

DATA ANALYTICAL PROGRAMMING (DAP)

CT050-3-M-DAP

UCMP1610DSBA

LECTURER: DR KALAI ANAND RATNAM

HAND IN DATE: 15TH MAY 2017

INDIVIDUAL ASSIGNMENT

NAME: PREMAWATHY KUMAR

STUDENT ID: TP028419

Table of Contents

[Table of Figures 2](#_Toc482568150)

[Acknowledgement 3](#_Toc482568151)

[1.0 Introduction 4](#_Toc482568152)

[1.1 Aim and Objectives 4](#_Toc482568153)

[1.2 Discussion on Data Set 5](#_Toc482568154)

[2.0 Methodology 7](#_Toc482568155)

[2.1 Data Understanding 7](#_Toc482568156)

[2.2 Data Preparation 8](#_Toc482568157)

[2.3 Analysis and Modelling 13](#_Toc482568158)

[2.3.1 To analyse overall crime in the United States (US) for year 2014 until mid-2015 13](#_Toc482568159)

[2.3.2 To analyse cities with highest crime ratio 13](#_Toc482568160)

[2.3.3 To calculate total frequency of crime based on selected 5 states 14](#_Toc482568161)

[2.3.4 To analyse highest murder occurrence cities and its state 16](#_Toc482568162)

[2.3.5 To analyse cities unsafe to travel alone 18](#_Toc482568163)

[2.3.6 To analyse highly populated (over 1 million) cities and states over crime rates 20](#_Toc482568164)

[2.4 Overall Evaluation 22](#_Toc482568165)

[Data Flow Diagram (DFD) 24](#_Toc482568166)

[Conclusion 25](#_Toc482568167)

[Reference 26](#_Toc482568168)

[Appendices 27](#_Toc482568169)

[SAS PROGRAM CODES 27](#_Toc482568170)

[SAS LISTINGS/ OUTPUTS 29](#_Toc482568171)

# Table of Figures

[Figure 1: Data Preparation 1 8](#_Toc482568172)

[Figure 2: Data Preparation 2 8](file:///C:\Users\Premawathy%20Kumar\Desktop\DAP.docx#_Toc482568173)

[Figure 3: Data Preparation 3 9](#_Toc482568174)

[Figure 4: Data Preparation 4 9](#_Toc482568175)

[Figure 5: Data Preparation 5 9](file:///C:\Users\Premawathy%20Kumar\Desktop\DAP.docx#_Toc482568176)

[Figure 6: Data Preparation 6 9](#_Toc482568177)

[Figure 7: Data Preparation 7 10](file:///C:\Users\Premawathy%20Kumar\Desktop\DAP.docx#_Toc482568178)

[Figure 8: Data Preparation 8 10](file:///C:\Users\Premawathy%20Kumar\Desktop\DAP.docx#_Toc482568179)

[Figure 9: SAS Libref 11](#_Toc482568180)

[Figure 10: SAS Import Data 11](#_Toc482568181)

[Figure 11: SAS crime\_data 11](#_Toc482568182)

[Figure 12: SAS sample code to merge columns 11](#_Toc482568183)

[Figure 13: SAS processed data 12](#_Toc482568184)

[Figure 14: SAS Total Crime Rate 13](#_Toc482568185)

[Figure 15: SAS Highest Crime Ratio Cities 13](#_Toc482568186)

[Figure 16: SAS Crime Summary 14](#_Toc482568187)

[Figure 17: SAS Pie chart representation 15](#_Toc482568188)

[Figure 18: SAS Cities High Murder cases 16](#_Toc482568189)

[Figure 19: SAS Average Mean for Murder 16](#_Toc482568190)

[Figure 20: SAS Bar Chart for Average Mean of Murder 17](#_Toc482568191)

[Figure 21: SAS Rape Summary 18](#_Toc482568192)

[Figure 22: Average Mean of Rape Cases 18](#_Toc482568193)

[Figure 23: SAS City Population More than 1 Million 20](#_Toc482568194)

[Figure 24: SAS State Population More than 1 Million 21](#_Toc482568195)

[Figure 25: Simple DFD to illustrate the flow of process 24](#_Toc482568196)

# Acknowledgement

The researcher would like to convey appreciation and gratitude to certain parties that have made the report a triumph. The researcher has achieved the goals of the report and research with large support and courage received from reliable people. First and foremost, the researcher would like to thank the lecturer, Dr Kalai Anand for his dedication and effort in giving the researcher guidance and knowledge about completing the report fulfilling the report requirements. The lecturer has always been willing to lend a helping hand and positive disposition motivated the researcher on completing the entire report.

