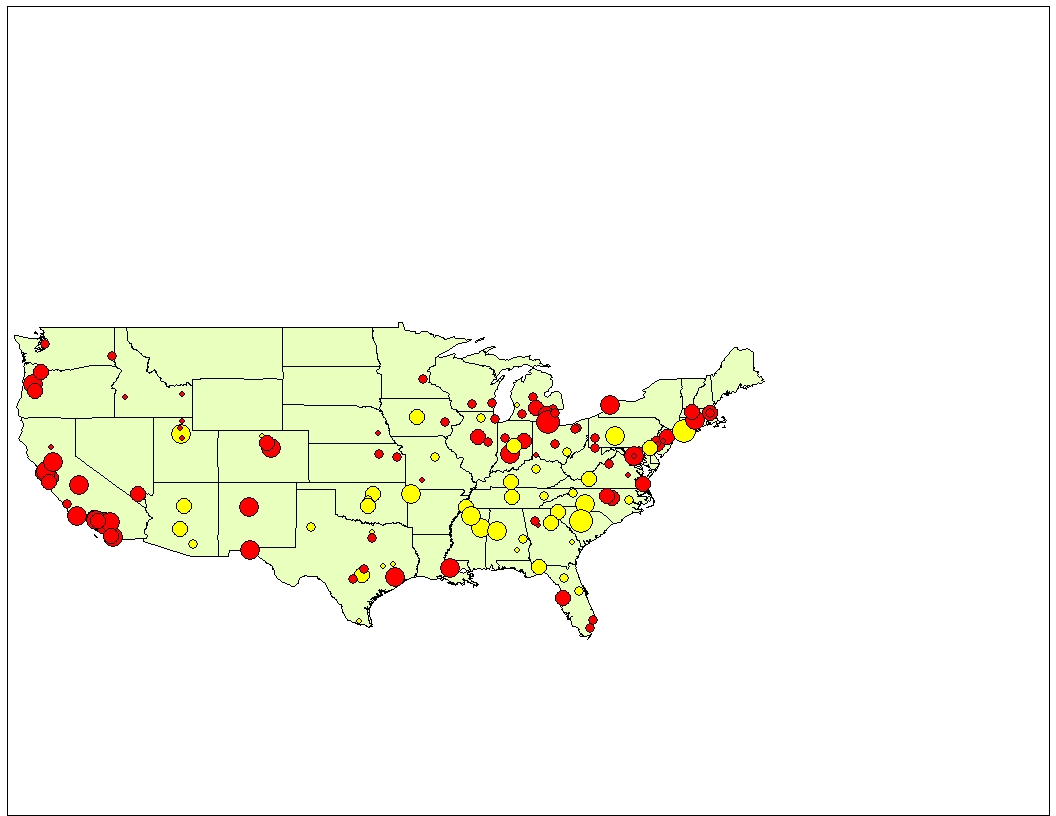
|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Estimate** | **Standard Error** | ***p*-value** |
| Intercept | -26.869231 | 11.581448 | 0.022\* |
| %Greek | 4.936018 | 1.438801 | <0.001\*\*\* |
| (%Greek)2 | -14.944523 | 5.604437 | 0.009\*\* |
| MarHouse | 0.008249 | 0.003806 | 0.032\* |
| %MPop | 99.143628 | 45.919314 | 0.033\* |
| (%MPop)2 | -94.785549 | 45.594575 | 0.040\* |
| Simps | 1.397558 | 0.341582 | <0.001\*\*\* |
| Simps2 | -0.248075 | 0.071203 | <0.001\*\*\* |
| Rural | 0.537093 | 0.272136 | 0.051 |
| Liquor | 0.003097 | 0.001220 | 0.012\* |
| Rural\*Liquor | -0.002993 | 0.001229 | 0.016\* |
| Rural\*Simps | -0.265700 | 0.141167 | 0.062 |

The *p*-value for the entire model was found to be 1.686e-05 and the adjusted R-squared value was 0.2131. All but two variables were proved to be significant with a *p*-value less than 0.05, which was our significance level. The two which were not (Rural and Rural\*Simps) had very close numbers though (0.051 and 0.062, respectively). There was a mix of positive and negative coefficient estimates. Percent of Greek life squared, percent of population that was male squared, Simpson’s Diversity Index squared, Rural times Liquor, and Rural times Simpson’s Diversity Index all had negative values. The linear terms of the squared variables were all positive. This implies there is a positive association between these variables (percent Greek, percent of population that was male, and Simpson’s Diversity Index) and the square root of crime rates. The negative sign for the squared terms suggests that although an increase in these variables increases the response variable, at some point the values will reach a threshold and the square root of crime will begin to decrease, keeping all other variables constant. Linear terms with negative coefficients (Rural\*Liquor and Rural\*Simps) imply a negative association with the square root of crime rates. Linear terms with positive coefficients denote a positive association with the square root of crime rates.

