

Data Analytical Programming

felicia tay sue ching  
TP044602

A report submitted in fulfilment of the

requirements for the course

Data Analytical Programming

CT050-3-M-DAP

Asia Pacific University of Technology & Innovation (APU)

School of Computing and Technology

15th May 2017

# DECLARATION OF ORIGINALITY AND EXCLUSIVENESS

I declare that this assignment report entitled

Data Analytical Programming  
 is the result of my own research work except as cited in the references.

|  |  |
| --- | --- |
|  |  |
| Name: | Felicia Tay Sue Ching |
|  | TP044602 |
| Date: | 15th May 2017 |

# ACKNOWLEDGEMENT

I would like to express my gratitude to Dr. Kalai Anand Ratnam, for inspiring me in crime analysis through this assignment and in providing quality guidance in many areas within Data Science & Analytics including SAS, without which the completion of this report would not be possible.

Felicia Tay Sue Ching

TP044602

# Table of Contents

[DECLARATION OF ORIGINALITY AND EXCLUSIVENESS i](#_Toc475183240)

[ACKNOWLEDGEMENT ii](#_Toc475183241)

[Table of Contents iii](#_Toc475183242)

[LIST OF TABLES iv](#_Toc475183243)

[LIST OF FIGURES v](#_Toc475183244)

[1. Introduction 1](#_Toc475183245)

[2. Problem Statement 2](#_Toc475183246)

[3. Aim and Objective 2](#_Toc475183247)

[4. Discussion on Data Sets and Methodologies 3](#_Toc475183248)

[5. Data Cleansing & Preparation 5](#_Toc475183248)

[6. Data Upload – Identify Missing Values 7](#_Toc475183248)

[7. Data Upload – Determine Sample Size (Top 5 States) 9](#_Toc475183248)

[8. Analysis : Property Crime 13](#_Toc475183248)

[9. Analysis : Violent Crime 26](#_Toc475183248)

[10. Analysis : Arson 33](#_Toc475183248)

[11. Conclusion 37](#_Toc475183262)

12. [References 38](#_Toc475183263)

APPENDICES

# LIST OF TABLES

[Table 1: Missing Values 8](#_Toc475183042)

[Table 2: Property Crime Analysis 13](#_Toc475183043)

[Table 3: Property Crime in New York 13](#_Toc475183044)

[Table 4: Larceny Theft by Type 14](#_Toc475183045)

[Table 5: Property Crime in Illinois 15](#_Toc475183045)

[Table 6: Property Crime in Texas 21](#_Toc475183045)

[Table 7: Violent Crime Analysis 23](#_Toc475183045)

[Table 8: Violent Crime in New York 23](#_Toc475183045)

[Table 9: Violent Crime in Illinois 26](#_Toc475183045)

[Table 10: Violent Crime in Texas 30](#_Toc475183045)

[Table 11: Total Fires, Deaths and Property Damage in the US 34](#_Toc475183045)

[Table 12: Intentional Fires Investigation and Suspected Motive 35](#_Toc475183045)

[Table 13: Intentional or Playing Fires by Age 36](#_Toc475183045)

# LIST OF **FIGURES**

[Figure 1: KDD methodology 3](#_Toc475183073)

[Figure 2: Application of KDD on Crime Analysis 4](#_Toc475183074)

[Figure 3: Scatter Plot for Violent Crime and Property Crime 10](#_Toc475183075)

[Figure 4: Pie Chart for Property Crime 12](#_Toc475183076)

[Figure 5: Distribution of Property Crime 17](#_Toc475183077)

[Figure 6: Larceny Theft Definition in Illinois 19](#_Toc475183078)

[Figure 7: Pie Chart for Violent Crime 22](#_Toc475183079)

[Figure 8: Distribution of Violent Crime 25](#_Toc475183080)

[Figure 9: Top Origin States for Illinois Crime Guns 27](#_Toc475183081)

[Figure 10: Murder Cases by State 29](#_Toc475183082)

[Figure 11: Gun Death Rate vs Gun Control 32](#_Toc475183083)

[Figure 12: Pie Chart for Arson 33](#_Toc475183084)

1. Introduction

Investigators worldwide are hoping to enforce laws through the advancements in technologies and big data analytics. The collection and analysis of data is important to both government and private sectors, to ensure that crime rates do not increase over time.

The usage of data to generate useful insights and predictions has its challenges, particularly in ensuring that the source of data is genuine. It is therefore important for government agencies to gather relevant and specific data for a better analysis. In this report, the source of data will be taken from the Uniform Crime Reports (UCR) of the FBI for the first 6 months (January – June) of 2014 and 2015 in cities with populations of 100,000 and over.

The UCR is the FBI’s widely used system for recording crimes and making policy decisions. Since 1930s, it has tracked data on the following crimes: murder, robbery, rape, aggravated assault, burglary, theft and vehicle theft. In 1979, the UCR started reporting on arson. The UCR website: https://ucr.fbi.gov/ represents a good source for law enforcements researchers to perform analysis on reliable uniform crime statistics.

2. Problem Statement

I have been assigned to carry out a study on the violent crime and property crime in the United Stated of America by the Headquarters of Federal Bureau of Investigations (FBI), Washington, D.C., United States. The scope of data set to be analyzed concerns the reported crimes for the first 6 months (January – June) of 2014 and 2015 in cities with populations of 100,000 and over.

3. Aim and Objective

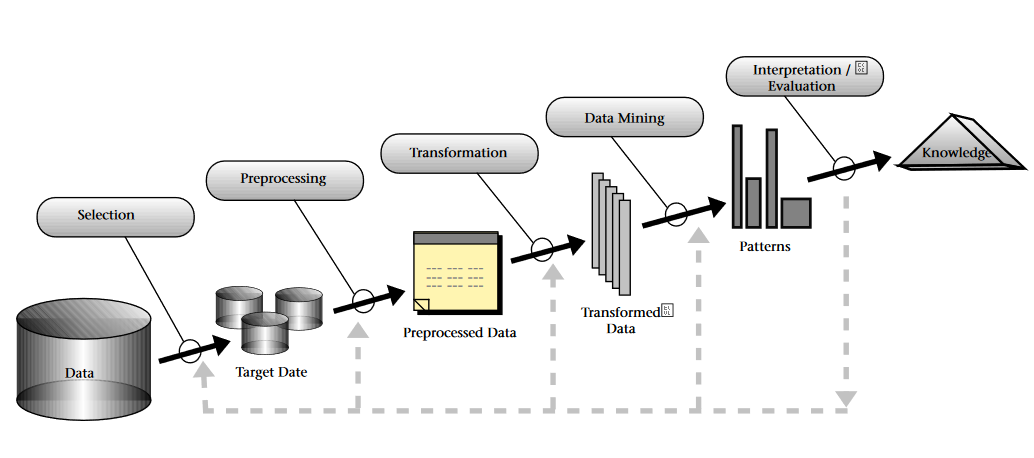
The main objective of this report is to analyze the preliminary data set from the Uniform Crime Reports of the FBI for the first 6 months (January – June) of 2014 and 2015 in cities with populations of 100,000 and over, make suitable recommendations / conclusions on the findings.

The following are the aims of this study:

* To identify top crime states by population, number of crimes and cities.
* To analyze further on the types of crimes and its trends from year 2014 to 2015.
* To make use of SAS for analysis and visualization to derive useful insights and make suitable recommendations to the readers of this report.

## 4. Discussion on Data Sets and Methodologies

The Knowledge Discovery in Databases (KDD) methodology is applied for this report, which starts from data selection, preprocessing & cleaning, transformation, data mining to deriving knowledge from interpretation / evaluation:



*Figure 1:* *KDD methodology (Fayyad, Piatetsky-Shapiro and Smyth 38-54)*

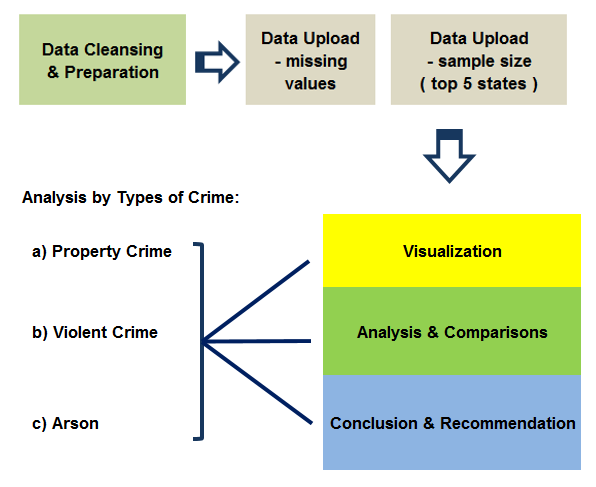
For this report, the purpose of data analysis within this data set is to track the highest crime locations by population and number of crimes, and drill down further into these hotspots to detect any potential cause for the high crime rates.

The first step is to reduce the scope of analysis to the largest crime areas, which will represent a sample size for population of the United States for further crime analysis.

To determine the top states with the most number of crimes, a separate transformed report is loaded into SAS for easier manipulation. The rationale to upload separate worksheets into SAS is because SAS stores the working files into temporary folders, which will be automatically cleared each time the session is logged off. Therefore, to ensure that worksheets are stored permanently for transformation within SAS, an upload is essential.

To apply the KDD methodology, the following steps are applied to the data set:

1. Data Cleansing and Preparation
2. Data Upload to identify missing values and other abnormal items
3. Data Upload to determine sample size & upload of the actual sample size
4. Analysis by types of crime through the techniques of visualization, analysis & comparisons, conclusion & recommendations



*Figure 2:* *Application of KDD on Crime Analysis*

5. Data Cleansing & Preparation

Data cleaning represents an important step prior to data analysis, this is because the analytical tool (SAS) needs the data set to be in an acceptable format before further changes can be made to the dataset.