Thank You.

# 1.0 Introduction

Crime trends could be related to demographics, cultural, economic, law enforcement trends and others. (Nevin, 2007). Gang violence are still norm in urban cities which underwrites crime rate in the United States (US) (James, 2015)

Numerous studies relate the common outset whereby irregularities in the country’s economic lead to higher crime rates. Impulsive recessions lead to unemployment, decline in household income, and poverty. This correlates to rise in crime rates whereby an individual is enforced towards circumstances of survival (Ajimotokin and Haskins, 2015). Unemployment due to lower economic status influences poverty that directs to criminal activities for sustainability (Grawert and Cullen, 2016)

Therefore, a research is done to analyse the crime rates and its occurrence in the US as well as to understand the possible cause of these crime incidences. This study is done to provide information to readers on to cities to avoid or to be more cautious during travels, household planning and others. Moreover, the suggestions are given to readers based on statics analysis done.

The Federal Bureau of Investigation (FBI) National Press Office released the crime statistics in United States (US) which estimates number of violent crimes has rose in 2015 and decrease in Property Crime with comparison to the year 2014 (Federal Bureau of Investigation, 2016). These stats are supported by a set of data by the FBI for further references to the study.

## 1.1 Aim and Objectives

The aim of this study is to analyse the provided data to study high crime rate occurrence area and create awareness, take precautionary actions to further decline the crime rates.

The objectives of these studies are to:

* To analyse overall crime in the United States (US) for year 2014 until mid-2015
* To analyse cities with highest crime ratio to
* To calculate total frequency of crime based on selected 5 states
* To analyse highest murder occurrence cities and its state
* To analyse cities unsafe for women to travel alone
* To analyse highly populated (over 1 million) cities and states over crime rates

## 1.2 Discussion on Data Set

The set of data displays the offenses reported to Law Enforcement in the United States. The data is organized by the states and its’ cities in which the population is at least over 100, 000 for each city. There are 43 states along with 260 cities listed in the dataset dated from the month January 2014 until June 2015. Most of the listed states are from the northern region of US.

The offenses are categorized per the frequent number of times the crime has happened whereby the classification are done for violent crime, murder, rape, robbery, aggravated assault, property crime, burglary, larceny-theft, motor vehicle theft and arson. In this context, the offenses of rape are divided into two sorts. This is because the rape incidents are categorized using the revised Uniform Crime Reporting (UCR) and legacy UCR definition of rape. Additionally, the data provided has some absent values of the crime occurrence in certain cities.

Based on reflection of the data, the total number of populations data tabulated are 78,920,616. The 2014 population figures are FBI estimates based on provisional data from the U.S. Census Bureau. The state of California has the utmost number of cities, 67, as followed by Florida, 30. The cities in California almost doubles the count of the cities in state of Florida.

As of the population of the cities, New York ranks the highest number of population with 8,473,938 for the year 2014. Erie, in the state of Pennsylvania marks the least in population consisting 100,403 numbers of civilian.

From the available data, property crime rates sets the highest figure of occurrence, followed by larceny theft with the approximate of 2.5 million and 1.7 million records respectively. Murder is reported as the least number of occurrences compared with other crimes in the listing. However, 6046 number of murders should be a matter of concern to diminish this figure.

In proportion to the case, many cities have witness plunge in crime rates by June 2015. Nevertheless, numbers of crime rates fluctuate around the mean in raise and decline of the figures.

Moreover, New York, in spite being the city with large population, the crime rates between year 2014 and 2015 has witnessed decline in crime rates. However, Los Angeles, with population over 3 million citizens, has drastic boost in crime rates for the documented duration with approximately 15 thousand added crime cases in 2015 compared with year 2014.

# 2.0 Methodology

A data mining method is crucial in discovering hidden valuable knowledge through analysis of large amount of data. As per this case study, Cross-Industry Standard Process for Data Mining (CRISP – DM) method is used as a base to complete the analysis. Thus, the following methods are done to produce finest analysis of the study.

Firstly, the objectives of the study are defined. Hence, this objectives and perspectives are converted into knowledge based on requirements of the researcher.

