Identifying Correlates of Homicide Rates in Michoacán, Mexico

---

A Thesis Defense

Presented to

the Faculty of the School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree

Master of Science (Systems and Information Engineering)

by

Carl M Kruger

May 2015

APPROVAL SHEET

The thesis

is submitted in partial fulfillment of the requirements

for the degree of

Master of Science

AUTHOR

The thesis has been read and approved by the examining committee:

Matthew Gerber

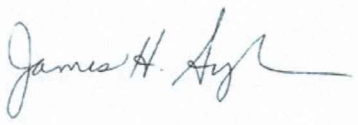
Advisor

Donald E. Brown

William T. Scherer

Kimberly A. Urban

Accepted for the School of Engineering and Applied Science:



Dean, School of Engineering and Applied Science

May

2015

Abstract

Violent crime rates in Mexico have reached pandemic levels in recent decades primarily due to corrupt and ineffective government and law enforcement officials as well as violent drug-trafficking organizations (DTOs) and other organized crime groups. The drug war is a large source of violent crimes in Mexico that has claimed over 100,000 lives, according to some estimates. This thesis will analyze social, economic, and demographic data, to determine what factors correlate with homicide rates. Previous research efforts by the Justice in Mexico Project (JMP) organization focused on understanding trends of homicide rates followed by subjective recommendations without any supporting empirical evidence. Using a large database of indicator variables, this thesis will provide a more objective explanation of high rates of homicides. The database consists of census-type data offered by the Information Bank on the *Instituto Nacional De Estradística Y Geografía* (INEGI) website. This thesis found that, using linear regression modeling techniques, a stepwise regression model performed best when estimating the homicide rates in the data. The most statistically significant indicators substantiated previous research efforts reinforcing the importance of economic and education reform as well as identifying the importance of social reform through organizations such as *Oportunidades*. The results developed in this thesis will include the identification of significant correlates of homicide rates in Michoacán, Mexico, which could assist the Mexican government in drafting and implementing reforms in education and economics aiming to decrease homicide rates.

Contents

Introduction

The United Nations Office on Drugs and Crime ranks Mexico as one of the most dangerous countries in the world. When comparing homicide rates, Mexico ranks in at number twenty-two with a homicide rate of 21.5 per 100,000 inhabitants. Putting this figure into perspective, the country with the highest homicide rate is Honduras with 90.4 homicides per 100,000 and the United States ranks 111 with 4.7 homicides per 100,000 [1]. A major source of violence in Mexico stems from the drug war. For years, Mexico has been a staging area for transporting contraband into the United States. In the 1920's, Mexican Cartels found a lucrative business smuggling alcohol into the U.S. during the Prohibition Era. Once alcohol was no longer illegal at the end of the Prohibition Era, Mexican Cartels sought other contraband to smuggle into the United States. This business evolved into what it is today, smuggling illicit drugs, illegal immigrants, and other forms of contraband into the U.S. [2]. Narcotics demand in the United States rates among the highest in the world. In 2008, 865 tons of cocaine were produced worldwide, of which 165 tons were consumed in the United States alone [3]. As drug use continues to increase in the United States, drug smuggling has no end in sight. With such a lucrative business, Mexican DTOs developed divisions and rivalries that often turn violent as they fight to gain and maintain control of prime smuggling routes into the United States. The result of this war has been a devastating wave of violence that has plagued Mexico for years.

In addition to violent DTOs, corruption in law enforcement and government officials contribute to violent crimes. DTOs bribe or coerce law enforcement and government officials to either facilitate operations or hinder those of rival DTOs. There are numerous examples of law enforcement officials incarcerating or kidnapping rival DTO members. Mexican federal investigations into police corruptions resulted in hundreds, if not thousands, of law enforcement officials at all levels being relieved of their duties due to direct ties with DTOs.

Additionally, some reports and investigations reveal elected officials are bribed by the promise of votes and power by simply turning a blind eye towards DTOs [2]. This creates a deleterious effect on the public's trust of the government. In a report by the *Comisión Nacional de los Derechoa Humanos* (CNDH), only 10% of crimes are reported and only 1% of reported crimes result in a conviction [4]. With a 99% impunity rate, criminals face a very low threat of being punished for committing their crimes and victims of crimes have a very low incentive of reporting crimes with the fear of retaliation due to the high level of corruption in the Mexican government. The level of corruption is such that Mexican citizens have little faith in their elected officials to seek justice. A side effect of these conditions has produced bands of citizens forming vigilante groups, taking matters into their own hands [5] [6].

The cumulative effects of these circumstances resulted in an increase in violent crimes to epidemic proportions in Mexico. Violence due to the drug war received international notoriety forcing the Mexican government to take a more proactive approach to curbing drug-related violence. In December 2006, President Felipe Calderon mobilized the Mexican military to engage in a major offensive with the drug cartels in the state of Michoacán. This was the start of the drug war, which continues today and has taken a devastating toll on Mexico, claiming well over 50,000 lives of cartel members, Mexican military and law enforcement, and innocent women and children [7].

Mexico's history is rife with upheaval, violence, and revolution. In 1519, Spain colonized the indigenous people of Mexico until a revolution in 1821, in which Mexico gained its independence as a monarchy known as the First Mexican Empire. This monarchy would be short-lived as it became the Republic of Mexico. Between the late 1820s to early 1850s, Mexico was led by a military man, President Antonio Lôpez de Santa Anna, who engaged with the United States in the two-year Mexican American War. Santa Anna was overthrown by Mexican Liberals in 1857 in a period known as La Reforma, in which the Mexican Constitution. This reform started a civil war between liberals and conservatives. Although the liberals successfully defeated the conservatives, this opened the door to a French invasion

in 1861 to collect on defaulted loans of the liberal government led by Benito Juárez. With the help of the French, the conservatives reestablished a monarchy, the Second Mexican Empire under Maximillian I, who would later be executed shortly after the French left in 1867. After the monarchy's collapse, Benito Juárez returned to the presidency in a period known as the Restored Republic. His predecessor, was overthrown by a liberal military man, Porfirio Diaz, who ushered a period of economic growth and stability. Diaz's time in power ended with the Mexican Revolution between 1910-1920, which established a new legal framework in the Constitution of 1917. The Mexican Revolution cost the country nearly 10% of its population. This was the last major turbulent event in Mexico's history. During and after World War II as a major ally to the United States, Mexico experienced economic growth and political stability under the Partido Revolucionario Institucional (PRI) political party [8].

In 2008, the United States, Mexico, and other Central American countries developed a security cooperation agreement in which they combined intelligence gathering and drug interdiction efforts in an attempt to combat the growing threats of drug trafficking. This agreement is known as the *Méridia Initiative*. Mexico has relied heavily on deployments of large military forces to areas plagued by drug violence. While this has been arguably effective to gain some control of those contested regions, this strategy is labor and resource intensive and cannot be sustained for prolonged periods. A more effective solution will rely on law enforcement action as a part of a holistic strategy that also includes economic, education, and diplomatic reforms in a wide-range of social factors.

This thesis employs regression techniques to identify correlating factors of homicide rates in Mexico. A literature review of previous research efforts are summarized in Section 2. The problem is formally defined in Section 3 as well as defining the thesis research objectives. The data used and analytic techniques developed in this thesis are described in Section 4. Results of all experiments and how they developed during the course of this thesis are in Section 5. Finally, Section 6 provides a summary of the results and their implications as well

as possible future work to be continued in this field.

Literature Review

2.1 Data Collection for Mexican Violence

The Justice in Mexico Project (JMP) is an organization that addresses a growing concern for security and human rights issues in Mexico due to drug violence, publishing annual reports on their trends and affects. Their objective is to use data analysis to promote policy solutions for the Mexican government to curb the effects of drug-related violence. Although the Mexican government collects information on drug-related violence, the data is not published for public consumption causing JMP to rely on media coverage from three national media outlets of drug-related homicides. Of the three media outlets, *Reforma*, is a more reliable resource due to its wide coverage with a high number of national correspondents as well as more conservative homicide counts.[1] *Reforma* uses six criteria when classifying a homicide as drug-related:

1. Use of high caliber and automatic weapons
2. Execution-style and mass casualty shootings
3. Decapitation or dismemberment of corpses
4. Indicative markings, written messages, or unusual configurations of the body
5. Presence of large quantities of illicit drugs, cash or weapons
6. Official reports explicitly indicating the involvement of organized crime.

This criteria prevents *Reforma* from conflating drug-related homicides with other homicides.

The *Reforma* data pose a few problems. While they provide sound criteria to classify a homicide as drug-related, lacking access to official reports or sensitive information that

[1] www.reforma.com

could classify a homicide as drug-related where *Reforma* would not. DTOs intimidate and assassinate journalists to censor coverage of DTO homicides. Because drug war violence received little media attention until recently, collecting data on DTO homicides did not start until November 2007. This limits analysis of the conditions in Mexico before the drug war received such attention.

JMP analyzes drug-related homicides by identifying geographic centers of violence and victim demographics. JMP draws few conclusions and recommendations for policy changes as a result of the data analysis but rather makes a subjective analysis of the increase in drug-related homicide as a result of corrupt Mexican politics. Politics in Mexico were primarily controlled by the Institutional Revolutionary Party with little competition. It was not until 2000 that other political parties developed a presence in national politics. D. A. Shirk, et. al. suggests that this change led to corruption as DTOs splintered into several smaller groups and leveraged politicians and law enforcement to target rival DTOs. While sound logic based on research and expert opinion, this conclusion is not substantiated by objective data analysis. This thesis hypothesizes an analysis of the general population during this time would provide greater insight into the causes of drug-trafficking and drug violence [9], [7], [10], [11].

2.2 Impact of Social and Environmental Factors on Crime

Most widely-used criminal forecasting objectives revolve around finding the probability of when and where a crime will occur. This provides law enforcement a tool to develop a more efficient and effective plan for operations with limited patrol resources. Alternative crime forecasting objectives analyze conditions of the environment that explain why criminal behavior occurs when and where it does. One such research effort proposed to identify the impact of foreclosures on neighborhood crime. Dan Immergluck and Geoff Smith researched the affects of abandoned and foreclosed homes on criminal activity in Chicago suburbs [12].