The Q-Q plot and histogram of residuals for the model were approximately normal. The plot of the fitted values versus residuals did not show evidence of non-constant variance. The residuals were evenly distributed and there were no signs of systematic patterns.

1. **Geographical Analysis**

Using the mapping software, ArcGIS, we mapped the Crime rates at all the schools and separated the schools between Rural and Urban cities (see Figure 1).



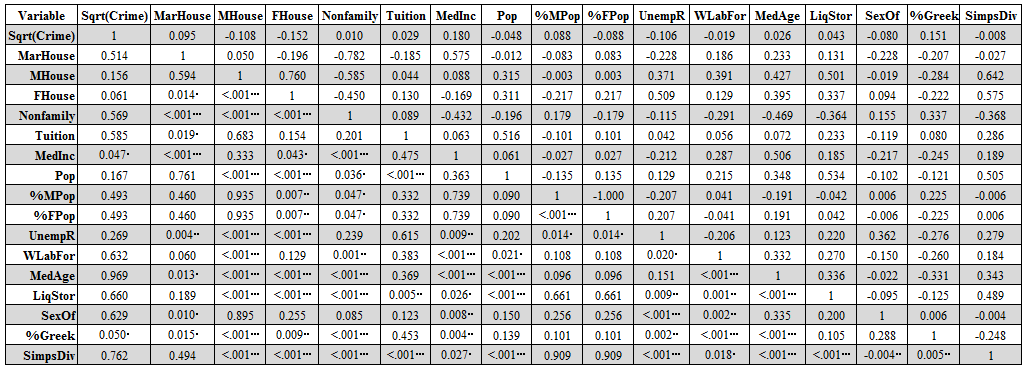
**Figure 1.** Crime Rates of the schools in Urban and Rural Areas

Figure 1 shows that most of the Institutions in Urban cities are located in western United States and that most of the Institutions in Rural cities are located in the southeast part of the United States. It was interesting to see that of the four schools in Utah the Institution in the Rural city had the highest crime rates on campus. The University of Utah had the highest crime rate while Utah State University, Weber State University, and Brigham Young University all had low crime rates and are in urban cities.

From the Model we used two interaction terms involving Rural vs. Urban. The Rural vs. Urban variable and the Simpson's Diversity Index variable as one of the interactions. Then we used the Rural vs. Urban variable and amount of Liquor stores in a county variable as the other interaction.

In order to analyze possible clusters in our data, we used SatScan and performed a purely spatial analysis scanning for clusters with high rates using the Discrete Poisson Model. In SatScan, we began by taking the largest distance between two observations and divided that distance in half and used it as the initial radius. We continued to divide the radii in half and used a radius of 445. Using this radius we found a small enough clusters to make observations. One cluster that we found had a radius of 281.60 and contained nine observations. Within this cluster, Penn State had crime rate of 3.319 which was different from the eight other observations. Another cluster with the radius of 145.66 and contained four observations. This cluster contained University of Utah which had a crime rate of 3.095 and essentially different from the other three observations. Ultimately, our observation of these clusters links with our observation that rural areas with many liquor stores have more crime.

**Table 3.** The correlation values between respective variables are found on the upper triangle. The lower triangle contains their corresponding *p*-values for the hypothesis test for statistically significant correlation. One asterisk denotes significance at the hundredth place and two asterisks represent significance at the thousandths place. Values with three asterisks were significant at a decimal place smaller than a thousandth.

****

1. **Conclusion**

The resulting model infers a positive relationship between the percent of students involved in Greek life on campus and the crime rates on campus, implying that if all other variables are kept constant a higher percent of students involved in Greek life on campus was a factor in increasing the crime rates on campus. Fox and Hellman studied the effect of percent of fraternities and sororities on crime on campus. They grouped this variable with others to test the influence of social organization on crime. The model showed significance in the effect of social organization with crime, which confirmed their thoughts that factors of social organization on campus negatively influenced the members of the community [?]. The model also showed that there is a positive relationship between the percent of married couples in a city and the crime on campus, implying that the higher the percent of married people in the city will increase the crime rates on the campus. This contradicts the results found by Bode and Snow. They found that the percent of married couples in a city decreased the amount of crime on a campus, but this could have been because of a difference in variables and data sets [?].

Along with the percent of married households, the model similarly showed that there was a positive relationship between percent of males in the city and crime on campus, meaning that the higher the percent of male population in the city the higher the crime on campus. This matched with the results from Henderson. He also found the percent of males in a city had a positive relationship with crime [?]. The model also inferred a positive relationship between the Simpson’s Diversity Index and crime on campus, implying that when all other variables are held constant, the increase in diversity of a city increases the crime in that city. The model also includes that there is a positive relationship between the amount of liquor stores in a county and crime on campus, suggesting that crime on campus occurs more often in counties that have more liquor stores. This significance matched results from Robinson and Roh.