## 2.1 Data Understanding

The data set of study consist of 15 columns which consists of State, City, Year, Population and type of crimes and its count of cases registered by FBI. The table consists of 520 rows of data in which there are few incomplete and missing values. The obtained data has few footnotes to its data that needs to be measured before proceeding, which are:

* The 2014 population figures are FBI estimates based on provisional data from the U.S. Census Bureau.
* The population for the city of Mobile, Alabama, includes 55,819 inhabitants within the jurisdiction of the Mobile County Sheriff's Department.
* Complete January through June data for 2014 are not available. Thus, the values are to be replaced later.
* The FBI determined that the agency's data were underreported. Consequently, those data are not included in the report.
* Arson offenses are reported by the Toledo Fire Department. Therefore, several figures are not included in this report

## 2.2 Data Preparation

As the data is thoroughly studied, the researcher requires to prepare the data whereby then the data can be utilized to produce clean analysis. In this phases, noisy data are removed, missing data are replaced and several changes are done to the data set without violating the provided data values to obtain desired dataset. The outcome of data preparation phase is the final data set

The original data set has been renamed to ‘*crime\_dataset*’ and format is ensured to be ‘.csv’ to be compatible with SAS

Several modifications are done to the dataset to ensure enhanced analysis of the data. The changes are:

1. The additional information provided in the dataset has been removed as the information are not required as attributes of the data.

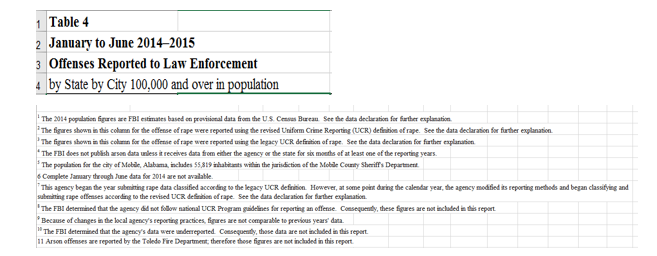


Figure 1: Data Preparation 1

1. ‘Rape (revised definition)’ and ‘Rape (legacy definition)’ attribute is merged as **‘Rape’**

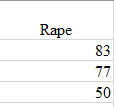


Figure 2: Data Preparation 2

1. The headers are relabelled and the column that consists year 2014 and 2015 has been named, ‘Year’



Figure 3: Data Preparation 3

1. The footnotes are removed

|  |
| --- |
|  |
| Figure 4: Data Preparation 4 |

1. The variables are set to numeric values to eliminate ‘,’

|  |
| --- |
| Figure 5: Data Preparation 5 |

1. Merged columns are removed

|  |  |
| --- | --- |
| Figure 6: Data Preparation 6 |  |

1. The missing values for population in the year 2015 is assumed unchanged from year 2014. Hence, the population values for 2015 is followed as per 2014

|  |
| --- |
| Figure 7: Data Preparation 7 |

1. The missing values of crime attributes has been replaced with 0. The missing values are due to confidentiality issues as mentioned above. Thus, these missing values are assumed as 0 to augment the analysis results.

|  |
| --- |
| Figure 8: Data Preparation 8 |

As the minor changes are done in Excel, the data is then uploaded in SAS to do further preparation. First, a folder ‘assignment1’ is created and the data set (crime\_data.csv) is uploaded to the folder. A library (DAP) is created and the data is imported to the library as follows:

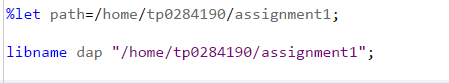


Figure 9: SAS Libref

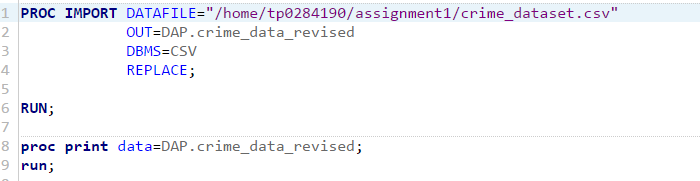


Figure 10: SAS Import Data

Then the imported data is sorted by year, and the output result is as follows:

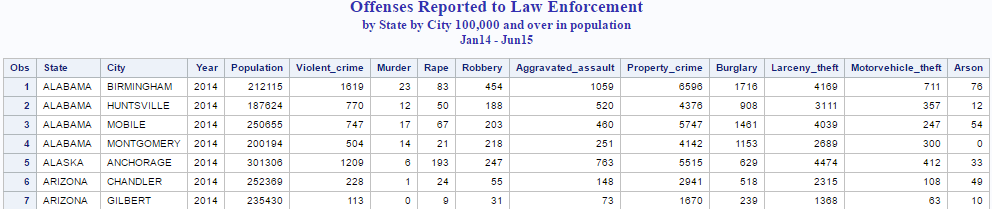


Figure 11: SAS crime\_data

The variables are then reduced whereby the column ‘Burglary’ and ‘Robbery’ are merged, and ‘Larceny Theft’ and ‘Motor Vehicle Theft’ are merged, classified as Stealing and Robbing respectively. This is because, both the variables define similar crime, hence it has been merged and renamed.