2015 data is not available at the initial data set. Therefore, 2015 data was downloaded from the 2016 release of January to June 2015–2016 Uniform Crime Report published at the following website: <https://ucr.fbi.gov/crime-in-the-u.s/2016/preliminary-semiannual-uniform-crime-report-januaryjune-2016/tables/table-4/table_4_january_to_june_2015_offenses_reported_to_law_enforcement_by_state_by_city_100-000_and_over_in_population/view>

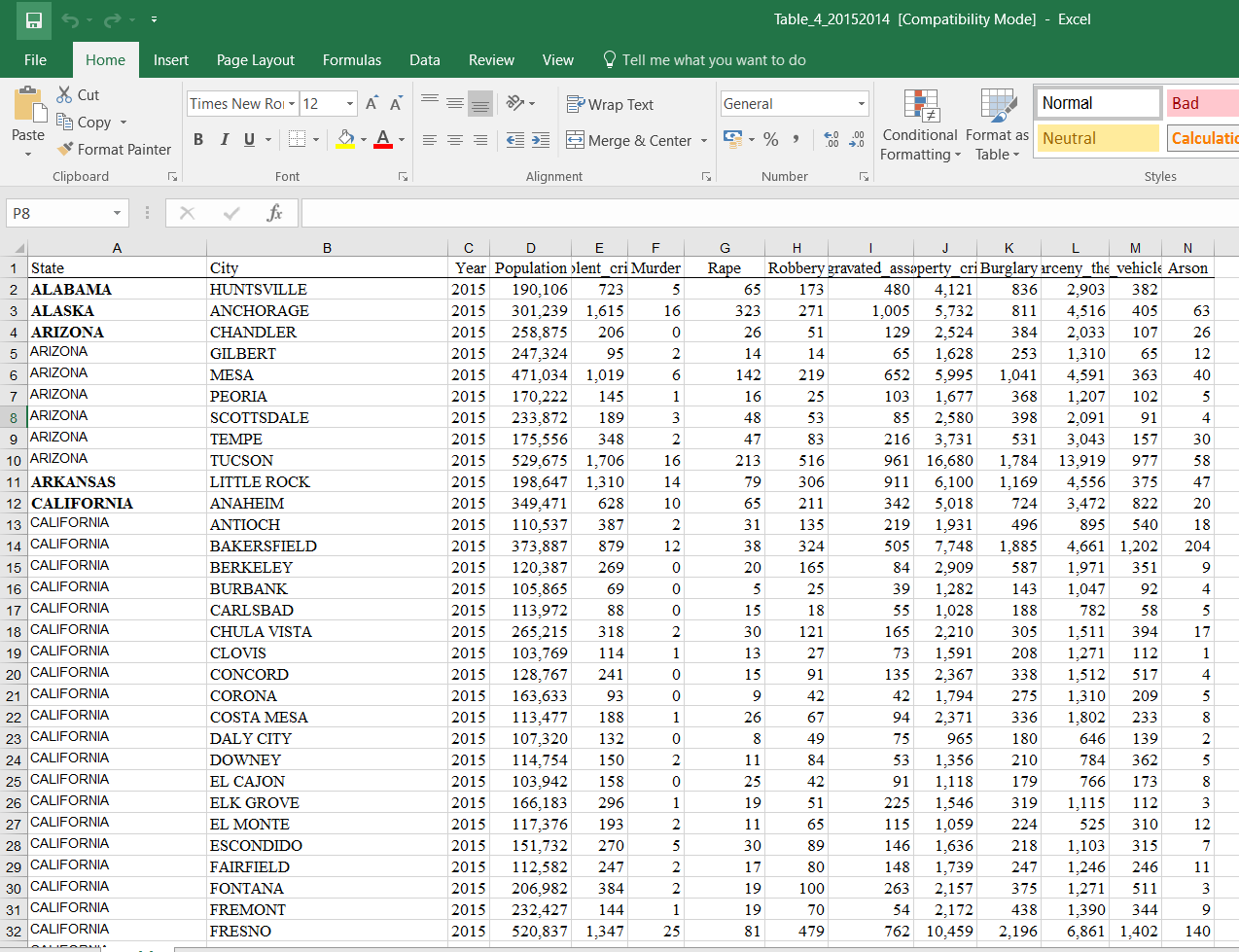
The following steps were taken to clean the data for both 2015 & 2016 data sets:

* Unmerge all merged cells using Excel ‘Merge and Center’ function -> ‘Unmerge Cells’
* Remove all notes at the bottom of the worksheet and Header Titles
* Replace all blank cells with the appropriate state (highlight Column A&B -> F5 -> Special -> Blanks -> = top cell Ctrl Enter) Copy paste value for both Columns A&B.
* Combine columns for Rape Legacy and Rape Revise
* Sort by year

For easier analysis, only one dataset is used to be uploaded into SAS for further analysis. Therefore, the 2014 data and 2015 data are consolidated into one new excel file -> rename as ‘Table4new’:

* Copy paste header
* Change header names to replace space with ‘\_’ and dash with ‘\_’
* Copy paste 2014 data from the 2015 data set
* Copy paste 2015 data from the 2016 data set
* Remove all numbers from header and description columns (Column A & B)

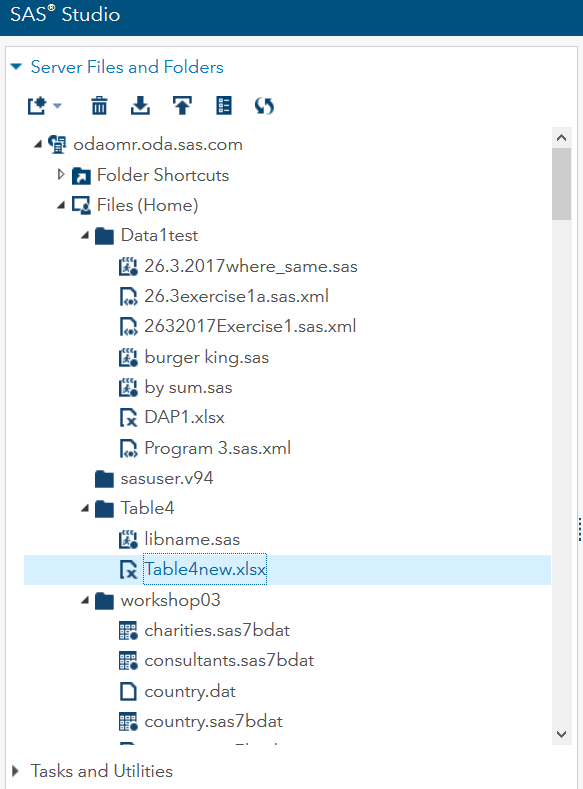
The final cleaned data set is presented in the following format:



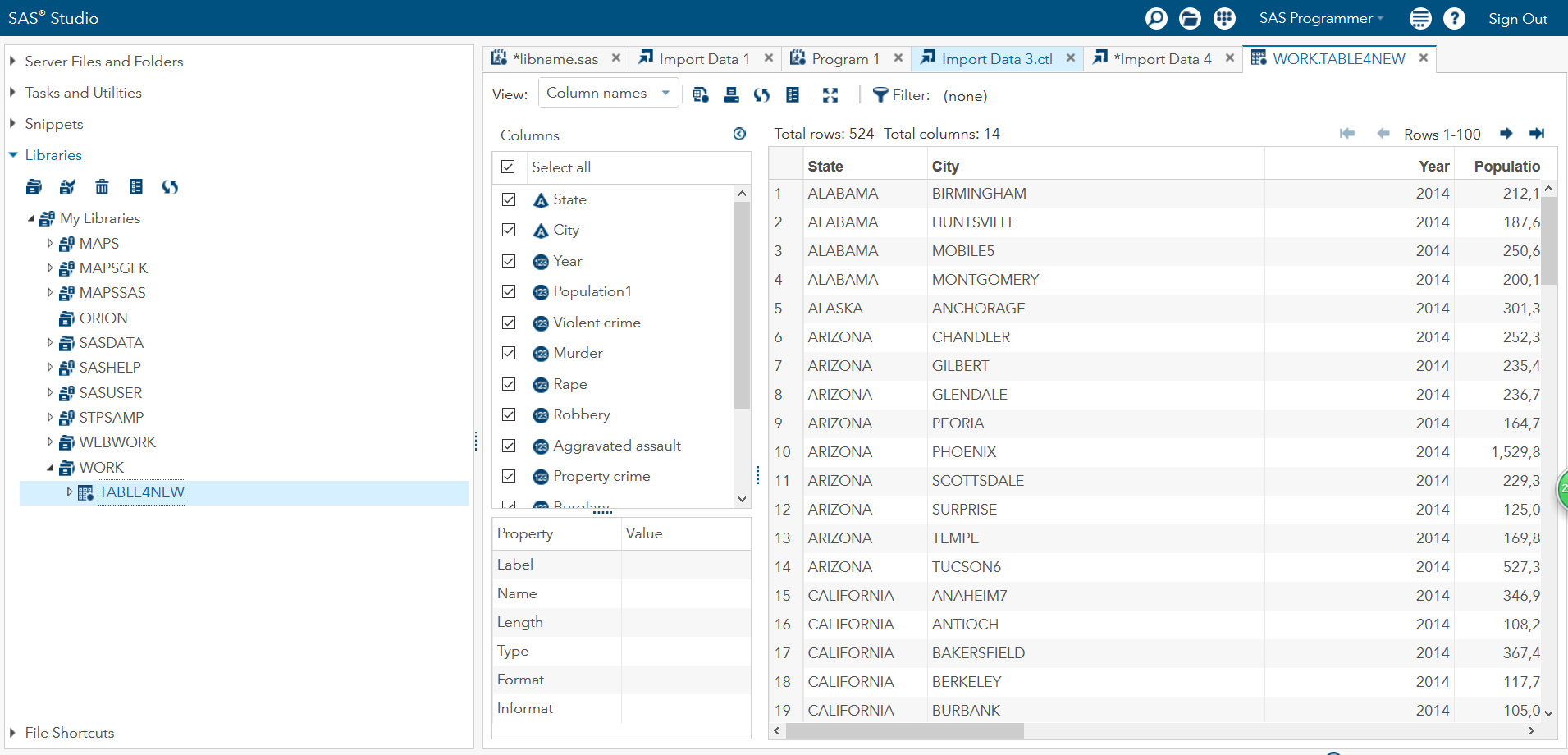
6. Data Upload – Identify Missing Values

The cleaned dataset is uploaded into SAS Studio:

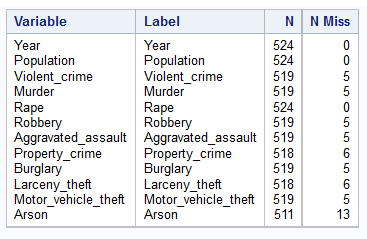
* Server Files and Folders -> create a new folder named ‘Table4’, click on upload button to upload the excel file Table4new.xlsx



* Then, run SAS codes in **Appendix 1** (proc import) to import data -> name the work folder as “WORK.table1415”.
* Go to Libraries to check that TABLE4NEW is now under ‘WORK’ folder



To check missing values, run SAS codes in **Appendix 2** (proc means) to generate the following table:



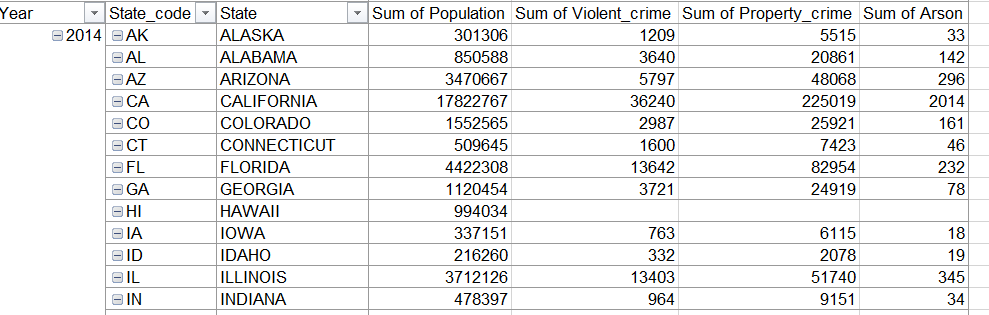
*Table 1: Missing Values*

From the table of contents, there are 13 missing values for Arson (13/511 = 2.5%), 5 missing values for Violent Crime (5/519 = 1.0%) and 6 missing values for Property Crime (6/518 = 1.2%). Since number of missing values over total observations are less than 5% for each main category that makes up total crime, the analysis will proceed with the assumption that missing values = 0.

7. Data Upload – Determine Sample Size (Top 5 States):

Prior to upload, go through the following steps:

* First, match the states with the US state code for abbreviations, copy the state codes from this website: <http://www.50states.com/abbreviations.htm>. Do a vlookup of state to state codes -> Copy paste value to set the state codes.
* Pivot the data by year, state code, state, population, violent crime, property crime, arson.