They wanted to measure the correlation between foreclosure rates and the levels of crime. They developed a model that relates foreclosures and other neighborhood conditions to neighborhood crime:

$$lnC_i = b_0 + b_1P_i + b_2B_i + b_3Z_i + b_4F_i + \varepsilon_i \tag{2.1}$$

The data consisted of population, P_i, and business statistics, B_i, resident characteristics, Z_i, and the foreclosure rates, F_i. Using ordinary least squares regression, the result of the research determined there was a statistically significant correlation between foreclosure rates and violent crime. Their findings led to the recommendation to develop regulations against aggressive mortgage lending, suggesting that the impact of foreclosures were not just financial but also social, as violent crimes increased in neighborhoods with high foreclosure rates [12]. Immergluck and Smith used regression techniques to show how social and environmental factors affected crime rates. Using similar regression techniques, this thesis aims to achieve the same objectives for homicide rates in Mexico.

Joel M. Caplan and Leslie W. Kennedy build on the concepts of hotspot mapping, environmental criminology, and problem-oriented policing to develop maps identifying areas of higher risk to crime based on social and environmental conditions [13]. Caplan and Kennedy summarize the steps used to develop their hotspot mapping tool, Risk Terrain Modeling (RTM). These steps require the user to select aggravating and mitigating factors subjectively to map the crime of interest, potentially developing a bias in the analysis. In a case study, they select social and environmental conditions such as bar locations, liquor store locations, gang member residence, etc. to develop statistical inferences on how these factors relate to shootings. It seems reasonable that areas in close proximity to bars, liquor stores, and gang member residence have a greater risk of shooting occurrences. While RTM serves the purpose of allowing law enforcement agencies to be proactive in their crime-prevention efforts, it restricts the analysis of the conditions to the factors that are of interest to the user instead of analyzing all factors [13]. However, selectively eliminating trivial or irrelevant variables

can help reduce the number of indicators in the data set and optimize regression techniques. This thesis will consider a selection criteria for eliminating such variables for the proposed research to reduce the number of variables after an initial assessment of all variables used in the data.

Problem and Objectives

3.1 Problem Description

The effects of DTO violence has devastated Mexico for decades. Both social scientists and law enforcement alike have examined violent criminal activity and taken steps to analyze and implement some kind of solution to minimize the affects and impacts of violent crime.

Social scientists researched in this thesis approach the problem from a human rights perspective. Aside from the staggering numbers of homicides and other violent crimes, DTOs also participate in human-trafficking, extortion, and kidnapping, as additional means to finance their illicit operations. With these other operations come very harsh realities of human rights violations to include sexual assault and rape. To compound the problem of these heinous human rights violations, corruption in the Mexico's judiciary is appalling. As noted earlier, Mexico's impunity rate is at an abysmal 1% [4]. Additionally, rampant corruption in law enforcement has developed a complete lack of trust between Mexico's citizens and those sworn to protect them. As such, social scientists focus on devising solutions in legislative reform and policy that targets corruption in law enforcement and the judiciary as well as improving overall environmental conditions in Mexico.

Meanwhile, law enforcement aim to develop more effective and efficient means of counter-narcotic operations. Since the induction of the war on drugs, Mexico has made gains in detaining high-level cartel leaders and interdicting large shipments of narcotics heading for the U.S.-Mexico boarder every year. However, their efforts do not have enough of an impact to slow the multi-billion dollar industry of the drug trade as a majority of illicit drugs eventually

Figure 3.1: Map of Michoácan, Mexico

end up in the hands of drug users all over the world [3].

The problem from a data scientist perspective is to unify the efforts of both these social scientists and law enforcement and provide a more focused effort on how best to implement a plan of action that will curb the effects of violence in Mexico. The Justice in Mexico Project (JMP) publishes annual reports on drug violence in Mexico, identifying how and where homicide rates change and linking those changes to significant events or dynamic changes in the Mexican government. JMP also analyzes the weaknesses in a variety of policy issues, concluding that reforms in certain policies will contribute to lowered homicide rates [9], [7], [10], [11]. Additionally, Sibel McGee, et. al., assessed the complex criminal networks of the drug cartels and identified failures in Mexican politics, the judicial system, economic policy, educational system, etc. Furthermore, the authors identified ways in which these failures contribute to the DTOs' successes [14]. These studies provide excellent insight into the problem of violent crimes in Mexico with similar recommendations about how to make Mexico safer. This thesis aims to substantiate these conclusions and recommendations with empirical evidence through data analysis. Although the drug war is one of the primary motivating factors and a significant source for Mexican homicides, this thesis considers homicides of all types.

Specifically, this thesis focused on the state of Michoacán, Mexico. Michoacán is a pacific-coastal state, shown in Figure 3.1, with 113 municipalities and the capital city of Morelia. On 11 December 2006, President Felipe Calderón sent 6,500 federal troops into Michoacán, initiating the war on drugs with the cartels. This initiative is known as Operation Michoacán. The *La Familia* cartel was the primary DTO that operated in Michoacán during the initiation of Operation Michoacán. After years of battling with military and law enforcement, *La Familia* suffered losses of cartel members and apprehension of some of the prime cartel leadership. What resulted was a struggle for power within the *La Familia* cartel, resulting in a division, where a new cartel, the Knights Templar, would eventually take control over Michoacán. In 2013, citizens of Michoacán took matters into their own hands as vigilante groups consisting of businessmen, farmers, and other professionals emerged to gain civil control of their area. The vigilante groups merged to form the United Command of Michoacán Self Defense Forces, claiming to have over 10,000 men armed with firearms. Michoacán continues to be relevant in the Mexican Drug War, as the Mexican military as well as emerging vigilante groups continue to engage the Knights Templar Cartel [15], [5], [6].

3.2 Objectives

The primary objective of this research is to investigate indicators that correlate with and how they affect homicide rates. As suggested by McGee et al, a holistic solution to curbing Mexican violence will include both high and low leverage interventions. Evaluating the effectiveness of the Méridia Initiative of 2008, the U.S. and Mexico rely heavily on a military and law enforcement response capability to seize control of an area determined to be heavily influenced by DTOs through counter-narcotic operations resulting in the seizure of illicit drugs and arrests of cartel members and leadership. These are low leverage interventions that, while necessary, are labor and resource intensive that cannot be sustained for long-term operations. High leverage interventions include reforms in systemic failures, such as the

failing economy and education system [14]. To summarize with an age-old adage "Give a man a fish and you feed him for a day. Teach a man to fish and you feed him for a lifetime." The current state of Mexican economy and education is in such disarray, that Mexican citizens are challenged to find jobs that provide sufficient income to meet basic needs. DTOs offer a lucrative alternative to earn income from the illicit drug trade. Policies that would bolster both the economy and education system would offer better, safer alternatives than working for DTOs. This thesis focuses on identifying these high leverage interventions.

3.3 Potential Impact

The ideal solution would present correlated indicators that the Mexican government can affect through legislative reform and policy changes. Political leaders in Mexico already developed a plan to reduce violence in Michoacán by enacting a series of programs targeting education reform, economic reform, and improvements in the judicial system, called *Plan Michoacán*. *Plan Michoacán* includes around 250 programs that target an exhaustive list of areas requiring improvement [11]. With results from this thesis and future work, a more focused effort on which areas would provide the most gain in violent crime reduction would show legislators and other government officials which programs will have the greatest affects. The techniques and procedures used in this thesis can be applied to other states in Mexico to reduce homicide rates and other crimes as well. The impact will not only affect public safety in Mexico but will also impact the public safety in the United States as the violence of the

DTOs tends to spill across the border, particularly in the border states [16].

Data and Methodology

4.1 Data

The data set used for this research comes from the Information Bank on the *Instituto Nacional De Estradística Y Geografía* (INEGI) website. INEGI publishes data from 1994-2013, consisting of 965 indicators placed into four categories:

1. Economy
2. Environment
3. Population, Households, and Housing
4. Society and Government

INEGI aggregates the data at the national, state, and municipality level. For the purposes of this thesis, the data are aggregated at the municipality level. The majority of INEGI's sources of data come from their own census surveys, but INEGI also relies on other various government agencies to compile data. Such offices as the Institute of Mexican Social Security (IMSS) and the Council of National Population (CONAPO) are examples of government agencies that provide data to INEGI.

4.1.1 Homicide Data

The response variable used for modeling and analysis in this thesis is homicide rates. The data source for homicide rates published by INEGI comes from the Office of the Attorney General of the State of Mexico. INEGI measures raw homicide counts at the municipality level, which we normalized as follows:

	2000	2001	2002	2003	2004	2005
Original INEGI Measurements	117,949					115,078
Most Recent Value Measurements	117,949	117,949	117,949	117,949	117,949	115,078
Interpolated Measurements	117,949	117,375	116,801	116,226	115,652	115,078

Table 4.1: Interpolated data of variables measured every five years.

$$r_m = \frac{c_m}{p_m/100,000} \tag{4.1}$$

where m indexes a particular municipality, r_m is the homicide rate in municipality, m, c_m is the homicide count in municipality, m, as reported by INEGI, and p_m is the population of municipality m. The homicide rate is defined as the number of homicides per 100,000 inhabitants.

4.1.2 Indicator Variables

INEGI measures most indicators annually, but some are measured in increments of five years. For the latter, this thesis treats these cases in two different methods and presents the results of both methods in different experiments. In the first method, years with no measurement assume the value of the most recent measurement. In the second method, years with no measurement assume an interpolated measurement in which the value changes by 20% of the difference between the last and next measured value provided by INEGI. For example, INEGI measures the population of Apatzingán at 117,949 in 2000 and 115,078 in 2005 with no measurements for 2001-2004. In the first method, the population of Apatzingán for years 2001-2004 assumes the value of 117,949. In the second method, The population of Apatzingán for 2001 will change by factor of 20% of the difference between 2005 and 2000, and 2002 will assume a 20% change of the difference between 2005 and 2000 from the newly interpolated value of 2001. This is repeated for years 2003 and 2004. A summary of these results can be seen in Table 4.1.

Any indicators with missing values were removed from the data set. Because data prior to 2000 contains many missing values, we focused on data from 2000-2010. Of the data in this interval, removing indicators with missing values leaves 291 indicator variables remaining for analysis. A list of these variables can be seen in Appendix A.

Because the response variable, homicide rates, is normalized to population, all indicator variables are normalized to the same metric: quantity per 100,000 inhabitants. The new variables are calculated using Equation 4.1, as described above for homicide rates.

Additionally, magnitudes of INEGI indicators vary by several orders. For example, Population is a raw count of residents that averages around 35,000; whereas, The Average Number of Occupants in Occupied Private Dwelling averages 4.5. Thus, the former indicator could dominate the effect of the latter in our statistical models. We re-normalized the rate per 100,000 inhabitants so that all observations in each indicator would have a mean of 0 and standard deviation of 1. We calculated the standard deviation for each indicator variable as follows:

$$x' = \frac{x - \mu}{\sigma} \tag{4.2}$$

where x' is the re-normalized value of x, μ is the mean of x, and σ is the standard deviation of x.