There are also two interaction terms in the model which showed significance. The urban versus rural variable and liquor stores interaction showed a negative relationship with crime on campus, suggesting that with all variables kept constant, rural cities with more liquor stores will cause more crime on college campuses than urban cities. The other interaction is Simpson’s Diversity Index and urban versus rural variable. This term had a negative relationship with crime. This implies that rural cities with poor Simpson's Diversity Index ratings is a cause of crime on campus.

1. **Discussion**

This study analyzed the relationship between the crime rates at large universities and the characteristics of the cities that surround each school. This work checked for spatial relationships between the location of colleges and crime rates at each school and whether urban versus rural areas had an effect on crime. We also modeled the relationship between crime rates of schools and influential factors, such as the demographics of the city, tuition of each school, and the percent of students involved in Greek life.

As we noted earlier, we wanted to investigate the relationship between the type of city that surrounds each university and the crime rates at those schools. Fox and Hamilton suggested almost thirty years ago that there was no relationship between the location of a school and its crime rates. In other words, we wanted to see if there now exists a “spillover effect” from the cities into the college communities. After computing the final model, we discovered, contrary to Fox and Hellman’s findings, that the type of city that a school is nested in has a strong relationship with the crime rate at each school. However, our findings may have differed from those of Fox and Hellman since the explanatory variables they used to test the relationship of the location and crime were much different from the ones we used. Also, they tested if schools that were more urban had more crimes. Unexpectedly, they found that they did not. They concluded that more urban schools did not have significantly larger crime rates because they invested in more security measures. Furthermore, their differences in results may have occurred from the different schools they used. On the other hand, similar to our finding, they learned that the percent of students that were involved in Greek life had a strong relationship with the crime rates at each school. Their model estimated that schools with larger percentages of students involved in Greek life had higher crime rates.

On another note, a study by Sloan in 1994 suggested that the racial makeup of the student body at each school had a strong relationship with the crime rates. We wanted to apply this study to the cities that surround each school. Similarly, we found that the Simpson’s Diversity Index rating for each city had a strong relationship with the crime rate at each school.

Other differences between our research and previous studies may have resulted from the types of crimes that were considered. For example, the crimes that were used in the study by McPheters were crimes that were reported to the police on that campus. This study was performed before the Clery Act was passed. Therefore, each school was not required to report their crimes and the type of crimes that occurred on campus to the Office of Postsecondary Education. However, like McPheters’ crime rates, our crime rates were expected to have discrepancies and errors according to the Office of Postsecondary Education.

We felt these findings were pivotal in the research of factors that affected the crimes at large universities around the United States. Our hope is that future researchers can take these results and investigate them further. For example, we hope researchers may investigate characteristics of Greek life that may be expected to influence crime rates on campuses or if Greek life on campuses causes a “spillover effect” in crime rates from the cities into the college community. Also we hope that researchers may study why rural cities with more diverse populations and more liquor stores are estimated to change the crime rates at large colleges. Crime at universities has been a topic that has been researched more thoroughly since the Clery Act was passed in the early 90’s and we hope our study may inspire even more researchers to search for other factors that may play a role in the high crime rates at schools.

Unfortunately, like all research, our study has its limitations. The first one being that there may have been errors in the crimes that were reported by each school. As we noted earlier, the Office of Postsecondary Education does say there are possible errors in their reported crimes and we have reason to believe that a few schools may have false reports. However, our model did not show enough evidence to say that this was the case. Furthermore, our study only focused on criminal offenses that occurred on campuses and the rates were calculated by the total number of crimes that occurred on campus. According to Teraguchi, hate crimes have been occurring more often at schools and hate crimes were not a focus of our study. Therefore, some schools may have appeared to have a safer community, but they may have large number of hate crimes or other crimes occurring around campus that were not investigated in our study. Also, since we used the total number of crimes, some universities may appear to have a dangerous environment, but they may have only had a large amount of small crimes such as burglary. On the other hand, other institutes may appear to be safe with a smaller crime rate, but they have a larger number of homicides, forcible rapes, and aggravated assault when compared to other schools. Finally, a large constraint was that we were unaware of how much security was involved with each campus. For instance, schools that were located in more urban areas or dangerous cities may have had greater securities to prevent outsiders from disturbing the campus community. On the other hand, schools in small cities or rural areas may have had less security, allowing more outsiders to create disturbances on campus. Therefore, the type and amount of security that was available on each campus may have a large effect on the crime rates at each school.

In order to do a more comprehensive study, we would like to perform analysis on the different types of crime rates at more diverse schools. We would like to investigate what types of crimes are occurring at each school and not refer to only the total number of criminal offenses that are occurring on campus. Furthermore, we would like to perform a study that looks at the types of security that are at each institute. However, the availability of this type of data may be limited. It would also be beneficial to look at the crime rates of each city and see if they are comparable with the crime rates of the schools they surround to perform a better study of the “spillover effect”. Meanwhile, to check the validity of our model, we would like to perform the same procedures on one of the datasets that was mentioned earlier.

**Acknowledgements**

The authors would like to thank the SUMSRI program at Miami University of Ohio for funding their project.