* Copy paste value for all data, fill up the blank cells for Year column. Take a step further to add some useful information into this table. The purpose of these columns is to facilitate further analysis by type of crimes by state. Crime rates are calculated to view number of crimes per capita, in this case per 100,000 people.
* Add in additional columns for ‘Total\_crime’, ‘Total\_crime\_rate’, ‘Violent\_crime\_rate’, ‘Property\_crime\_rate’, ‘Arson\_rate’. The formulas are:
  + ‘Total\_crime’ (addition of violent crime, property crime and arson)
  + ‘Total\_crime\_rate’ (Total\_crime / Population \* 100,000)
  + ‘Violent\_crime\_rate’ (Violent\_crime / Population \* 100,000)
  + ‘Property\_crime\_rate’ (Property\_crime / Population \* 100,000)
  + ‘Arson\_rate’ (Arson / Population \* 100,000)
* Copy paste value to override formulas and load the worksheet into SAS studio as ‘Crime\_rate\_by\_state’. Run the codes in **Appendix 3** (proc import) to import data into temporary work folder named “WORK.Crime\_Rate\_by\_State”.

(\*Note on crime rates: Throughout this report, crime rates were not discussed due to time constraints. Comparison was made between total number of crimes by top 5 states vs number of crimes ranked by high crime rates, which is too low to treat as sufficient analysis for the purpose of this report.)

* Next, run the SAS codes in **Appendix 4** (proc sgplot) to generate a scatter plot for violent crime and property crime by state codes:



*Figure 3:* *Scatter Plot for Violent Crime and Property Crime*

From the above scatter plot, it is observed that for the two largest categories of crimes - Violent Crime and Property Crime, the top 5 states by crime are:

* California
* Texas
* New York
* Florida
* Illinois

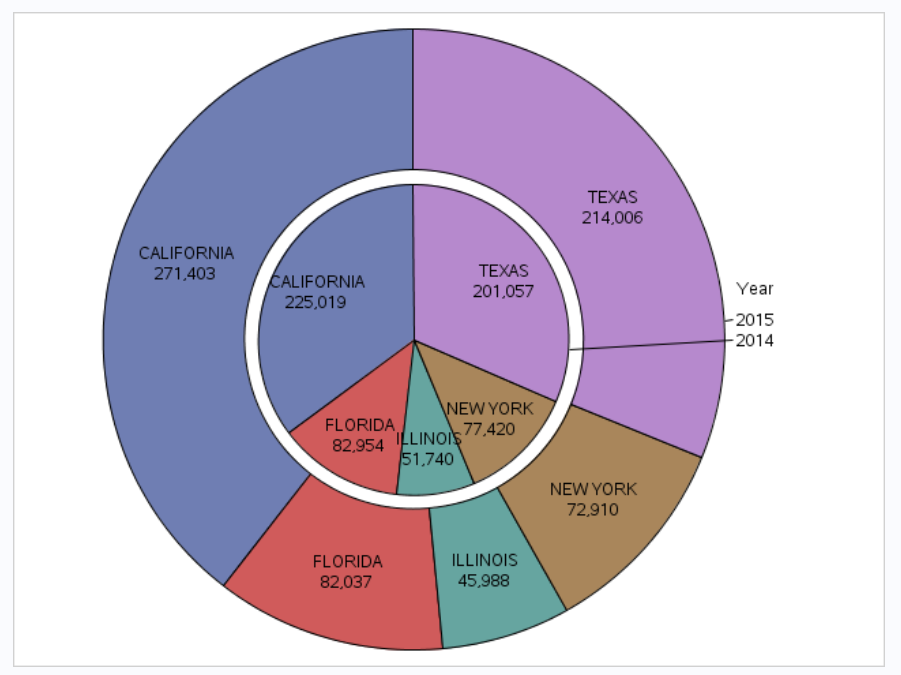
Therefore, the sample size is determined by the top 5 states with the highest number of crime cases.

To enable a better data analysis within SAS Studio, a separate set of data for this sample of top 5 states will be loaded into SAS:

* Filter only California, Texas, New York, Florida, Illinois
* F5 -> select visible cells -> copy paste to a new worksheet -> save as ‘top5states’
* Load into SAS studio, import data into temporary work folder.
* Run the SAS codes in **Appendix 5** (proc import) to import the top 5 states, name work folder as “WORK.TOP5STATES”.

## 8. Analysis : Property Crime

To view total cases of property crime by top 5 states, generate a pie chart through running the SAS codes in **Appendix 6** (proc template).

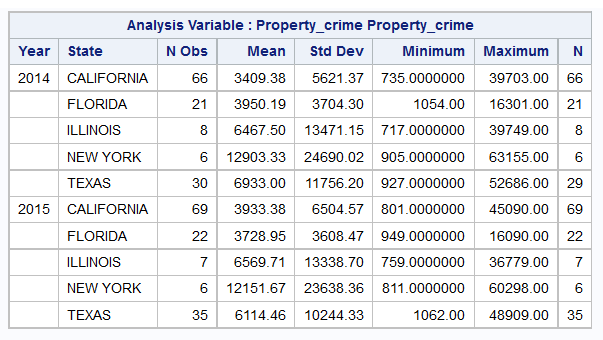


*Figure 4:* *Pie Chart for Property Crime*

The pie chart above shows the number of property crimes by the top 5 states. For both years 2014 & 2015, California has the highest number of property crimes (2015: 271,403 cases, 2014: 225,019 cases), followed by Texas (2015: 214,006 cases, 2014: 201,057 cases), Florida (2015: 82,037 cases, 2014: 82,954 cases), New York (2015: 72,910 cases, 2014: 77,420 cases), and Illinois (2015: 45,988 cases, 2014: 51,740 cases).

For both California and Texas, property crime had increased from year 2014 to 2015, where New York, Illinois and Florida decreased during the same period. The greatest increase came from California, by 46,384 cases or 21%.

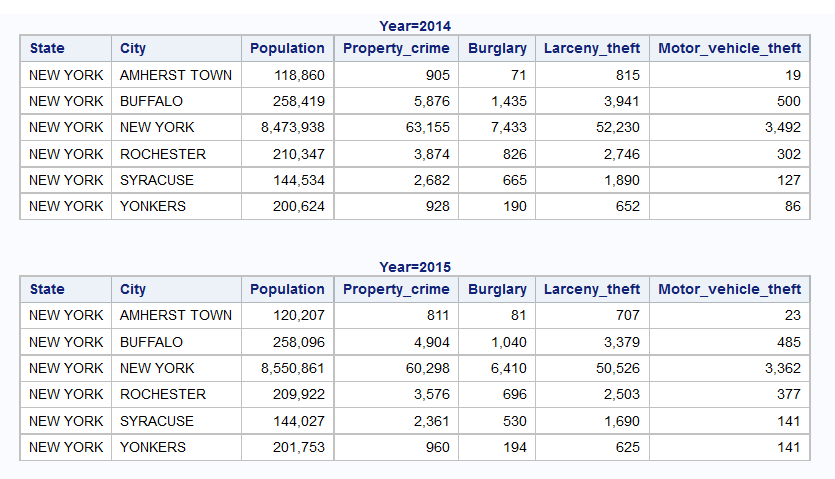
For further details of property crime for each state, run the SAS codes in **Appendix 7** (proc means) to generate the number of observations, mean, standard deviation, minimum and maximum:



*Table 2:* *Property Crime Analysis*

From the table above, we can see that under the Maximum column, New York has an extreme distribution of property crime records, with one of its cities with property crime as high as 63,155 cases in year 2014 and 60,298 cases in year 2015.

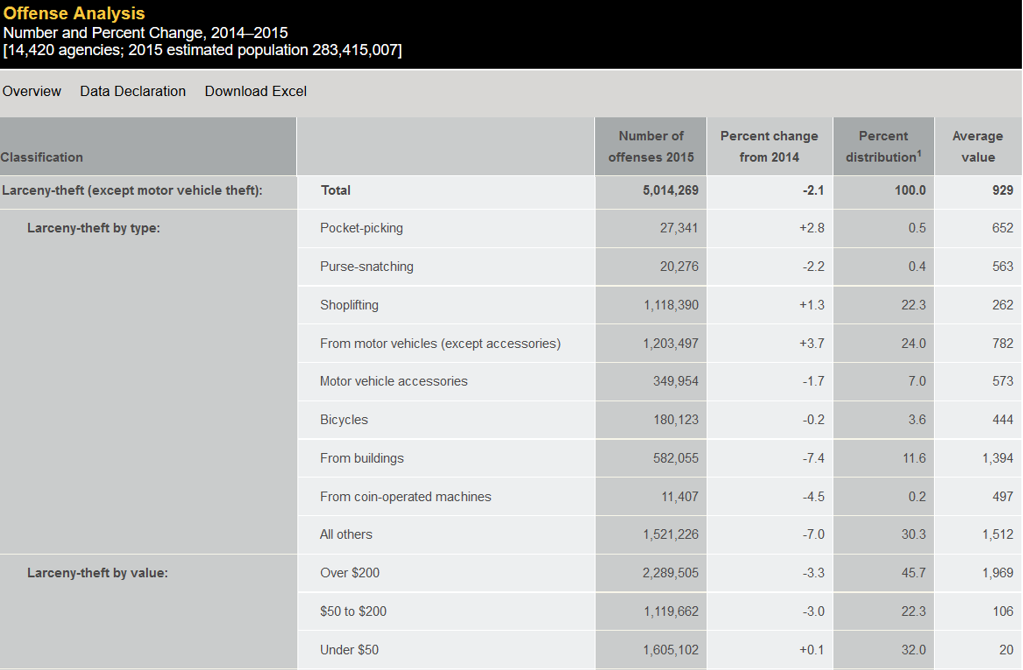
To find out which City from New York has the extreme distribution, run the SAS codes in **Appendix 8** (proc sort & proc print) to zoom into New York -> select the following variables: State, City, Population, Property\_crime, Burglary, Larceny\_theft, Motor\_vehicle\_theft:



*Table 3:* *Property Crime in New York*

From the table above, the high property crime record is coming from New York City, and the type of crime that has the highest cases under Property Crime is Larceny Theft (2015: 50,526 cases, 2014: 52,230 cases). There is a decrease of Larceny Theft by 5.5% from year 2014 to 2015, although population has increased slightly by 0.9% (2015: 8,550,861 people, 2014: 8,473,938 people).