4.2 Analytic Approach and Experimental Setup

4.2.1 Linear Regression

Focusing on 2000-2010, there are only 11 observations per municipality. With 291 indicator variables, there are not enough observations to develop linear regression models. Thus, this thesis combined observations from all 113 municipalities, resulting in a data set with 1,243

observations. This thesis fit a linear regression model to the homicide and indicator variables as follows:

$$y = \beta_0 + \sum_{i=0}^{k} \beta_i x_i \tag{4.3}$$

where y is the homicide rate, β_0 is the intercept and β_i is coefficient of the indicator variable, i, and x_i is the value of the indicator variable.

Variable Inflation Factor

In order to test models for multicollinearity, this used the Variance Inflation Factor (VIF) from the Companion to Applied Regression R package [17]. Any model with a VIF over a threshold of 5 is considered to have too much multicollinearity. This thesis developed a routine that removed the variable with the maximum VIF in a stepwise manner until the VIF of the resulting model was under the threshold.

Cross-Validation

After building the linear regression model, this thesis will utilize a 10-fold cross-validation of the data to see how well the model estimates homicide rates. The Mean Square Error (MSE) of the cross-validation results will define the models' performances. Cross-validation will confirm that our linear regression models generalize effectively. This thesis utilizes a cross-validation package in R [18].

4.2.2 Principal Component Regression

Due to the large number of indicator variables present in the data set, this thesis also explores Principal Component Regression (PCR) to develop models. PCR will provide another method in which to address the potential issue of multicolinearity that could be present with as many variables as are used in this data set. The first step is to conduct Principal Component

Figure 4.1: Map of Chihuahua, Mexico

Analysis (PCA) and determine how many principal components capture 90% of the variance in the data. Because our indicator variables are normalized with a mean of 0 and standard deviation of 1, this thesis will utilize principal component analysis with covariance. Once determined, this thesis builds PCR models using the "pls" package in R and a 10-fold cross validation of the models, using MSE as the same metric of performance as the linear regression models [19].

This thesis will handle variable analysis a little differently since PCR uses principal components as weighted sums of the original variables in the models. This thesis will the same statistical significance threshold of 0.05 to determine which principal components are most statistically significant in order to analyze the variables. The variables are analyzed by taking the absolute values of the loadings of the selected, statistically significant principal components and taking the sum of those loadings. The variables with the largest sum of the loadings are determined to be most important in accounting for the most variance in the data.

4.2.3 Validation Set: Chihuahua, Mexico

Since part of the objective is to determine if the results developed in the models can be applied to other states in Mexico, this thesis selected another candidate state to act as the validation set for the regression models. Chihuahua, Mexico, was selected due to its relevance in the Mexican Drug War, see map in Figure 4.1. Chihuahua shares a substantial border with Texas and is a historically prime smuggling route for illicit drugs, undocumented aliens, and other contraband. Because of this, border towns and cities, particularly Ciudad Juárez, Chihuahua, have experienced some of the worse violence in Mexico as some of Mexico's most dangerous cartels, namely the Juárez and Sinaloa Cartels, battle for control of these strategic towns and cities.

The data set for Chihuahua, Mexico is treated in the same manner in which the Michoacán data set is treated, as described in Section 4.1. With 67 municipalities and focusing on the same period, years 2000-2010, the resulting data set consists of 737 observations with 315 indicator variables.

Results

5.1 Experiment Setup

As the models evolved based on diagnostics and changes in the data set, this thesis ran all linear regression models through the same sequence of routines. First, this thesis produces a main effects model and conducts a stepwise regression on the main effects model, comparing p-values, R^2, and Adjusted R^2 values. The p-value of the Model Utility Test determines the statistical significance of the model, using a threshold of 0.05 for a statistically significant model. Additionally, both the R^2 and adjusted R^2 values determine how well the data fit the statistical model. Next, this thesis runs both the main effects and stepwise regression models through the routine that tests for multicollinearity, as described in Section 4.2.1.

	p-value	R^2	Adjusted R^2
Main Effects Model	< 0.05	1.0	1.0
Stepwise Regression Model	< 0.05	1.0	1.0

Table 5.1: Linear regression results of homicide rate (dependent variable) against various demographic, educational, economic, and public safety variables.

Analyzing the p-values, R^2, and Adjusted R^2 values as well as the diagnostic plots for each of the resulting models, this thesis selects the best model and conducts a 10-fold Cross-Validation, analyzing the MSE. Further investigation of the model is determined by the MSE results. Finally, this thesis analyzes the p-values of each indicator variable used in the models to determine if the indicator is statistically significant using the same threshold of 0.05. Analyzing the statistically significant indicator variables will help try to draw some insight into variables that highly correlate with homicide rates.

5.2 Results

5.2.1 Linear Regression Models

This thesis developed linear regression models utilizing all 291 indicator variables and the first method of assigning values for years with no measurements in variables measured every five years, see Section 4.1.2. The stepwise regression of the main effects model, reduces the number of indicator variables to 67. Table 5.1 shows the results of linear regression on both models. Both models are statistically significant according to their p-values and the R^2 and adjusted R^2.

In the main effects model, 16 of the 291 variables are statistically significant. In the Stepwise Model, 52 of the 67 variables are statistically significant. The R^2 and Adjusted R^2 values are suspiciously high. A quick analysis of the indicator variables shows that the most statistically significant variables are other crime variables and seem to have a very dominating performance on correlating with homicide rates. In both models, the coefficients of the other crime variables (e.g., assaults, sexual offenses, burglary, etc.) were often orders of magnitude

	p-value	R^2	Adjusted R^2	Cross Validation
Main Effects Model	< 0.05	0.3863	0.3704	484
Stepwise Regression Model	< 0.05	0.4900	0.4742	425

Table 5.2: Linear regression results of homicide rate (dependent variable) with other crime rate variables and select economic variables removed.

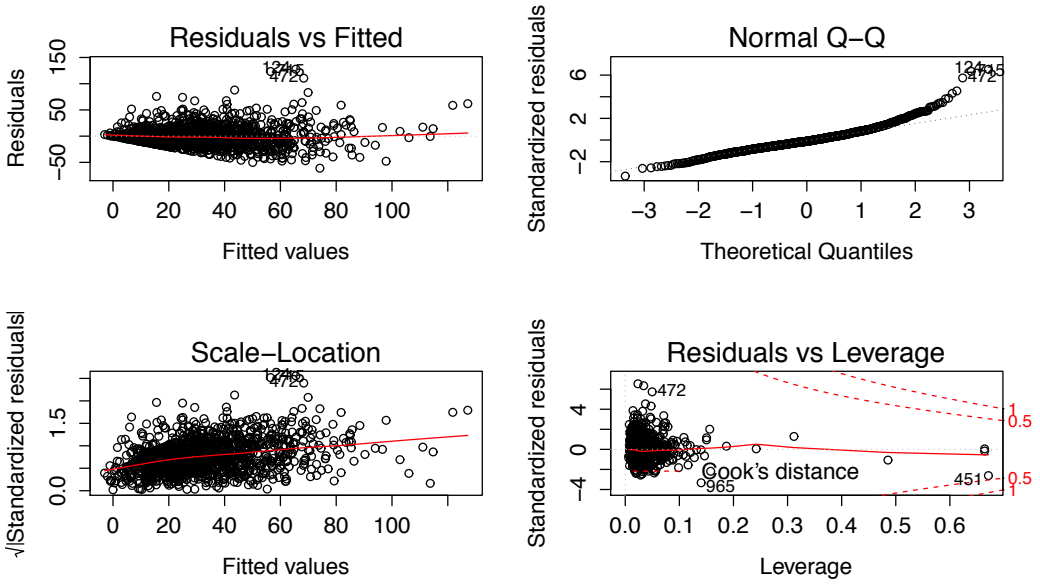


Figure 5.1: Diagnostic Plots for Stepwise Regression Model with other crime and economic variables removed.

greater than those of non-crime variables (e.g., education variables). Since our intent is to identify social and environmental correlates of homicide that might be addressed through policy changes we removed the other crime variables. While analyzing the remaining indicator variables, we noticed the majority of economic variables are oriented around publishing statistics of various economic industries in Michoacán, Mexico, and did not provide statistics of individual economic strength or poverty levels, so we removed these variables as well.

With the new indicator dataset, a stepwise regression reduced the main effects model to 38 variables. We calculated the p-values, R^2, and adjusted R^2 values to determine the model's accuracy. Table 5.2 shows these statistics for the model. The diagnostic plots in Figure 5.1 show some evidence of heteroscedasticity and non-normality in our data. This is

Municipality	MSE	Min H.R.	Mean H.R.	Max H.R.	H.R. > 100
Acuitzio	28.6	0.0	21.4	49.7	0
Ocampo	48.4	10.6	23.8	37.2	0
La Piedad	50.1	10.6	31.2	49.4	0
Hidalgo	67.2	12.2	29.2	50.8	0
Contepec	74.3	10.0	28.3	52.1	0
Aquila	1348.2	13.5	71.4	153.1	2
Taretan	1648.5	22.6	67.8	179.0	2
Nocupétaro	1773.8	13.1	85.0	128.0	5
Churintzio	2396.3	0.0	60.8	179.7	1
Coahuayana	2753.2	21.5	112.5	193.2	7

Table 5.3: MSE and Homicide Rate Statistics by Municipality

also evident in the Residuals vs. Leverage Plot as the influential points in the data consist of extremely high homicide rate observations.

We used 10-fold cross-validation to evaluate how well the model performed on different subsets of data, using Mean Squared Error (MSE) to evaluate our model's performance. The MSE was 425. In order to investigate what might account for such high MSEs, we broke down the results by municipality. The five municipalities with the highest MSEs recorded some of the highest homicide rates. The 95% confidence interval of all homicide rate observations falls between 0 and 85.4. Over 30% of all observations of the five municipalities with the highest MSE results recorded homicide rates above 100. Meanwhile, the five municipalities with the lowest MSE results remained well within this 95% confidence interval, (Table 5.3). This suggests that the model does not perform well in cases of extremely high homicide rates. In cases with more moderate homicide rates the model performs better.

As such, we developed models in which we removed all outlier observations from the data set and developed models using the same techniques. Using a box plot of all homicide rate observations, Figure B.1, we determined that the outlying observations occurred when the homicide rate was 96.8 or greater, as indicated by the plotted points above the upper whisker, removing 62 observations from the data set.