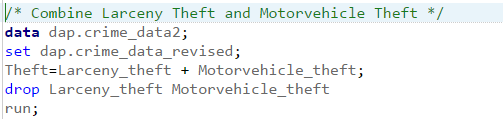


Figure 12: SAS sample code to merge columns

Next, the processed data has been reduced to 12 variables with 520 observations as follows:

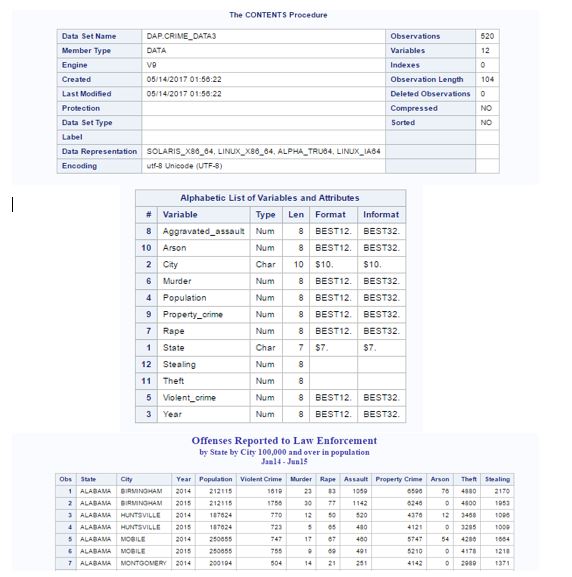


Figure 13: SAS processed data

## 2.3 Analysis and Modelling

This phase consists of thorough discussion on the objectives, analysis and outcome of the analysis. The analyses are closely discussed on how the output relates to the overall aim of the study. The followings are the objectives stated and its discussions:

### 2.3.1 To analyse overall crime in the United States (US) for year 2014 until mid-2015

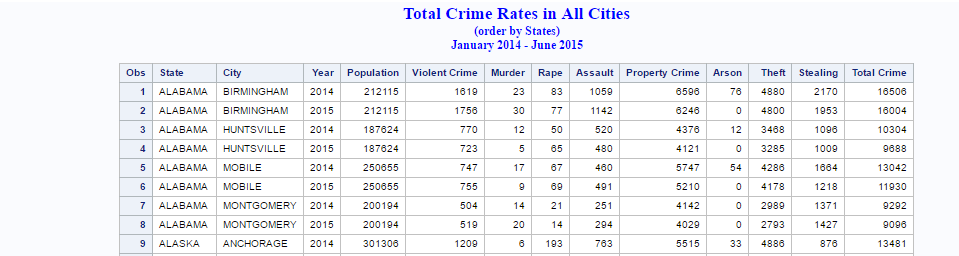


Figure 14: SAS Total Crime Rate

|  |  |
| --- | --- |
| Analysis: | The figure above shows tabulated data of the Total crime rates sorted by state and its cities. |
| Discussion: | Based on this analysis, the researcher can understand the overall crime occurrence for respective cities. Besides, comparison of crime rate between both years can be done to identify increase or decline of crime activities. |