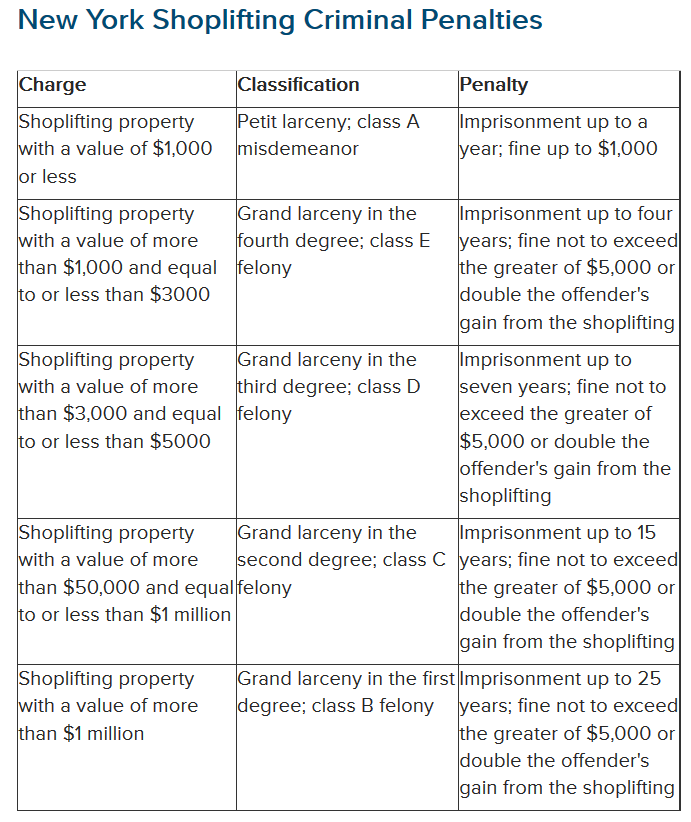
Larceny Theft is defined as “the unlawful taking, carrying, leading, or riding away of property from the possession or constructive possession of another. Examples are thefts of bicycles, motor vehicle parts and accessories, shoplifting, pocket-picking, or the stealing of any property or article that is not taken by force and violence or by fraud. Attempted larcenies are included. Embezzlement, confidence games, forgery, check fraud, etc., are excluded.” ("Larceny-Theft").



*Table 4:* *Larceny Theft by Type ("Table 23")*

The above table is taken from the UCR website on the classification of crimes and its percentage change from 2014 to 2015. From the types of larceny theft, it is observed that the majority of larceny theft is made up of shoplifting and from motor vehicles. The value of larceny theft ranges from under $50 (1.6million cases) to $50 to $200 (1.1million cases) and above $200 (2.3million cases). In other words, 2.7 million cases is below $200.

In New York, shoplifting is charged as petit larceny if the value of the items stolen is $1,000 or less, and as grand larceny if the value is above $1,000 (Stine). The following table shows the type of penalties for each category of shoplifting charge in New York. The penalties are quite severe and can go up to 25 years of imprisonment.



*Table 5: Source: (Stine)*

Shoplifting represents an economic loss for retailers. Walmart used to have a zero tolerance policy on shoplifting, however has changed their shoplifting policy since July 2006 to let shoplifters go if the item shoplifted is below $25 (Barbaro).

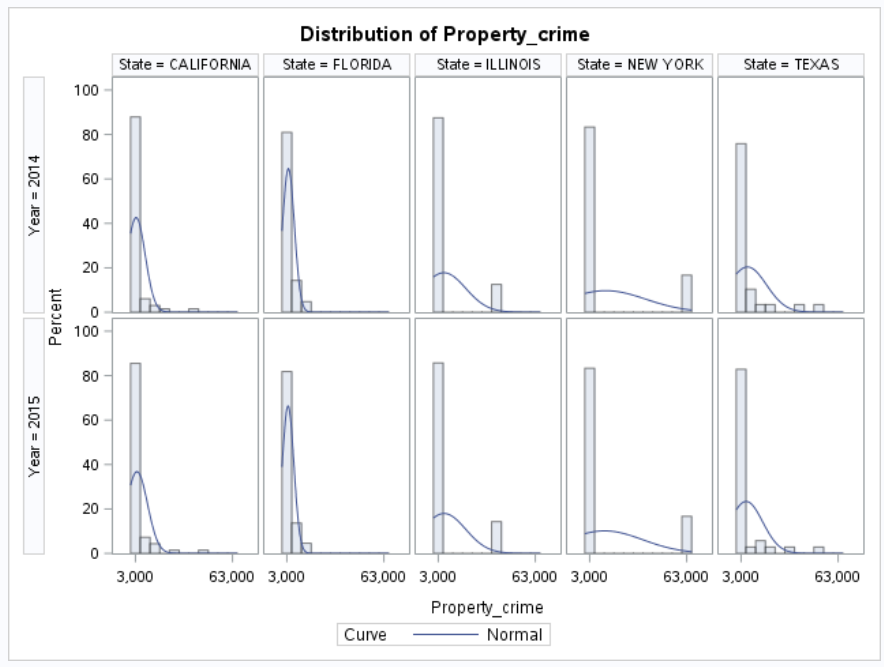
Prosecuting shoplifters represents a cost to the police authorities, as the wage of the police officer may not be covering the cost of the shoplifted item. Therefore, it can be said that retailers often tolerate petite shoplifters instead of prosecuting them because the prosecution represents a reluctance from the authorities due to volume of cases and cost consideration, as well as higher resources to be allocated by the retailer for shoplifting detection. Therein lies the irony where cost is concerned, which does not encourage both authorities and retailers in persecuting shoplifters.

**Recommendation:**

It is recommended that law enforcers work together with major retailers in New York city, to focus efforts on crime prevention through setting up deterrents within the retail stores. As shoplifting represents stock loss, retailers would be glad to reduce their stock loss but not without effort. Suggested deterrents include:

1. Offering top notch customer service, ie. making sure employees greet customers right away upon entering the stores and be nearby if they need help. The presence of employees nearby will deter those who do not intend to shop from lingering too long.
2. Keeping shelves organized and clean. If displays are kept visible and tidy, employees will be aware if something goes missing. Often, shoplifters are able to steal because no one notices that the stock has gone missing.
3. Keep a record of high stock loss item and display these items in plain sight. Retailers should encourage employees to take extra precaution in these items, and should place the items where everyone can see them, for example at the front of the store or near the cash register.
4. Encourage employees to walk around frequently. If each alley has employees wandering up and down, shoplifters can be deterred as they would not want to be caught.

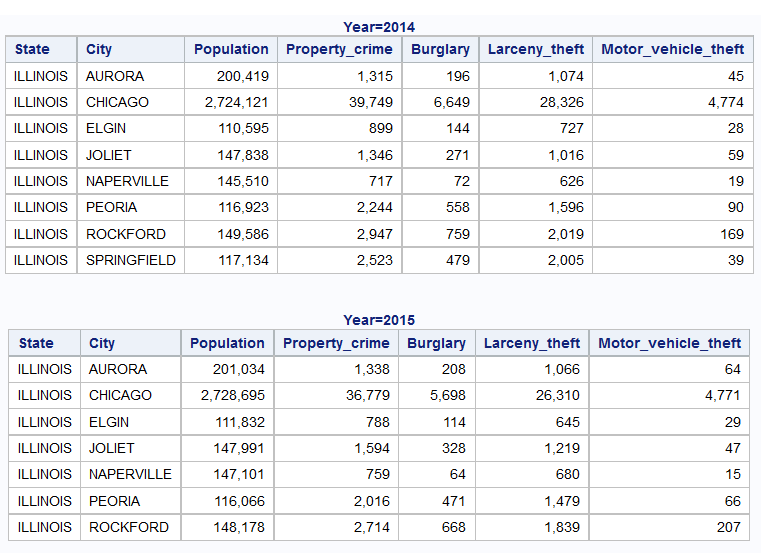
Another method of analyzing the distribution of Property Crimes can be performed through generating histograms through running the SAS codes in **Appendix 9** (proc univariate). The following histograms by state are generated:



*Figure 5:* *Distribution of Property Crime*

From the above distribution graphs, it is observed that Illinois, New York and Texas has a wider horizontal distribution. Referring back to *Table 2: Property Crime Analysis*, the analysis indicated that there are cities that have extreme distribution of Property Crime.

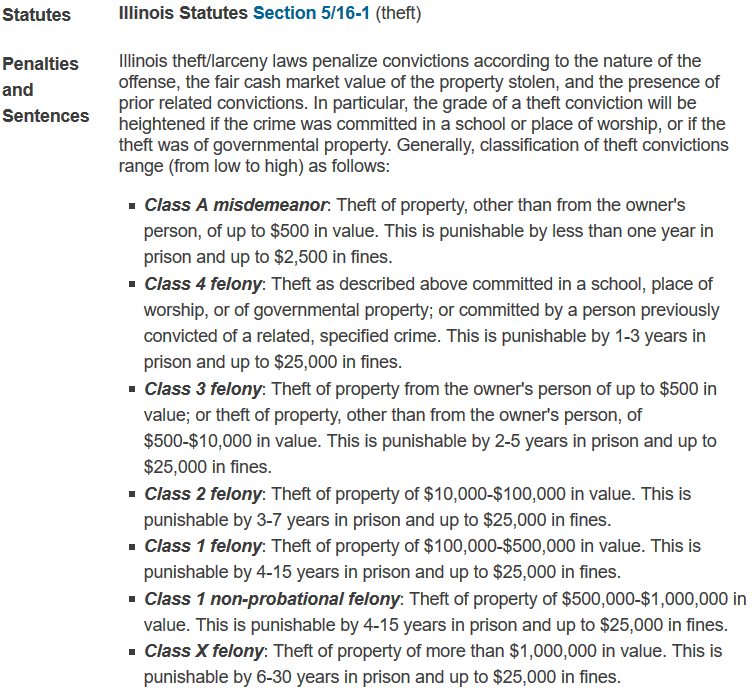
To drill down further, run the SAS codes in **Appendix 10** (proc sort & proc print) to zoom into Illinois -> select the following variables: State, City, Population, Property\_crime, Burglary, Larceny\_theft, Motor\_vehicle\_theft. The following table is generated:



*Table 5:* *Property Crime in Illinois*

From the list of data above, this high property crime record is coming from Chicago City, and the type of crime that has the highest cases under Property Crime is Larceny Theft (2015: 26,310 cases, 2014: 28,326 cases).

In Illinois, Larceny Theft is classified as Felony is the theft is $500 or more ("Illinois Theft / Larceny Laws - Findlaw"). The following table shows the types of felony and penalties in Illinois, which can reach up to 30 years in prison.



*Figure 6:* *Larceny Theft Definition in Illinois ("Illinois Theft / Larceny Laws - Findlaw")*

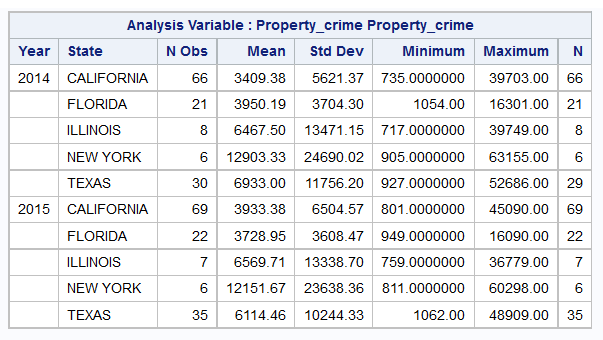
However, effective 16th December 2016, shoplifters will no longer be charged with felony for stealing less that $1000 (Demarest). This move was put into effect by the Cook Country administration, to prioritize state resources as they realized that much of the Cook County jail space was occupied by shoplifters who were spending a lot of time in jail, and they were often homeless, mentally ill or suffering from addiction (Demarest).

From the illustration above, there exists a conflict of interest between retailers on shoplifting losses and politicians on saving jail space. The contributing factors to shoplifting might also be due to homelessness, mental illness and addiction. As of 2015, there is an estimation of 140,000 homeless people living in Chicago, including thousands of public school students living in shelters, tents and parks (Goudie).