First, we developed a main effects model and ran a stepwise regression of that model. Testing for multicollinearity, we ran both models through the same routine described in

	p-value	R^2	Adjusted R^2	Cross-Validation
Main Effects Model	< 0.05	0.331	0.312	298
Stepwise Regression Model	< 0.05	0.406	0.386	268

Table 5.4: Linear regression results of homicide rate (dependent variable) with outlier observations removed.

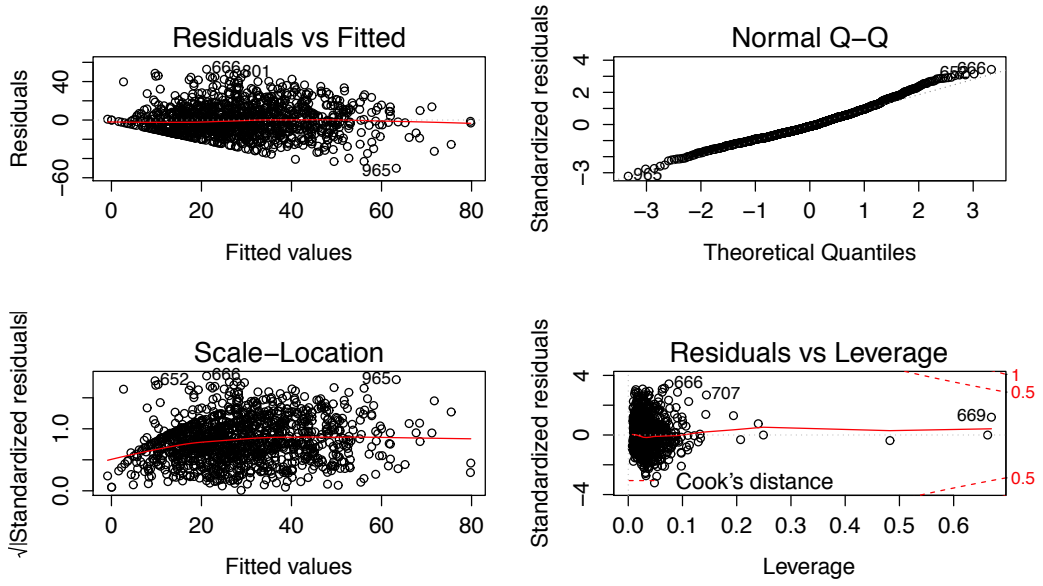


Figure 5.2: Diagnostic Plots for Stepwise Regression Model with outlier observations removed

Section 4.2.1. The resulting models were both statistically significant, see Table 5.4 for p-values, R^2, and Adjusted R^2 values. The Main Effects and Stepwise Regression models were reduced to 31 and 37 variables, respectively. A 10-fold cross-validation shows an improvement in performance compared to prior models, see Table 5.4.

Focusing our analysis of the Stepwise Regression model, having performed better than the Main Effects Model, 19 of the 37 variables were statistically significant. The diagnostics of the model are shown in Figure 5.2. While there is still evidence of heteroskedasticity, the diagnostic plots show some improvement from the previous model with the removal of the outliers, removing very influential points.

As discussed in Section 4.1.2, this thesis employed a second method of assigning values in

	p-value	R^2	Adjusted R^2	Cross-Validation
Stepwise Regression Model	< 0.05	0.502	0.469	246

Table 5.5: Linear regression results of homicide rate (dependent variable).

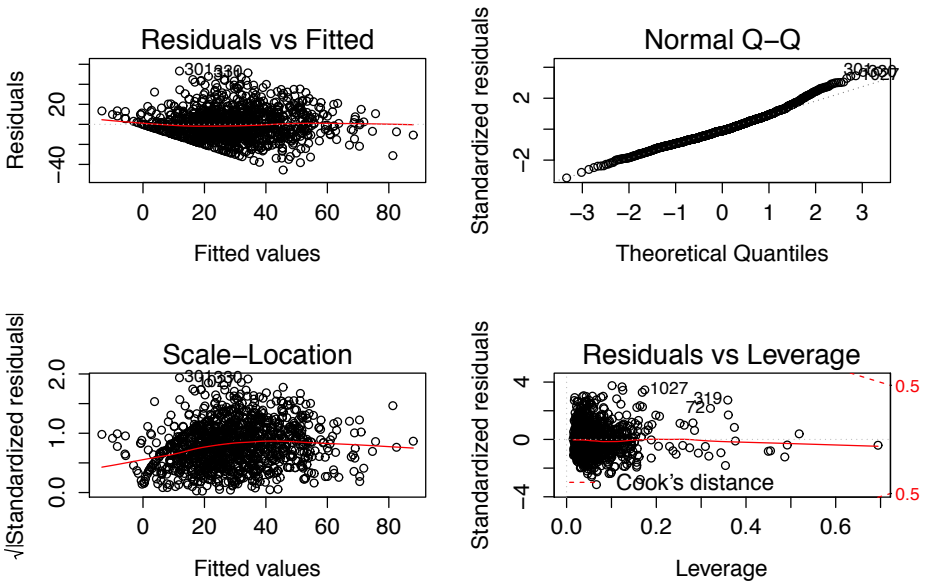


Figure 5.3: Diagnostic Plots for Stepwise Regression Model

years where no measurements are taken for those variables that are measured every five years. With this new data set was treated in the same manner as described above. The resulting model is a stepwise regression of the main effects model, see Table 5.5 for model statistics. Of the 73 variables used, 33 were statistically significant. The diagnostics of the model are shown in Figure 5.3. While there is still some heteroskedasticity present in the Residuals vs. Fitted plot, it appears to be slightly improved. All other plots have little noticeable difference betwen those in Figure 5.2. This is the best performing model with the highest R^2 and Adjusted R^2 as well as lowest MSE from cross-validation.

Table 5.6 shows a summary of the most interesting, statistically significant variables, their coefficients, and statistical significance. The total population of men being positively correlated with homicide rates could suggest that a larger male population accounts for higher homicide rates. Likewise, the number of men in the CERESO Federal Prison is negatively

Variable	Coefficient	Stat. Sig.
Total population men	4.7666	< 0.001
Men in the CERESO Federal Prison	-3.5714	< 0.001
Graduates in High School	-1.8059	0.021
Women Over 5 Years Without Schooling	3.0998	< 0.001
Average Occupants in Private Dwellings	-4.5354	< 0.001
Population in Family Homes	-3.5898	< 0.001
Beneficiary Families by the *Oportunidades* Human Development Program	3.4625	< 0.001

Table 5.6: Variable Statistics

correlated with homicide rates, suggesting that when more male criminals are incarcerated in federal prison, the homicide rates decrease. The model also identified several education variables that were statistically significant. With the number of high school graduates being positively correlated and number of women over 5 years of age without schooling being negatively correlated with homicide rates both demonstrate how a more educated population has lower homicide rates. Some interesting variables suggest how a population with strong family connections can decrease homicide rates. The average occupants in private dwellings is negatively correlated with homicide rates. These homes with a larger average occupants could represent areas with larger extended-family homes. This could suggest that these are areas with stronger family connections that could be more resistant to criminal activity. Likewise, the population in family homes also being negatively correlated with homicide rates could also suggest stronger families as well.

One of the most surprising and interesting variables is the Beneficiary Families by the *Oportunidades* Human Development Program, showing more families signed up as beneficiaries resulted in higher homicide rates. *Oportunidades* is Mexico's relative equivalent to the social welfare program in the United States. Founded in 2002, *Oportunidades* provides financial assistance to families in need, and in return, families commit to regular school attendance for school-aged children, health clinic visits, and nutritional support. While the social welfare program in the United States has been heavily criticized by conservative opponents for

	p-value	R^2	Adjusted R^2	Cross-Validation
Box Cox Step	< 0.05	0.324	0.288	10.8

Table 5.7: Linear Regression Model with Box Cox Transformation of the Response Variable

abuse of loopholes by beneficiaries and political leverage by liberally-leaning politicians, *Oportunidades* is credited with being very effective in improving economic, education, and medical conditions of areas it has provided aid. While this thesis does not draw the conclusion that increase aid by *Oportunidades* is a correlating factor of homicide rates, this is evidence of areas that are experiencing economic and social hardships, resulting in increased homicide rates. Future work in this area should investigate what effects *Oportunidades* aid has on homicide rates in subsequent years.

The stepwise regression model of the dataset with other crime variables, economic variables, and outlier observations removed is the best performing model developed in this thesis. With an MSE of 268, it is by no means an ideal model, but it does provide some interesting insight into correlating factors of homicide rates. This model will be used on the validation data set representing Chihuahua, Mexico.

Removing outliers from the dataset is not the most practical method in modeling analysis. Because this thesis aims to identify correlates of homicide rates, removing the observations with the highest homicide rates leaves out the observations from which this thesis can learn the most. The issue with the response variable, homicide rates, in linear regression is that it violates the assumption that the response variable is Gaussian. The box plot, Figure B.1, shows how homicide rates violates this assumption with outlier observations. By conducting a Box Cox transformation of the response variable, this thesis can leave those outlier observations in the data set and develop linear regression models that will provide better insight into high rates of homicides.

Figure 5.4 shows a Box Cox Plot that will determine if the response variable needs a logarithmic transformation. Based on the 95% confidence interval of λ in the Box Cox Plot, the response variable needs a logarithmic transformation. Applying the logarithmic

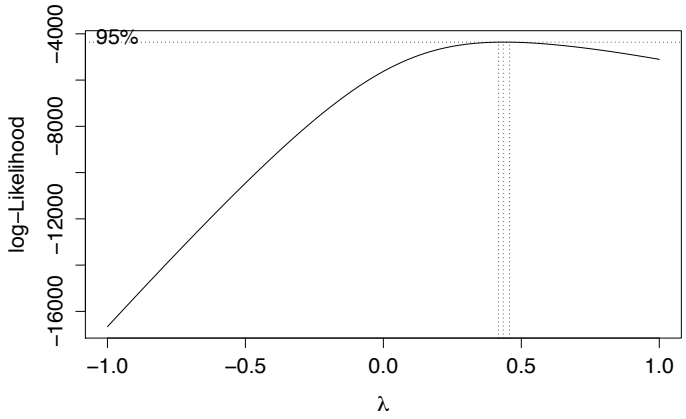


Figure 5.4: Box Cox Plot

	Min	Median	Mean	Max
Original Response	0.0	27.6	32.9	207.0
Transformed Response	-9.21	3.32	2.15	5.33

Table 5.8: Summary of Homicide Rates for Michoacan, Mexico before and after Box Cox Transformation

transformation of the response variable, this thesis first develops a stepwise regression of the main effects model. Next, this thesis runs the resulting stepwise regression model through the VIF sequence to remove the effects of multicollinearity. The statistics of this resulting model is summarized in Table 5.7 with the diagnostic plots shown in Figure 5.5. The R^2 and Adjusted R^2 are lower than the model summarized in Table 5.5. While the MSE of the cross validation is considerably lower, it is important to remember that this MSE is based on the transformed response. Table 5.8 summarizes the response variable before and after the transformation to put this number into perspective.

The diagnostics of the model, Figure 5.5, show several problems. The Residuals vs. Fitted and Scale-Location plots show a pattern indicating a lack of fit for the model. With points far from the fitted line in the Normal Q-Q Plot, this is an indication of a failure of the Gaussian assumption for the error term. Overall, this model performs very poorly compared to the previous linear regression models.

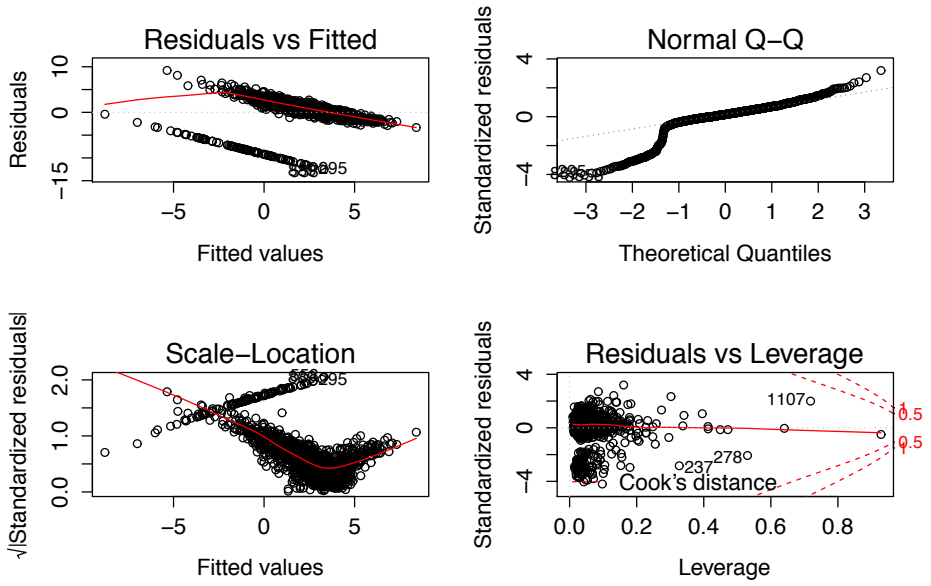


Figure 5.5: Diagnostic Plots for Stepwise Regression Model

	p-value	R^2	Adjusted R^2	Cross-Validation
PCR 90%	< 0.05	0.529	0.5	415.9
PCR 99%	< 0.05	0.628	0.574	263.1

Table 5.9: Principal Component Regression Model

5.2.2 Principal Component Regression Models

Using the same data set used in the linear regression models, this thesis also developed Principal Component Regression models. Looking at Figure C.1, the first seventy-one components capture 90% of the variance in the data. Building a PCR model with the first seventy-one principal components, the model is statistically significant. See Table 5.9 for model statistics. Although the R^2 and Adjusted R^2 are better, the cross-validation MSE is worse. Even when considering a PCR model that captures 99% of the variance of the data, requiring 157 principal components, the model's performance is comparable to the best performing linear regression models. Analyzing the statistical significance and coefficients of the principal components in the PCR model with 157 principal components, the sixth and nineteenth principal components were most statistically significant with the largest magnitude

coefficients at -3.374 and -2.814, respectively. Taking the absolute values of all loadings in these two principal components an summing them across all variables, the top twenty-five variables are summarized in Table 5.10. There are several groups of related variables. Most prominent are six education variables indicating an importance of basic education through secondary school through those various variables. There are four public safety variables that could indicate a potential connection to DTOs: three firearms-related variables and one narcotics-related variable. Finally, there are also three agriculture variables in the top twenty-five. These three groups of variables could provide some insight and emphasize the conclusions by JMP.

5.3 Validation Data Set: Chihuaha, Mexico

One objective is to see if the resulting model for Michoacán, Mexico, can be applied to other Mexican states with similar results. Since Chihuahua, Mexico, is similar in nature to Michoacán due to the high level of violence and DTO activity. This thesis treated the resulting data set for Chihuahua, Mexico, identically as the Michoacán data set. Homicide rates in Chihuahua, Mexico are generally higher than those in Michoacán, Mexico. A boxplot of Chihuahua's homicide rates is shown in Figure B.2, showing the outliers above the upper whisker. Removing outliers results in a data set includes 701 observations with 315 variables covering the same period, years 2000-2010.

This thesis developed a linear regression model using the variables from the model developed with Michoacán, Mexico, data set. There are several variables from the model that are not available in the Chihuahua data set. These variables are summarized in Table 5.11. This is more than likely due to some missing values in observations in these variables. Two of these variables were statistically significant in the Michoacán model, so this may be a factor in this model's performance.

The results of the model are summarized in Table 5.12. While the model is statistically

Variable
Agencies of the public prosecutor of the common law
Agents of the Public Ministry of the common law
Suspects in courts of first instance in criminal matters of federal jurisdiction
Persons sentenced of first instance in criminal matters of federal jurisdiction
Population in non-family households
Claimant population to health services of the ISSSTE[1]
Total teachers in basic education and higher secondary
Schools in basic education and higher secondary
Suspects for firearms crimes in the courts of the federal jurisdiction
Alleged offenders for offences relating to narcotics in the courts of the federal jurisdiction
Lodging establishments
Sources of water supply
Primary schools
The production value of the rest of the crops[2]
Sentenced for firearms crimes in the courts of the federal jurisdiction
Occupants in refuge
Occupied private dwellings that have a refrigerator
Population of 5 and more years without schooling
Men of 5 and more years without schooling
Population of 5 and more years that speak indigenous languages
The total agricultural production value
Sentenced for carrying banned weapons in the courts of the common law
The production value of pork
Schools in preschool
Total sown area

Table 5.10: Top 25 Variables in PCR model accounting for 99% of the variance in the data

significant, the R^2 and Adjusted R^2 are about the same as the Michoacán model, but the MSE for cross-validation is higher by a nearly a factor of 5 compared to the Michocán model. The diagnostic plots show significant signs of heteroskedasticity in the Residuals vs. Fitted plot. The Normal Q-Q plot suggests the data is not normally distributed, as there is a fat tail at the end of the plot. Some of the most important statistically significant variables were either not present in the Chihuahua data set (e.g. Beneficiary Families of the *Oportunidades* Human Development Program and Men in the CERESO Federal Prison) or were not statistically significant (e.g. Graduates in High School). Not all the statistically significant variables agreed between the two models. For example, Population in Family Homes is negatively

Variable
Students in preschool
Beneficiary Families by the *Oportunidades* Human Development Program
Men in the CERESO Federal Prison
Harvested Area, National Cultures
Total Harvested Area
The Total Agricultural Production Value
The Production Value of Beef
The Production Value of Pork
The Production Volume of Poultry
The Production Volume of Coal
The Production Volume of Bovine Milk
Civil Aviation Flights

Table 5.11: Variables missing from Chihuahua, Mexico data set.

	p-value	R^2	Adjusted R^2	Cross-Validation
Stepwise Regression Model	< 0.05	0.457	0.405	1190

Table 5.12: Chihuahua, Mexico linear regression results of homicide rate (dependent variable) with outlier observations removed.

correlated with the Michoacán model where it was positively correlated in the Chihuahua model. This thesis can sufficiently conclude the Michoacán model does not work well with the Chihuahua, Mexico data set.

That being said, this thesis can apply the same procedures to develop a model specific to the Chihuahua data set. The results of the models are summarized in Table 5.13. The stepwise regression model shows improvement in all model statistics and cross-validation, but the results are still not sufficient to draw substantial conclusions. The diagnostic plots show significant issues with heteroscedasticity and non-normality in the data, see Figure 5.7.

An interesting observation noted in the Chihuahua data set that was not present in the Michoacán data set is a substantial number of observations with no homicides recorded in a given year. Nearly 25% of the observations recorded a homicide rate of zero. Less than 9% of Michocán's observations recorded a zero for homicide rates. Removing all observations with zero homicide rates and applying the same procedures, the resulting model statistics are summarized in Table 5.14. There is very little improvement in the model statistics and

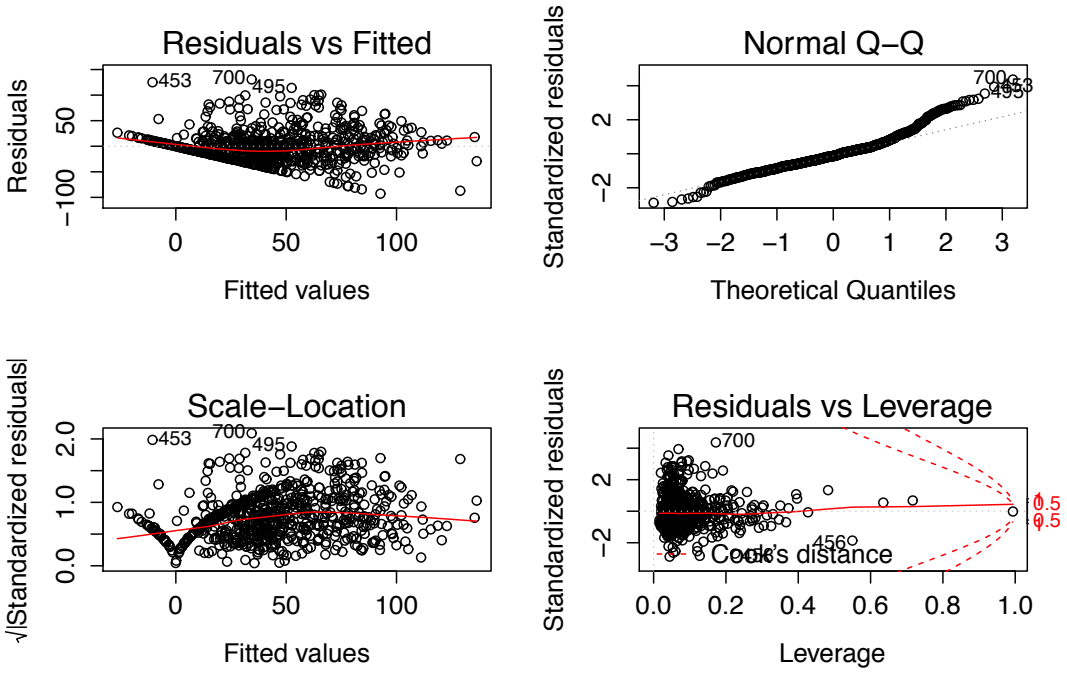


Figure 5.6: Diagnostic Plots for Chihuahua, Mexico Stepwise Regression Model with outlier observations removed

	p-value	R^2	Adjusted R^2	Cross-Validation
Main Effects Model	< 0.05	0.244	0.221	1444
Stepwise Regression Model	< 0.05	0.451	0.385	1373