**Recommendation:**

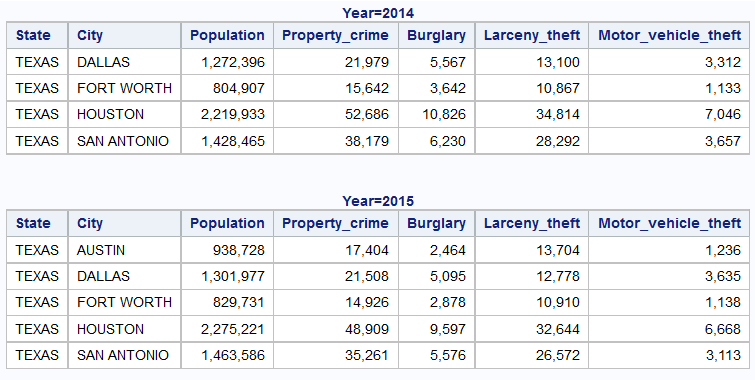
It is recommended that law enforcers advocate to politicians and retailers to solve the issues of shoplifting through other non-conventional methods, in the case of Chicago, sponsoring or setting up shelters for the homeless students and families. This move not only helps boost the politician ratings and retailer’s reputation, the minimal amount of money spent on shelters will help to improve the living quality of the city and hopefully reduce Larceny Theft.

Next, identify the extreme distribution within Texas through the above similar steps. From the previous *Table 2: Property Crime Analysis*, there are at least 29 observations for Texas.



Therefore, a slight modification is made to the SAS codes to zoom into Texas -> add in a where function for “Propery\_crime>10000”, since it is already made known in the table above that the maximum is more than 10,000 cases. Run the SAS codes in **Appendix 11** (proc sort & proc print).

Select the following variables: State, City, Population, Property\_crime, Burglary, Larceny\_theft, Motor\_vehicle\_theft The following table is generated:



*Table 6: Property Crime in Texas*

From the above table, it is observed that both Houston (2015: 32,644 cases, 2014: 34,814 cases) and San Antonio (2015: 26,572 cases, 2014: 28,292 cases) contributes to the high Larceny Theft within Texas.

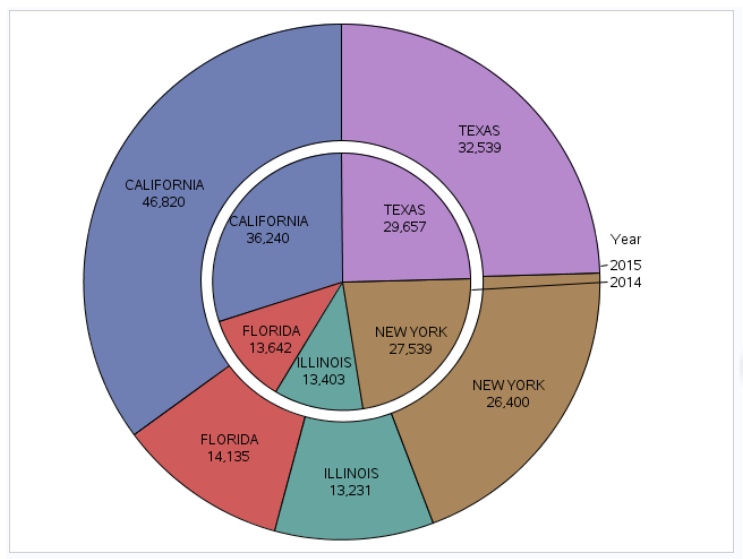
In San Antonio, Walmart has introduced a program called “Restorative Justice” to first-time offenders, where these shoplifters are given a chance to take an online interactive educational program where they are taught how to reduce debt and create a budget (Knapp).

**Recommendation:**

It is recommended that retailers come up with creative methods in preventing stock loss from shoplifting. For example, the educational program by Walmart serves as an example to prevent recurring shoplifting.

## 9. Analysis : Violent Crime

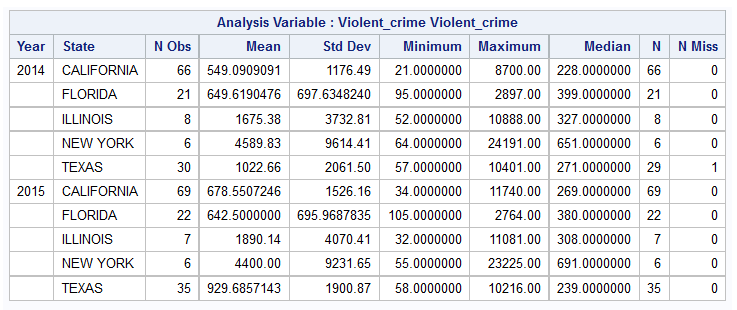
To view total cases of violent crime by top 5 states, generate a pie chart through the running the SAS codes in **Appendix 12** (proc template).



*Figure 7: Pie Chart for Violent Crime*

The pie chart above shows the number of violent crimes by the top 5 states. For both years 2014 & 2015, California has the highest number of violent crimes (2015: 46,820 cases, 2014: 36,240 cases), followed by Texas (2015: 32,539 cases, 2014: 29,657 cases), New York (2015: 46,820 cases, 2014: 36,240 cases), Florida (2015: 14,135 cases, 2014: 13,642 cases) and Illinois (2015: 13,231 cases, 2014: 13,403 cases).

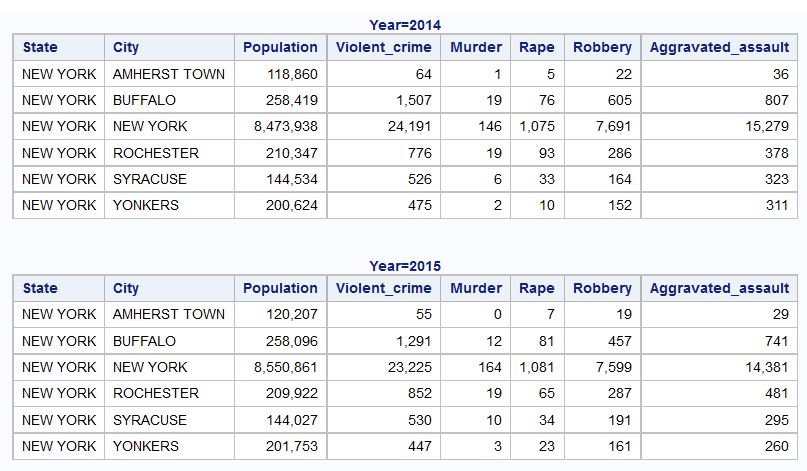
For further details of violent crime for each state, run the SAS codes in **Appendix 13** (proc means) to generate the number of observations, mean, standard deviation, minimum and maximum:



*Table 7: Violent Crime Analysis*

From the table above, we can see that under the Maximum column, New York has an extreme distribution of violent crime records, with one of its cities with violent crime as high as 24,191 cases in year 2014 and 23,225 case in year 2015.

To find out which City from New York has the extreme distribution, run the SAS codes in **Appendix 14** (proc sort & proc print) to zoom into New York -> select the following variables: State, City, Population, Violent\_crime, Murder, Rape, Robbery, Aggravated\_assault. The following table is generated:



*Table 8: Violent Crime in New York*

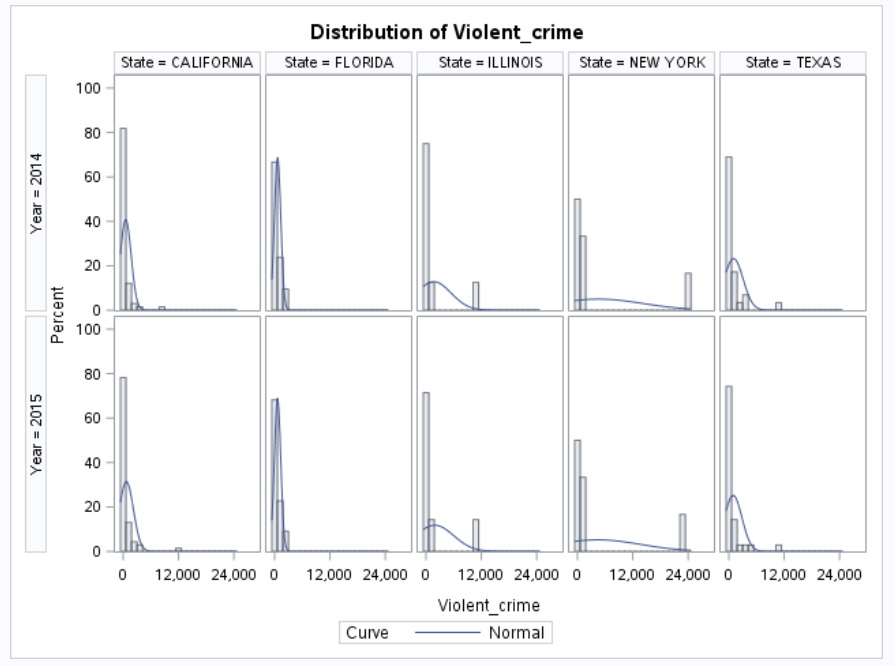
From the list of data above, this high violent crime record is coming from New York City, and the type of crime that has the highest cases under Violent Crime is Aggravated Assault (2015: 14,381 cases, 2014: 15,279 cases). There is a decrease of Aggravated Assault by 5.9% from year 2014 to 2015, although population has increased slightly by 0.9% (2015: 8,550,861 people, 2014: 8,473,938 people).

Aggravated Assault is defined as “A person commits aggravated assault when, with intent to cause serious physical injury to a person whom he or she knows or reasonably should know is a police officer or peace officer performing his or her duties, the defendant causes such injury by means of a deadly weapon or dangerous instrument; or a defendant 18 years old or more commits third-degree assault upon a person less than 11 years old and has been previously convicted of the same offense within the last 3 years.” (New York Assault And Battery Laws - Findlaw)

**Recommendation:**

It is recommended for law enforcers to focus efforts in reducing Aggravated Assault cases in New York City, either through awareness campaigns or increasing penalties in this type of crime within New York city. Further investigative efforts should be put into surveying the improvements that was carried out in year 2015 that enabled a reduction of 5.9% cases of Aggravated Assault in New York city.

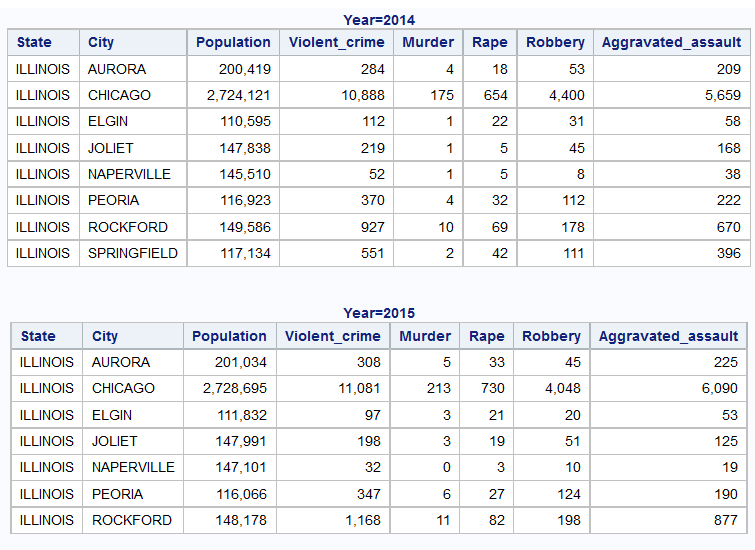
Another method of analyzing the distribution of Violent Crimes by state can be performed through generating histograms through the running the SAS codes in **Appendix 15** (proc univariate).



*Figure 8: Distribution of Violent Crime*

From the above distribution graphs, it is observed that Illinois, New York and Texas has a wider horizontal distribution. Similarly to the Summary Statistics table, the graphs indicate that there are cities that have extreme distribution of Violent Crime.

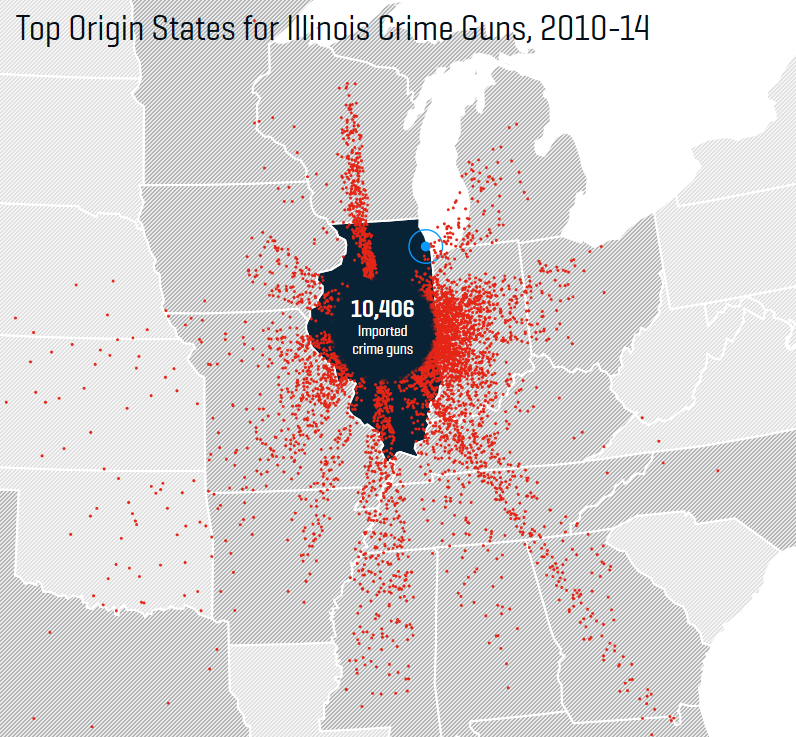
To drill down further, run the SAS codes in **Appendix 16** (proc sort & proc print) to zoom into Illinois -> select the following variables: State, City, Population, Violent\_crime, Murder, Rape, Robbery, Aggravated\_assault. The following table is generated:



*Table 9: Violent Crime in Illinois*

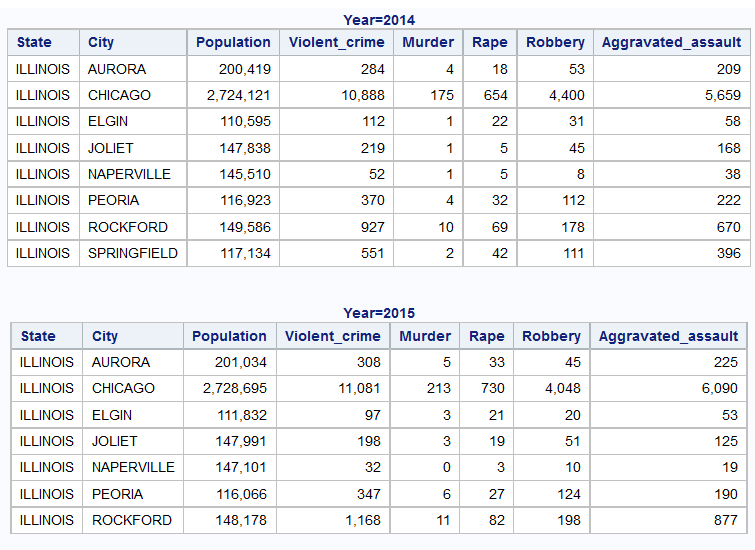
From the list of data above, this high violent crime record is coming from Chicago City, and the type of crime that has the highest cases under Violent Crime is Robbery (2015: 4,048 cases, 2014: 4,400 cases) and Aggravated Assault (2015: 6,090 cases, 2014: 5,659 cases). Both Robbery and Aggravated Assault contributes to the majority crimes within Chicago. The city of Chicago is described as having the toughest gun laws in the US, yet has high levels of gun violence (Berman). This is mostly caused by gun trafficking from other states with looser gun rules (Berman). The following diagram shows the traffic of guns going into Illinois from the top 15 states from year 2010 to 2014 (Spies and Fuhrman):

* Indiana 3,269
* Mississippi 1,002
* Wisconsin 898
* Missouri 780
* Kentucky 503
* Tennessee 459
* Georgia 453
* Ohio 448
* Texas 444
* Florida 439
* Michigan 365
* Iowa 344
* Alabama 315
* Arkansas 280
* Others 407

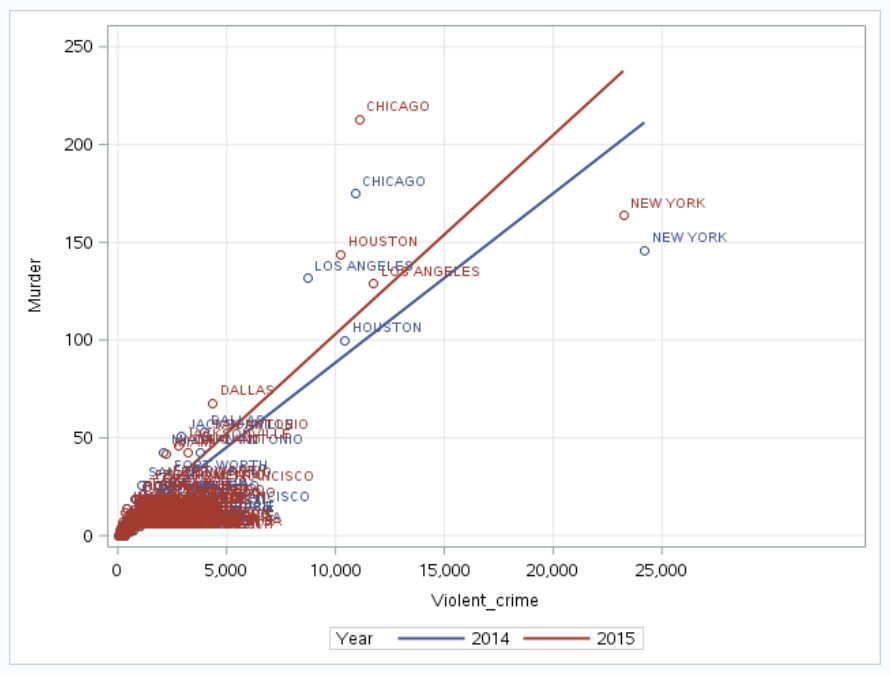


*Figure 9:* *Top Origin States for Illinois Crime Guns (Spies and Fuhrman)*

The high inflow of illegal guns to Illinois leads to not only Aggravated Assault, but also a high number of murder and robbery cases in Chicago, as shown in *Table 9: Violent Crime in Illinois* tabulated earlier on violent crime types by city for Illinois:



To compare murder cases across the rest of the cities for the top 5 states, run the SAS codes in **Appendix 17** (proc sgplot) to generate a scatter plot for total murder cases over total number of violent crimes by city:



*Figure 10:* *Murder Cases by State*

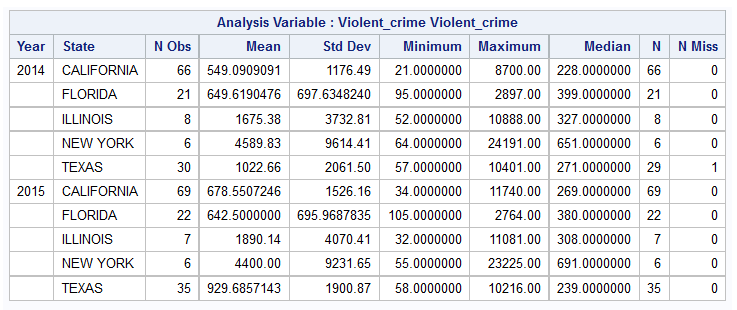
From the scatter plot, murder numbers for Chicago city is the highest, which coincides with the inflow of illegal guns into the state of Illinois. The residents of Chicago is faced with daily reports of shootings, with police forces fearing for their own lives, of which city officials are trying to confront with recruiting more police officers and enforce new laws (Smith and Smith). Although gun violence may seem to have become a culture to the average Chicago resident, children and families are fearing for their lives and losing hope in the officials to help reduce violence (Smith and Smith).

Perhaps the greatest difficulty for the police officials in Chicago is in finding ways to overcome the gang members, who solve interpersonal conflicts through murder (Smith and Smith). It takes a tough personality to set the rules with these gang members. If there are no such leadership within the police force, then other methods need to be taken into consideration. Residents would have to find ways in their own initiatives to leave this high crime area and search for other safer avenues.

**Recommendation:**

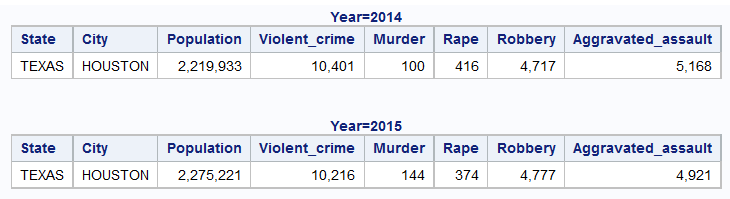
From the above statistics of gun trafficking, it is recommended for policy makers on gun ownership to work collaboratively across the various affected states, especially states where the originating guns resides the most, in this case Indiana, Mississippi, Wisconsin, Missouri etc. By restricting the supply of illegal guns to Illinois, the rate of violent crimes committed within Chicago can hopefully be reduced.

Next, identify the extreme distribution within Texas through the above similar steps. From *Table 7: Violent Crime Analysis*generated earlier on, there are at least 30 observations for Texas.



Therefore, a slight modification is made to the code to zoom into Texas -> add in a where function for “Violent\_crime>10000”, since it is already made known in the table above that the maximum is more than 10,000 cases. Run the SAS codes in **Appendix 18** (proc sort & proc print).