Table 5.13: Chihuahua, Mexico linear regression results of homicide rate (dependent variable) with outlier observations removed.

cross-validation is worse. Overall, procedures and techniques developed in this thesis do not

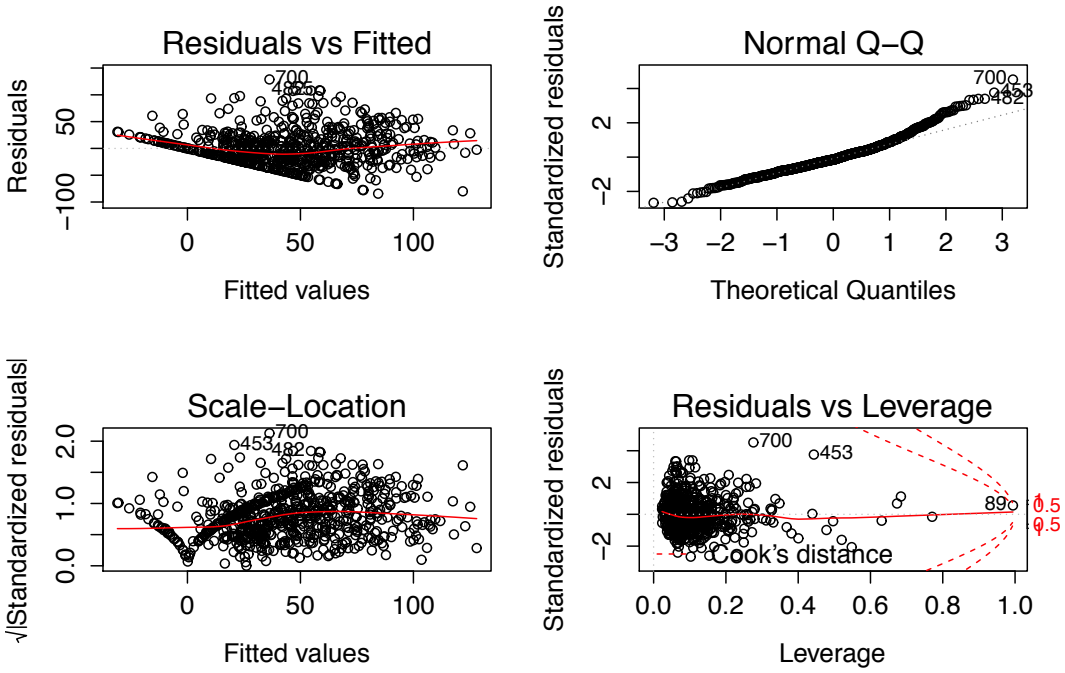


Figure 5.7: Diagnostic Plots for Chihuahua, Mexico Stepwise Regression Model with outlier observations removed

	p-value	R^2	Adjusted R^2	Cross-Validation
Main Effects Model	< 0.05	0.195	0.182	2487
Stepwise Regression Model	< 0.05	0.542	0.461	6338

Table 5.14: Chihuahua, Mexico linear regression results of homicide rate (dependent variable) with homicide rates observations of zero removed.

provide a viable solution with the Chihuahua data set.

Conclusions, Limitations, and Future Work

6.1 Conclusions

Before discussing the conclusions drawn from the models analyzed in the results section, it is important to note that this thesis is unable to generalize the results and the conclusions

drawn from them with any confidence due to serious limitations in the data, discussed below. The results of these models offer some potentially useful insight and provide a foundation upon which to conduct further analysis using other data analytic techniques as it applies to the issue of violence in Mexico. These results may also provide other social scientists evidence to research this problem and find a better understanding of the underlying driving forces behind human behavior as it relates to violent crime in Mexico.

6.1.1 Substantiating Previous Research

Our research corroborates research done by McGee, et al., which posits that economic and education reform are paramount in developing a population that will resist the allure of criminal business [14]. The data shows that observations with a higher population with a high school education correlates with lower homicide rates, as is evident with the variable Graduates in High School, being negatively correlated with homicide rates. Additionally, the model suggests that female basic education will have an impact on homicide rates as well. This is evident by the variable Women Over 5 Years without Schooling, which is positively correlated with homicide rates. Education reform could be focused on improving high school education through quality and quantity of graduates as well as the retention rate of students.

Economically, the number of Beneficiary Families by the *Oportunidades* Human Development Program is positively correlated with homicide rates. As the principal anti-poverty program in Mexico, *Oportunidades* dedicates financial aid to poor families in both urban and rural areas. This variable suggests that observations with higher numbers of beneficiary families are living in areas that are experiencing wide-spread economic hardships. With these poor economic conditions, these areas have increased homicide rates. Economic reform could focus on improving conditions in areas with large numbers of beneficiary families in *Oportunidades*.

6.1.2 New Insight

Neither McGee, et al., nor JMP address the male population as a primary concern for criminal behavior. This is evident in two statistically significant variables highlighted from the model. First, the total male population is positively correlated with homicide rates. Based on prison statistics for Michoacán, Mexico, over 85% of inmates are male.[1] This suggests criminal activity is predominantly committed by male citizens. Second, the number of men in in the CERESO Federal Prison is negatively correlated with homicide rates. As discussed in Section 1, corruption in the government is rampant, including the judicial system [4]. In observations where more men are incarcerated in the CERESO prison system displayed lower homicide rates. This could be indicative of areas with a more effective law enforcement and judicial system that arrests and incarcerates men who would otherwise commit more heinous crimes like homicides. Analyzing these two variables should provide some incentive to investigate further root causes of these crimes and develop solutions to prevent men from committing crimes.

The variable this thesis found most interesting and not previously addressed in other research efforts is the impact on social welfare programs such as *Oportunidades*. The number of beneficiary families of the *Oportunidades* Human Development Program is positively correlated with homicide rates. Initially, this thesis postulated that these resources were being exerted for the wrong reasons and to the wrong people. However, it makes sense that *Oportunidades* would focus more of their efforts to more troubled regions throughout Mexico. *Oportunidades* receives high accolades for being very effective at providing financial aid for the purpose of improving not only economic conditions but also education and health care in regions that receive aid. They are effective because they place conditions on receiving financial aid. These conditions are focused on improving conditions for children in regards to education and health care. In order to maintain beneficiary status through *Oportunidades*, parents must send their kids to school with regular attendance, health checkups, and attend informative

[1] These statistics provided by INEGI Bank of Information for 2009-2011.

talks at health centers. Ceilings are placed on the amount of financial aid received to prevent families from abusing the system by having more children for the purpose of receiving more financial aid. *Oportunidades's* financial assistance results in stronger, healthier children and a higher retention rate at all levels of education. The possibility of corruption is greatly reduced, as parents received financial aid directly from the central government [20]. This is far different from the social welfare program in the United States, where programs are run at state and federal levels of government. Critics of the welfare programs suggest this opens the potential of corruption and abuse of programs by both beneficiaries and administrators alike. The concept of social welfare programs such as *Oportunidades* offers new insight into developing solutions to reduce homicide rates and certainly requires more investigation.

The concept of the effects of the male population and social welfare programs supports the notion posited by McGee et al., in which low leverage interventions by law enforcement need to be complimented by high leverage interventions in social and economic reform. In Fiscal Year 2014, the United States Customs and Border Protection Agency (CBP) takes credit for seizing 2,366,639 pounds of illicit narcotics with a street value of $25,087,159 along the southwest border in the states of Texas, New Mexico, Arizona and California. CBP also seized more than $237 million in unreported currency related to the drug-trade. Additionally, aircrews were able to detect and interdict smuggling attempts that resulted in the seizure of 112,224 pounds of cocaine with an estimated street value of $8.4 billion [21]. These staggering statistics of drug seizures and the monetary loss to DTOs are but a fraction of the overall drug industry, valued at over $143 billion by some estimates [3]. These are low-leverage interventions according to McGee et al. These interventions are expensive, time-consuming, resource intensive, and, in the best cases, only temporarily disrupt drug trade operations. More organizations such as *Oportunidades* could be critical in providing this needed social reform and provide the necessary high-leverage interventions to compliment the efforts of military and law enforcement to combat DTO operations. Investing into more social reform programs such as *Oportunidades* requires little effort when compared to counter-narcotic

operations. These high-leverage interventions operate at a slower-rate, but the long-term effects have the potential to greatly improve the situation. *Oportunidades* offers aid in areas struggling economically and improve the social environment by keeping children in school longer and increasing their earning potential in the economy. Effects of low-leverage interventions with complimenting high-leverage interventions will have the greatest overall impact.

6.2 Limitations

While plenty of research efforts have been dedicated to qualitative analysis of Mexico's culture of violent crime, this thesis is a preliminary effort to providing empirical evidence and analysis. Logical analysis would conclude that Mexico's struggling education system and economy makes a population vulnerable to coercion at that hands of the DTOs. There are many factors determined by social scientists that have a significant impact on violent crime in Mexico that are not available in the data set used in this thesis nor are they measured and recorded in a way that would be useful based on the techniques used in this thesis. An example of this is the number of fatherless families in Mexico. INEGI does not provide records that convey this statistic but would otherwise provide some potentially revealing correlation with violent crime. This is only one example of other socio-cultural factors that are not recorded in this data set. The methods and techniques employed in this thesis provide foundation on which to develop better data sets and apply more robust techniques for data analysis as it applies to identifying correlates of homicide rates in Mexico.