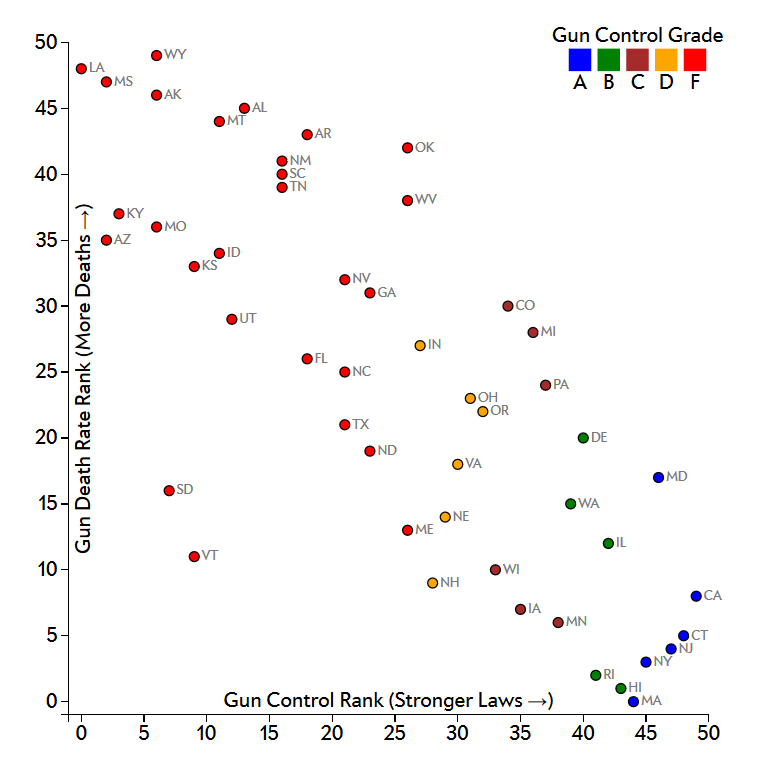
Select the following variables: State, City, Population, Violent\_crime, Murder, Rape, Robbery, Aggravated\_assault. The following table is generated:



*Table 10: Violent Crime in Texas*

From the above table, it is observed that both Robbery (2015: 4,777 cases, 2014: 4,717 cases) and Aggravated Assault (2015: 4,921 cases, 2014: 5,168 cases) contributes to the majority crimes within Houston. Compared to Illinois, Texas is a state where legalized guns purchased is encouraged as a culture (Schiller). The largest group of gun license holders in Texas are 55 years old, approximately 22.871 people out of the 1 million statewide gun license holders (Fernandez).

In the scatter plot below, each dot represents a state in the US, where each state is ranked for its gun control laws (with A the strongest control to F the weakest control), and its 2013 gun deaths rate rank from lowest at 0 to highest at 50. Illinois, California and New York is at the bottom right quadrant, where gun controls are high and gun death rank is low. Texas has a lower gun death rate compared to Florida. Both Florida and Texas have the lowest gun control grades. In conclusion, states with lower gun control tend to have more gun deaths.



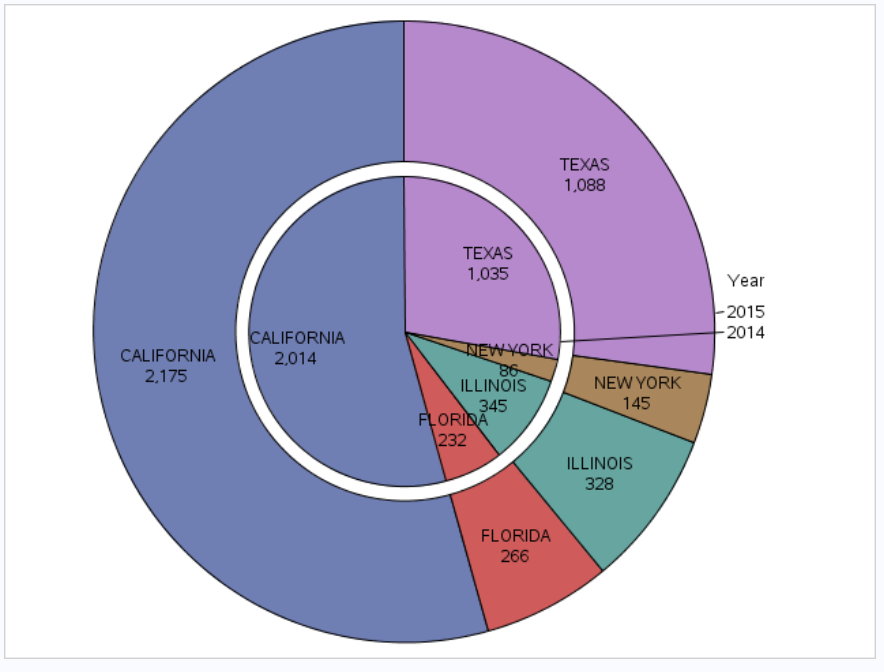
*Figure 11:* *Gun Death Rate vs Gun Control (Source: Kirk)*

**Recommendation:**

From the observations on gun control laws vs gun deaths, it is recommended that Texas law enforcement implement tighter controls on gun control. Passing laws for legalizing gun ownership does not aid in keeping violent crime rates low. There might be a stronger link between gun control laws and crime rates. It is best to come up with policies for reducing guns ownership within the Texas community for the purpose of fighting crime rates.

## 10. Analysis : Arson

## To view total cases of arson by states, generate a pie chart through running the SAS codes in **Appendix 19** (proc template):



*Figure 12:* *Pie Chart for Arson*

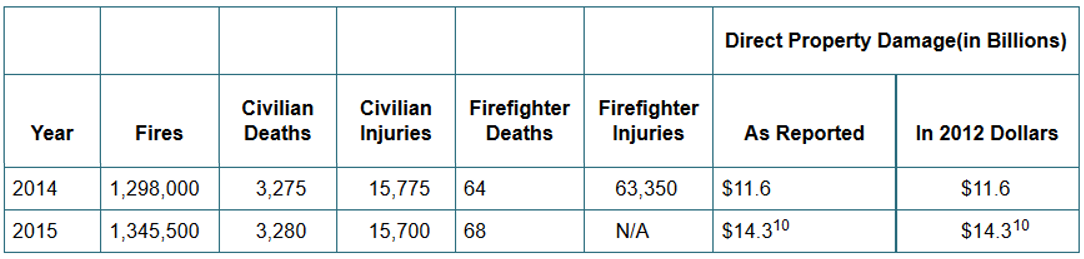
The pie chart above shows the number of arsons by the top 5 states. For both years 2014 & 2015, California has the highest number of arsons (2015: 2,175 cases, 2014: 2,014 cases), followed by Texas (2015: 1,088 cases, 2014: 1,035 cases), Illinois (2015: 328 cases, 2014: 345 cases), Florida (2015: 266 cases, 2014: 232 cases) and New York (2015: 145 cases, 2014: 86 cases).

The UCR Program defines arson as any willful or malicious burning or attempting to burn, with or without intent to defraud, a dwelling house, public building, motor vehicle or aircraft, personal property of another, etc. ("Arson").

Because of the nature of fire, which causes a wide spread of effects to nature, families and homes caught in fire, although the number of arson is the least among all categories of crimes, it causes more damage than the average crime. There are investigations that pointed out that arson has been under-reported as resources are not enough to cover all fire incidents, and there is a disconnection between fire-fighters and arson investigators (Thomas Hargrove).

Since the category Arson has the least data, and there are suspicions in the validation on these data, further exploration on the data set for Arson is not conducted through SAS. Instead, other sources of data is taken into this analysis on Arson, to facilitate a more complete view of the potential impact of the crime numbers that might be associated with Arson, for example the total fires reported or types of behavior associated with intentional fire setters.

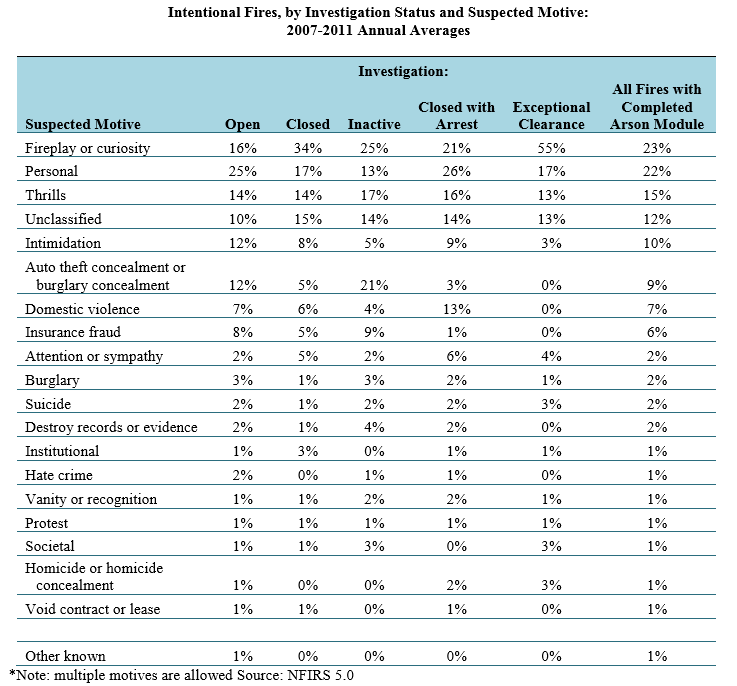
The following table shows the total fires reported in the US for year 2014 and 2015. The financial loss is $14.3 billion dollars in year 2015, with 3,280 deaths and 15,700 injuries. If arson is under-reported, the arsonists that have not been convicted would be able to roam freely, which might lead to second or third crime commitments and may lead to greater loss to the society and economy of the United States.



*Table 11:* *Total Fires, Deaths and Property Damage in the US*

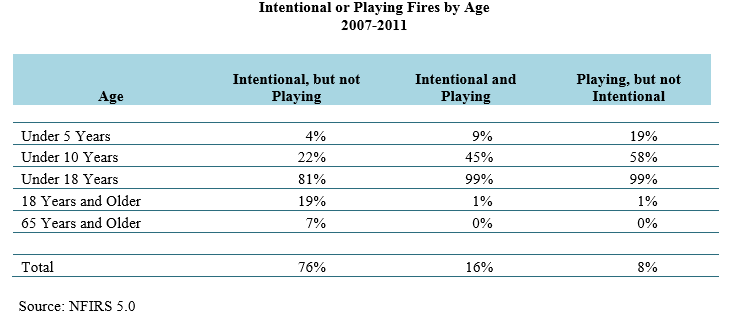
*("NFPA Statistics - The U.S. Fire Problem")*

The following table shows the suspected motives of starting intentional fires, which include curiosity, personal and thrills among the top 3, which makes up to more than 50% of the total motives. These motives suggest that it can be difficult for the authorities to track arsonist based on these subjective motives, which is hard to detect in individuals.



*Table 12:* *Intentional Fires Investigation and Suspected Motive (Campbell)*

From the intentional or fire playing category, most of the fire setters are below 18 years old, which are juvenile fire setters who might be suffering from some form of emotional disturbance.



*Table 13:* *Intentional or Playing Fires by Age (Campbell)*

**Recommendation:**

It is recommended for law enforcers and arsonist investigators to work together with juvenile psychologist or experts in emotional behaviors. This is because there is a relationship between fire setting and age, and these juvenile fire setters who are potential arsonists should be detected and educated through some form of behavior intervention. Those who come into contact with a potential arsonist might be alerted through various symptoms or observations highlighted by the juvenile psychologist and arsonist investigators. In this way, the public can be educated into protecting themselves from being exposed to arsonist and hopefully prevent the arsonist from further arsons commitments.

# 

# Conclusion

The analysis of property crime and violent crime by the top 5 states and its cities have provided some useful insights, particularly on violent crimes, where the influence of gun violence to the population heavily relates to the number of aggravated assault and murder cases. Crime prevention is advocated through increasing awareness of the public and tightening laws on gun controls on the neighboring states of the top crime state. Although this will take a lot of effort, especially from the central government bodies, ultimately these efforts will go into providing a better life for the families and children of the states where crimes are the highest.

While violent crimes call for government initiatives, property crimes can be prevented through the support of the private sector. Retail organizations in particular, play an important role in sponsoring programs to reduce larceny theft among civilians. The ethical concern on theft seems to be remedied best when the potential crime committers can be educated to agree to the terms of compliance in order to avoid getting caught.

The availability of data for analysis to derive useful insights, can help policy makers to target potential crime committers by age, location and behaviors. This is important for the general public because each individual might be limited in their own prevention efforts, as opposed to concentrated and planned efforts in targeted crime prevention demographics, which has the same aim, in helping residents to live in a better and safer environment.