The models developed and analyzed provide some interesting insight, but the performance of these models is less than ideal. The data set poses some potential issues that could offer significant improvement in the current models. First, corruption and coercion limit reporting of important data in this report. Most notable of these under-reported variables are crime statistics including the response variable used in this thesis, homicide rates. The Mexican

government no longer publishes crime-related deaths. It is uncertain if this is due to pressure internal to the Mexican government or if this is a result of coercion from DTOs [22] [23]. Recorded measurements of homicides on the INEGI Bank of Information end in 2010. Missing four years of the most recent data presents some issues being able to analyze the current level of violence in Mexico. Next, the data set lacks some potentially revealing variables that would provide more substantial evidence of correlating variables with homicide rates. For example, the data set lacks individual economic statistics, such as poverty levels, income levels, etc. Variables such as these are measured at the national level but not at the state and municipality level. Because variables with missing values were removed from the data set, this also limits the analysis of some potentially revealing variables. Leaving these variables in the data set would not allow this thesis to compute the linear regression models. This is a significant limitation in the data set. Finally, the limited number of observations per municipality severely limits the ability to conduct a temporal analysis of the data. This thesis proposes some potential solutions to overcome this challenge in order to incorporate a temporal analysis for this problem.

6.3 Future Work

While this thesis focuses on all homicides, the primary motivation is to investigate DTO violence in Mexico. The current data available does not aggregate DTO-related homicides at the municipality level. Additionally, the DTO-related homicides rely on a journalist tally based on an unofficial criteria to classify a homicide as DTO-related. As discussed in Section 2.1, the most reliable of three Mexican media outlets that keep a tally of DTO-related homicides is *Reforma*. While *Reforma* uses sound criteria to classify DTO-related homicides, this is not an official count and only covers journalistic reporting of DTO-related homicides dating back to 2006. The Mexican Government maintains a count of DTO-related homicides that has not been made publicly available in recent years [22]. The Mexican government

declared that classifying a homicide as DTO-related or not was ineffective in determining why victims were murdered. The potential exists to extend work with United States and Mexican federal law enforcement agencies. Since the *Méridia Initiative*, the United States and Mexico have increased cooperation as a shared responsibility to combat the illicit drug trade and DTO violence. Obtaining data that classifies homicides as DTO-related homicides could provide some important insight for the purposes of this research.

The insatiable demand for narcotics in the United States is a major problem fueling DTO operations. Incorporating this concept into modeling and analysis could help reveal some information not previously explored. High-leverage interventions could be applied to the demand side of this business with the potential to reduce the effectiveness and lucrative business of the drug trade. It would also be interesting to see how DTO operations have evolved in recent years as some states in the U.S. have begun to legalize recreational use of marijuana. DTOs started smuggling alcohol during the Prohibition Era as a very lucrative business until Prohibition ended. Should legalization of marijuana continue at the current rate in the United States, it will be interesting to see how this affects DTO operations. In any circumstance, the demand for drugs in the United States is a major factor in the drug war that needs to be considered in future work.

Another major aspect lacking in this thesis is temporal analysis. Due to limitations in how many observations available in the data used in this thesis, temporal analysis was not a viable option to pursue. Should such data be available at monthly or quarterly intervals instead of annually, future work may be able to conduct temporal analysis. The scope of the research may have to be narrowed as well. While this thesis relies on census data measured annually or less frequently, future work could focus on a single municipality or city, collecting measurements through surveys and and other methods with a focus on more specific indicator variables. The ground work established in this thesis could provide a starting point for determining what indicator variables to measure. Not only could this reveal social and environmental factors correlated with DTO-related homicide rates but could also identify

seasonal variations in DTO activities. For example, the Taliban fighting season in Afghanistan is directly related to the harvesting season for the poppy plants, as they rely heavily on income from opium production. Temporal analysis could reveal similar revelations as DTO activities also rely on an illicit cash crop, marijuana. These revelations could provide more focus for drug interdiction operations by Mexican military and law enforcement. Already a resource-intensive operation, law enforcement are constantly trying to get a step ahead of the DTOs through more effective counter-narcotic operations.

Bibliography

[1] Global Study on Homicide 2013: Trends, Contexts, Data. Technical report, United Nations Office on Drugs and Crime, Vienna, 2013.

[2] Ed Vulliamy. *Amexica: War Along the Borderline*. Farrar, New York, 2010.

[3] A. Buggle. After 40-Year Fight, Illicit Drug Use at All-Time High. July 2013.

[4] Liliana Alcantara. CNDH ve impunidad en 99[Online; posted 15-December-2008].

[5] N. P. Flannery. Mexico Media Roundup: Another Month of Violence And Vigilantes In Michoacan. March 2014.

[6] P. Villegas. Mexico Targets Gang That Infiltrated the Mining Industry. March 2014.

[7] et al. C. Molzahn. Drug Violence in Mexico: Data and Analysis through 2011. *Justice in Mexico Project, Trans-Border Institute, Joan B. Kroc School of Peace Studies*, 2012.

[8] B. Kirkwood. *History of Mexico*. Greenwood Press, Westport, Connecticut, March 2000.

[9] D. A. Shirk. Drug Violence in Mexico: Data and Analysis from 2001-2009. *Justice in Mexico Project, Trans-Border Institute, Joan B. Kroc School of Peace Studies*.

[10] et al. C. Molzahn. Drug Violence in Mexico: Data and Analysis through 2012. *Justice in Mexico Project, Trans-Border Institute, Joan B. Kroc School of Peace Studies*, 2013.

[11] et al. K. Heinle. Drug Violence in Mexico: Data and Analysis through 2013. *Justice in Mexico Project, Trans-Border Institute, Joan B. Kroc School of Peace Studies*.

[12] D. Immergluck and G. Smith. The Impact of Single-Family Mortgage Foreclosure on Neighborhood Crime. *Housing Studies*, 21(6):851–866, 2006.

[13] Joel. M. Caplan and Leslie W. Kennedy (Eds.). Risk Terrain Modeling Compendium. *Rutgers Center on Public Security*, 2011.

[14] et al. S. McGee. Mexico's Cartel Problem: A Systems Thinking Perspective. *Applied Systems Thinking Institute, Analytic Services, Inc.*, 2011.

[15] J Villamil. The War in Michoacan: A Brief Chronology. January 2014.

[16] S. Longmire. *Cartel: The Coming Invasion of Mexico's Drug Wars*. MacMillan, New York, 2011.

[17] John Fox and Sanford Weisberg. *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, second edition, 2011.

[18] John H. Maindonald and W. John Braun. *DAAG: Data Analysis And Graphics data and functions*, 2014. R package version 1.20.

[19] M. Bjørn-Helge et al. *pls: Partial Least Squares and Principal Component Regression*, 2013. R package version 2.4-3.

[20] Paying for Better Parenting. October 2006.

[21] CBP Border Security Report: Fiscal Year 2014. Technical report, Department of Homeland Security, U.S. Customs and Border Protection, 2014.

[22] Elyssa Pachico. Mexican Government Stops Publishing Data on Crime-Related Deaths, August 2012. [Online; posted 12-August-2012].

[23] Tracy Wilkinson. Sinaloa, one of Mexico's Most Violent States, Limits Crime Coverage, August 2014. [Online; posted 1-August-2014].

Appendices

List of Variables

The tables below list all variables used for modeling and analysis for both candidate states.

Road traffic in urban area
Fatal traffic accidents
Non-fatal traffic accidents
Transit damage-only accidents
Road traffic in urban and suburban areas
Land traffic accidents by collision with fixed object
Land traffic accidents by collision with pedestrians
Land traffic accidents by collision of vehicle
Accidents where the driver responsible is man
Accidents where the responsible driver is female
Accidents where the liable driver escaped
Total collection of fixed assets
Adults in primary education for adults
Adults in secondary education for adults
Airports
Travel agencies
Agencies of the public prosecutor of the common law
Agencies of the Federal Public Ministry
Agents of the Public Ministry of the common law
Agents of the Public Ministry in the Federal jurisdiction
Students in preschool
Pupils in primary
Students in high school
Students in special education
Preschool graduates
Elementary school graduates
High school graduates
Students stock in preschool
Students in primary stock
Students in secondary stocks
Trees planted
New cars sold to the public
Cars registered in circulation
Passengers reported in circulation trucks
New trucks sold to the public
Commercial trucks and vans in circulation
Capacity of the Social rehabilitation centers
Installed capacity of the water treatment plants in operation
Digital community centers e-Mexico

Table A.1: All Variables

Labor disputes
Resolved labor disputes
External consultations in the IMSS
External consultations in other institutions
External consultations issued by the popular insurance
Registered rooms lodging
Infant deaths
Infanty deaths, unspecified sex
Infant deaths, male
Infant deaths, female
General deaths
General deaths, unspecified sex
General deaths, men
General deaths, women
Complaints received in environmental matters
Divorces
Median age
Middle-aged men
Middle-aged women
Graduates in Bachelor of open system
Gross expenditures of municipalities
Expenditures of municipalities. Final availability
Expenditures of municipalities. Public investment
Expenditures of municipalities. Other applications
Expenditures of municipalities. Personal services
Net disbursements of the municipalities
Preschools
Primary schools
Secondary schools
Lodging establishments
Preparation and service of food and beverages establishments
Beneficiary families by the Oportunidades human development program
Sources of water supply
Average grade of schooling of population 15 years and over
Homes
Households with female leadership
Households with male headship
Men of 5 and more years with no specified instruction
Men of 5 and more years with post-primary instruction

Table A.2: All Variables (cont.)

Men of 5 and more years with preschool
Men of 5 and more years with primary
Men of 5 and more years attending school
Men of 5 and more years not attending school
Men of 5 and more years that do not specify if they attend school
Men of 5 and more years without schooling
Men of 6 and more years that do not specify whether they are literate
Men of 6 and more years who are illiterate
6 and more men who are literate
Internal men of the common law in the CERESO
Internal men of the federal jurisdiction in the CERESO
Strikes exploded
Solved strikes
Use in primary index
Use secondary index
Retention in primary index
Retention in school
Revenues of municipalities
Income for municipalities. Federal and State contributions
Income for municipalities. Initial availability
Income for municipalities. Taxes
Income for municipalities. Other sources
Income for municipalities. Federal and State shares
Net income of the municipalities
Internal of the common law in the CERESO
Length of the federal highway network of quota with State administration
Length of the federal highway network of federal administration fee
Length of the federal highway network of quota with particular management
Marriages
Motorcycles registered in circulation
Women of 5 and more years with no specified instruction
Women of 5 and more years with post-primary instruction
Women of 5 and more years with preschool
Women of 5 and more years with primary
Women of 5 and more years attending school
Women of 5 and more years that do not specify if they attend school
Women of 5 and more years without schooling
Women of 6 and more years that do not specify whether they are literate
Women of 6 and more years who are illiterate

Table A.3: All Variables (cont.)

Women of 6 and more years who are literate
Births
Births, unspecified sex
Births, men
Births, women
Occupants in detached house
Apartment building occupants
Occupants in premises not built for room
Occupants in refuge
Mobile home occupants
Occupants in house or room in roof
Occupants in house or room in vicinity
Occupiers in private homes
Occupiers in private homes of unspecified class
The Telegraph Network offices
Other terrestrial traffic accidents in urban and suburban areas
Teachers in special education
Teachers in preschool
Teaching staff in primary
Teaching staff in secondary
Total employed personnel per economic unit.
Total employed personnel.
Paid staff.
Water purification plants in operation
Population 18 years and over with professional level
Population 18 years and over with graduate
Population of 5 and more years
Population of 5 and more years with no specified instruction
Population of 5 and more years with post-primary instruction
Population of 5 and more years with preschool
Population of 5 and more years with primary
Population of 5 and more years attending school
Population of 5 and more years that speaks indigenous languages
Population of 5 and more years do not attend school
Population of 5 and more years not specified if he attends school
Population of 5 and more years not specified if it speaks the indigenous language
Population of 5 and more years that speaks indigenous languages
Population of 5 and more years without schooling
Population of 6 and more years
Population of 6 and more years not specified if it can read and write
Public institutions of social security claimant population
Popular insurance claimant population
Claimant population to health services
Claimant population to health services of IMSS

Table A.4: All Variables (cont.)

Claimant population to health services of the ISSSTE
Claimant population to health services in another institution
Claimant population to PEMEX health services
Claimant population health men services
Claimant population to women health care
Population in households
Population in households of type not specified
Population in family homes
Population in non-family households
Population which did not specify its status as rightful claimant
Total population
Total population of 6 and more years who are illiterate
Total population of 6 and over who are literate
Total population men
Total population women
Percentage of fatal road traffic accidents
Percentage of area covered by woods and forests
Percentage of population from 15 to 29 years
Percentage of population 15 to 29 years old men
Percentage of population 15 to 29 years old women
Percentage of the population 60 and over years
Percentage of the population 60 and over years men
Percentage of the population 60 and over years women
Population without title to health services
Criminal suspects by damage to the things reported in the courts of the common law
Alleged offenders for offences relating to narcotics in the courts of the federal jurisdiction
Suspects for crimes of tax fraud in the courts of the federal jurisdiction
Suspects for crimes of firearms in the Federal law of the federal jurisdiction
Suspects for crimes provided in the General Law of population of the federal jurisdiction
Alleged criminals by looting reported in the courts of the common law
Suspects for fraud in the courts of the common law
Suspects for killing reported in the courts of the common law
Suspects for violating family obligations recorded in the courts of the common law
Suspects for injuries reported in the courts of the common law
Suspects for other crimes reported in the courts of the common law
Suspects for other crimes reported in the courts of the federal jurisdiction
Suspects for carrying banned weapons reported in the courts of the common law
Suspects for robbery reported in the courts of the common law
Suspects for violation in the courts of the common law
Suspects in courts of first instance of the common law
Suspects in courts of first instance in criminal matters of federal jurisdiction
Total gross output by total employed personnel.
Total gross output per economic unit.
Total gross output.
Average of children born alive of the women of 12 years and over
Average of occupants in occupied private dwellings
Gender relationship
Total remuneration per occupied person paid.

Table A.5: All Variables (cont.)

Sentenced for burglary reported in the courts of the common law
Sentenced by damage to the things reported in the courts of the common law
Sentenced for offences relating to narcotics registered in the courts of the federal jurisdiction
Sentenced for crimes of tax fraud registered in the courts of the federal jurisdiction
Sentenced for crimes of firearms provided in the Federal law of the federal jurisdiction
Sentenced for crimes provided in the General Law of population registered of the federal jurisdiction
Sentenced for looting reported in the courts of the common law
Sentenced for concealment reported in the courts of the common law
Sentenced for killing reported in the courts of the common law
Sentenced for injuries reported in the courts of the common law
Sentenced for other offences registered in the courts of the common law
Sentenced for other offences registered in the courts of the federal jurisdiction
Sentenced for carrying banned weapons reported in the courts of the common law
Sentenced for theft reported in the courts of the common law
Sentenced for rape reported in the courts of the common law
Sentenced in the courts of first instance in criminal matters of the common law
Reported in the Court of first instance in criminal matters of federal jurisdiction sentenced
Harvested area from the rest of national cultures
Total harvested area
Reforested
Area sown to grain corn
Area sown to the rest of national cultures
Total sown area
Average size of household
Average size of households with boss man
Average size of households with boss woman
Literacy rate of 15-24 years women
Literacy rate of persons aged 15 to 24 years
Literacy rate of 15-24 year old men
People with conviction rate
Telegrams transmitted
Total students approved in basic education and higher secondary
Total graduates in basic education and higher secondary students
In basic education and higher secondary schools
Total stock of pupils in basic education and higher secondary
Total expenditure for consumption of goods and services.
Total income by providing goods and services.
Total dependent of the company staff.
Total of teachers in basic education and higher secondary
Total non-dependent of the company staff.
Total occupied individual dwellings
Total remuneration.
Urban permanent and casual workers affiliated to IMSS
Economic units.
In the IMSS medical units
Medical units in other institutions
Value added gross census by total employed personnel.

Table A.6: All Variables (cont.)

Census gross value added.
The total agricultural production value
The production value of beef carcass
Value of production of beef carcass of galliformes
The production value of carcass of pigs
The production value of total carcass
The production value of maize grain
The production value of the rest of the crops
Registered in-service motor vehicles (excluding motorcycles)
Victims injured in traffic accidents
Victims killed in traffic accidents
Occupied private dwellings
Private homes inhabited with 1 to 4 occupants
Private homes inhabited with 5 to 8 occupants
Private homes inhabited with 9 and more occupants
Private homes inhabited with different ground floor
Occupied private dwellings which have water from the public network in the field of housing
Occupied private dwellings that have a computer
Occupied private dwellings are equipped with drainage
Occupied private dwellings that have electric power
Occupied private dwellings which have excused or health
Occupied private dwellings have washing machine
Occupied private dwellings that have a refrigerator
Occupied private dwellings that have a television
The production volume of beef carcass
The production volume of beef carcass of galliformes
The production volume of carcass of pigs
The production volume of coke
The production volume of bovine milk
The production volume of corn grain
The production volume of iron pellets
Annual delivered volume of drinking water
Civil aviation flights

Table A.7: All Variables (cont.)

Box Plots

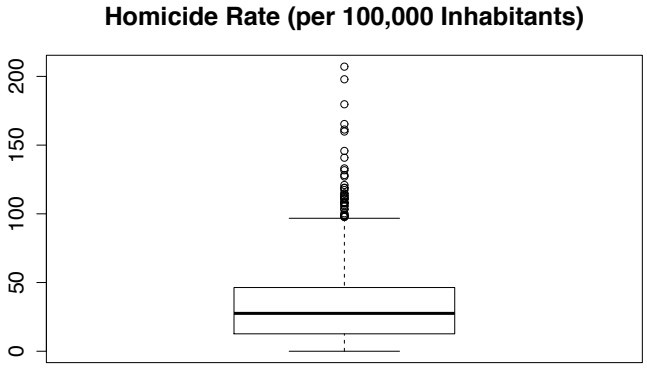


Figure B.1: Box Plot of Homicide Rates Michoacán

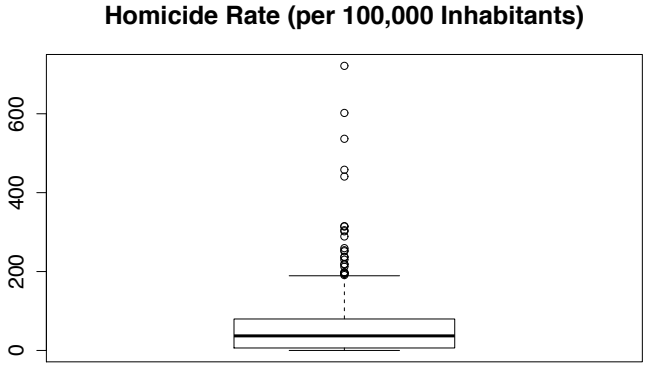


Figure B.2: Box Plot of Homicide Rates Chihuahua, Mexico

Principal Component Regression Plots

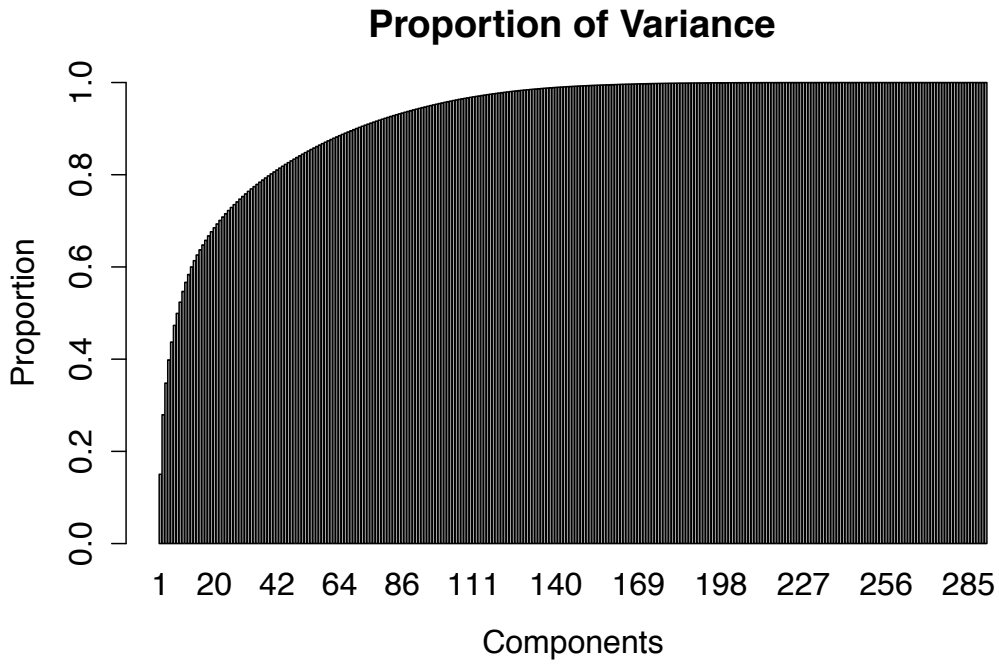


Figure C.1: Plot of cumulative variance with each additional principal component