# References

"Arson". *FBI*. N.p., 2010. Web. 13 Apr. 2017.

"Illinois Theft / Larceny Laws - Findlaw". *Findlaw*. N.p., 2017. Web. 12 Apr. 2017.

"Larceny-Theft". *FBI*. N.p., 2017. Web. 15 Apr. 2017.

"New York Assault And Battery Laws - Findlaw". *Findlaw*. N.p., 2017. Web. 17 Apr. 2017.

"NFPA Statistics - The U.S. Fire Problem". *Nfpa.org*. N.p., 2016. Web. 14 Apr. 2017.

"Table 23". *FBI*. N.p., 2017. Web. 15 Apr. 2017.

Barbaro, Michael. "Some Leeway For The Small Shoplifter". *Nytimes.com*. N.p., 2006. Web. 14 Apr. 2017.

Berman, Mark. "Chicago Has Tough Gun Laws, But People Can Still Buy Guns Elsewhere And Bring Them There". *Washington Post*. N.p., 2016. Web. 19 Apr. 2017.

Campbell, Richard. "NFPA Report - Intentional Fires". *Nfpa.org*. N.p., 2014. Web. 15 Apr. 2017.

Demarest, Erica. "Shoplifting Under $1,000 Will No Longer Be A Felony, Kim Foxx Says". *DNAinfo Chicago*. N.p., 2016. Web. 12 Apr. 2017.

Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From Data Mining To Knowledge Discovery In Databases". *AI Magazine* Fall (1996): 38-54. Print.

Fernandez, Manny. "Licensed To Carry: Handguns In Texas". *Nytimes.com*. N.p., 2016. Web. 20 Apr. 2017.

Goudie, Chuck. "Chicago's New Face Of Homelessness". *ABC7 Chicago*. N.p., 2015. Web. 12 Apr. 2017.

Kirk, Chris. "This Chart Shows That States With Tighter Gun Control Laws Have Fewer Deaths". *Slate Magazine*. N.p., 2015. Web. 20 Apr. 2017.

Knapp, Deborah. "Walmart Cracks Down On Shoplifting By Taking Justice Into Its Own Hands". *KENS*. N.p., 2016. Web. 12 Apr. 2017.

Schiller, Dane. "Texas Leads The Nation In Number Of Gun Background Checks". *Houston Chronicle*. N.p., 2015. Web. 20 Apr. 2017.

Smith, Julie, and Mitch Smith. "As Chicago Murder Rate Spikes, Many Fear Violence Has Become Normalized". *Nytimes.com*. N.p., 2016. Web. 13 Apr. 2017.

Spies, Mike, and Emily Fuhrman. "Watch How Chicago Gets Flooded With Thousands Of Crime Guns". *The Trace*. N.p., 2017. Web. 19 Apr. 2017.

Stine, Stephen. "New York Shoplifting Charges: Penalties And Defense | Criminal Law". *CriminalDefenseLawyer.com*. N.p., 2016. Web. 12 Apr. 2017.

Thomas Hargrove, Scripps News. "Arson In America: What's The Alarming Reality?". *WFTS*. N.p., 2013. Web. 12 Apr. 2017.

# APPENDICES

**Appendix 1:**

**SAS Code (import WORK.table1415)**

FILENAME REFFILE '/home/tp0446020/Table4/Table\_4\_20152014.xls';  
  
PROC IMPORT DATAFILE=REFFILE  
 DBMS=XLS  
 OUT=WORK.table1415;  
 GETNAMES=YES;  
RUN;  
  
PROC CONTENTS DATA=WORK.table1415; RUN;

**Appendix 2:**

**SAS Code (check missing values)**

proc means data=WORK.TABLE1415 chartype n nmiss vardef=df;

var Year Population Violent\_crime Murder Rape Robbery Aggravated\_assault

Property\_crime Burglary Larceny\_theft Motor\_vehicle\_theft Arson;

run;

**Appendix 3:**

**SAS Code (import WORK.Crime\_Rate\_by\_State)**

FILENAME REFFILE '/home/tp0446020/Table4/Crime\_rate\_by\_state.xlsx';

PROC IMPORT DATAFILE=REFFILE

DBMS=XLSX

OUT=WORK.Crime\_Rate\_by\_State;

GETNAMES=YES;

RUN;

PROC CONTENTS DATA=WORK.Crime\_Rate\_by\_State; RUN;

**Appendix 4:**

**SAS Code (****generate scatter plot for violent crime & property crime by state codes)**

proc sgplot data=WORK.CRIME\_RATE\_BY\_STATE;

scatter x=Property\_crime y=Violent\_crime / group=Year datalabel=State\_code

datalabelattrs=(size=7) transparency=0.0 name='Scatter';

xaxis grid;

yaxis grid;

run;

**Appendix 5:**

**SAS Code (import WORK.TOP5STATES)**

FILENAME REFFILE '/home/tp0446020/Table4/top5states.xlsx';

PROC IMPORT DATAFILE=REFFILE

DBMS=XLSX

OUT=WORK.TOP5STATES;

GETNAMES=YES;

RUN;

PROC CONTENTS DATA=WORK.TOP5STATES; RUN;

**Appendix 6:**

**SAS Code (generate pie chart for property crime by top 5 states)**

proc template ;

define statgraph WebOne.Pie;

begingraph;

layout region;

piechart category=State response=Property\_crime / group=Year groupgap=2%

start=90 datalabellocation=INSIDE;

endlayout;

endgraph;

end;

run;

proc sgrender template=WebOne.Pie data=WORK.TOP5STATES;

run;

**Appendix 7:**

**SAS Code** **(generate analysis for property crime)**

proc means data=WORK.TOP5STATES chartype mean std min max n vardef=df;

var Property\_crime;

class Year State;

run;

**Appendix 8:**

**SAS Code (drill down property crime in New York)**

proc sort data=WORK.TOP5STATES out=WORK.SORTTEMP;

by Year;

where State="NEW YORK";

run;

proc print data=WORK.SORTTEMP label noobs;

var State City Population Property\_crime Burglary Larceny\_theft Motor\_vehicle\_theft;

by Year;

run;

**Appendix 9: SAS Code (generate distribution of property crime by state)**

%macro DEHisto(data=, avar=, classVar=);

%local i numAVars numCVars cVar cVar1 cVar2;

%let numAVars=%Sysfunc(countw(%str(&avar), %str( )));

%let numCVars=%Sysfunc(countw(%str(&classVar), %str( )));

%let cVar1=%scan(%str(&classVar), 1, %str( ));

%let cVar2=%scan(%str(&classVar), 2, %str( ));

proc sql noprint;

select count(distinct &cVar1) into :nrows from &data;

quit;

proc sql noprint;

select count(distinct &cVar2) into :ncols from &data;

quit;

proc univariate data=&data noprint;

var &avar;

class &cVar1 &cVar2;

histogram &avar / nrows=&nrows ncols=&ncols

normal(noprint); run;

%mend DEHisto;

%DEHisto(data=WORK.TOP5STATES, avar=Property\_crime, classVar=Year State);

**Appendix 10:**

**SAS Code (drill down property crime in Illinois)**

proc sort data=WORK.TOP5STATES out=WORK.SORTTEMP;

by Year;

where State="ILLINOIS";

run;

proc print data=WORK.SORTTEMP label noobs;

var State City Population Property\_crime Burglary Larceny\_theft Motor\_vehicle\_theft;

by Year;

run;

**Appendix 11:**

**SAS Code (drill down property crime in Texas)**

proc sort data=WORK.TOP5STATES out=WORK.SORTTEMP;

by Year;

where State="TEXAS" AND Property\_crime>10000;

run;

proc print data=WORK.SORTTEMP label noobs;

var State City Population Property\_crime Burglary Larceny\_theft Motor\_vehicle\_theft;

by Year;

run;

**Appendix 12:**

**SAS Code (generate pie chart for violent crime by top 5 states)**

proc template ;

define statgraph WebOne.Pie;

begingraph;

layout region;

piechart category=State response=Violent\_crime / group=Year groupgap=2%

start=90 datalabellocation=INSIDE;

endlayout;

endgraph;

end;

run;

proc sgrender template=WebOne.Pie data=WORK.TOP5STATES;

run;

**Appendix 13:**

**SAS Code (generate analysis for violent crime)**

proc means data=WORK.TOP5STATES chartype mean std min max median n nmiss

vardef=df qmethod=os;

var Violent\_crime;

class Year State;

run;

**Appendix 14:**

**SAS Code (drill down violent crime in New York)**

proc sort data=WORK.TOP5STATES out=WORK.SORTTEMP;

by Year;

where State="NEW YORK";

run;

proc print data=WORK.SORTTEMP label noobs;

var State City Population Violent\_crime Murder Rape Robbery Aggravated\_assault;

by Year;

run;

**Appendix 15: SAS Code (generate distribution of violent crime by state)**

%macro DEHisto(data=, avar=, classVar=);

%local i numAVars numCVars cVar cVar1 cVar2;

%let numAVars=%Sysfunc(countw(%str(&avar), %str( )));

%let numCVars=%Sysfunc(countw(%str(&classVar), %str( )));

%let cVar1=%scan(%str(&classVar), 1, %str( ));

%let cVar2=%scan(%str(&classVar), 2, %str( ));

proc sql noprint;

select count(distinct &cVar1) into :nrows from &data;

quit;

proc sql noprint;

select count(distinct &cVar2) into :ncols from &data;

quit;

proc univariate data=&data noprint;

var &avar;

class &cVar1 &cVar2;

histogram &avar / nrows=&nrows ncols=&ncols

normal(noprint);

run;

%mend DEHisto;

%DEHisto(data=WORK.TOP5STATES, avar=Violent\_crime, classVar=Year State);

**Appendix 16:**

**SAS Code (drill down violent crime in Illinois)**

proc sort data=WORK.TOP5STATES out=WORK.SORTTEMP;

by Year;

where State="ILLINOIS";

run;

proc print data=WORK.SORTTEMP label noobs;

var State City Population Violent\_crime Murder Rape Robbery Aggravated\_assault;

by Year;

run;

**Appendix 17:**

**SAS Code (generate scatter plot for murder by city)**

proc sgplot data=WORK.TOP5STATES;

/\*--Fit plot settings--\*/

reg x=Violent\_crime y=Murder / nomarkers group=Year name='Regression';

scatter x=Violent\_crime y=Murder / group=Year datalabel=City

datalabelattrs=(size=7) transparency=0.0 name='Scatter';

xaxis grid;

yaxis grid;

run;

**Appendix 18:**

**SAS Code (drill down violent crime in Texas)**

proc sort data=WORK.TOP5STATES out=WORK.SORTTEMP;

by Year;

where State="TEXAS" AND Violent\_crime>10000;

run;

proc print data=WORK.SORTTEMP label noobs;

var State City Population Violent\_crime Murder Rape Robbery Aggravated\_assault;

by Year;

run;

**Appendix 19:**

**SAS Code (generate pie chart for arson by top 5 states)**

proc template ;

define statgraph WebOne.Pie;

begingraph;

layout region;

piechart category=State response=Arson / group=Year groupgap=2% start=90

datalabellocation=INSIDE;

endlayout;

endgraph;

end;

run;

proc sgrender template=WebOne.Pie data=WORK.TOP5STATES;

run;