### 2.3.2 To analyse cities with highest crime ratio

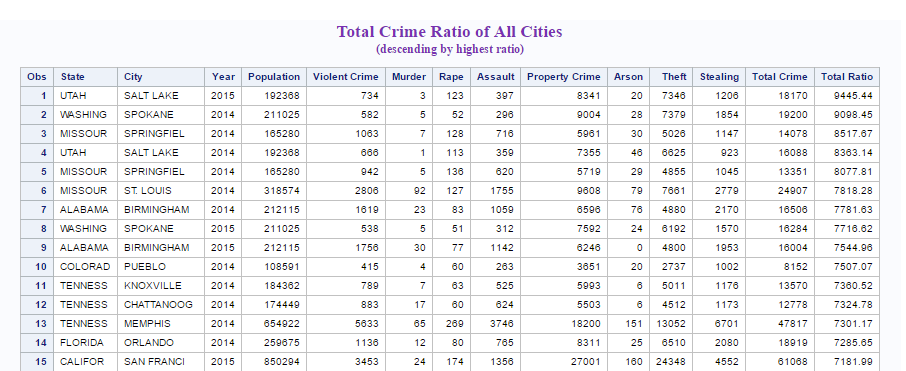


Figure 15: SAS Highest Crime Ratio Cities

|  |  |
| --- | --- |
| Analysis: | This analysis is done to view the overall ratio of total crime, sorted descending by highest crime ratio. |
| Discussion: | Based on this analysis, the researcher has selected 5 states to narrow the data towards states with highest crimes. Through this analysis, variables can be manipulated to obtain better results and information desired by the researcher.  In addition, the selected 5 states are California, Missouri, Texas, Utah and Washington.  These 5 states are selected based on the cities with high crime ratio. Moreover, states like Utah and Missouri has cities with top ratio in top 15 observations.  California is chosen due to its peak fraction of cities, followed by Texas. Finally, Washington is also chosen as a study attribute whereby Washington is the Capital of US. In spite being the capital, the crime rates in this state is no different from other states with high criminal activities. |

### 2.3.3 To calculate total frequency of crime based on selected 5 states

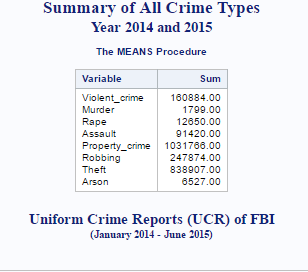


Figure 16: SAS Crime Summary

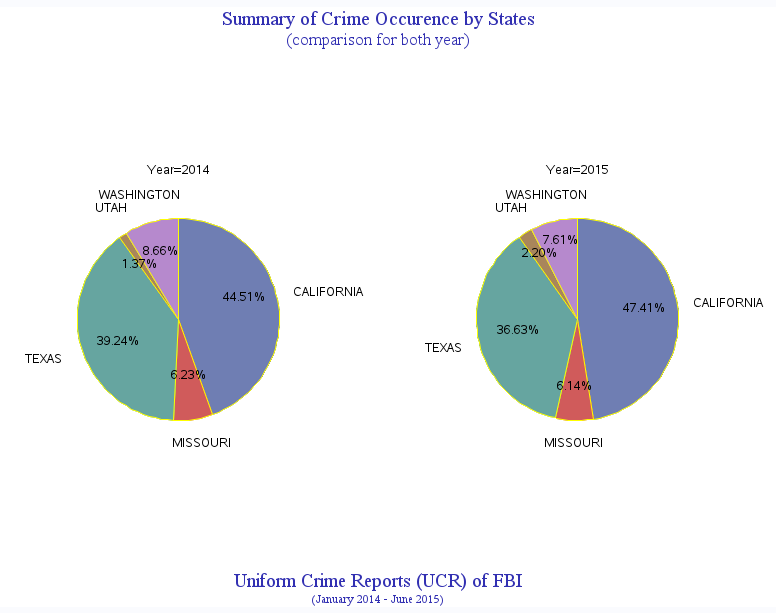


Figure 17: SAS Pie chart representation

|  |  |
| --- | --- |
| Analysis: | The analysis is done to calculate the total frequency of the selected states |
| Discussion: | The summary of all crime types for both years illustrates Property Crime has the maximum number of occurrence, followed by theft which includes both Larceny and Motor Vehicle theft. Based on the summary, a chart is produced to show comparison on the incline and decline of crimes in the following states.  California and Utah has increase in crime rates whereby the other states illustrate decline in crime rates.  However, these statistical observation does not rationalise overall crime rates will continue to rise or decline in upcoming years whereby other factors such as political, economic conditions may affect these statistics. |

### 2.3.4 To analyse highest murder occurrence cities and its state

Foremost, top 10 cities with high murder cases are sorted to compare the pattern of repetition of the cities.

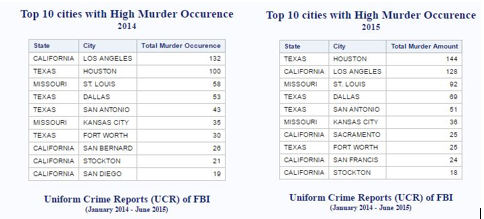


Figure 18: SAS Cities High Murder cases

Then based on this output, average occurrence in states is found as an overall analysis to recognize the state high murder rates.

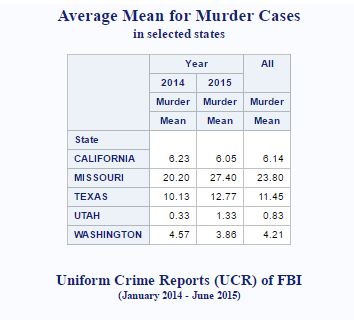


Figure 19: SAS Average Mean for Murder

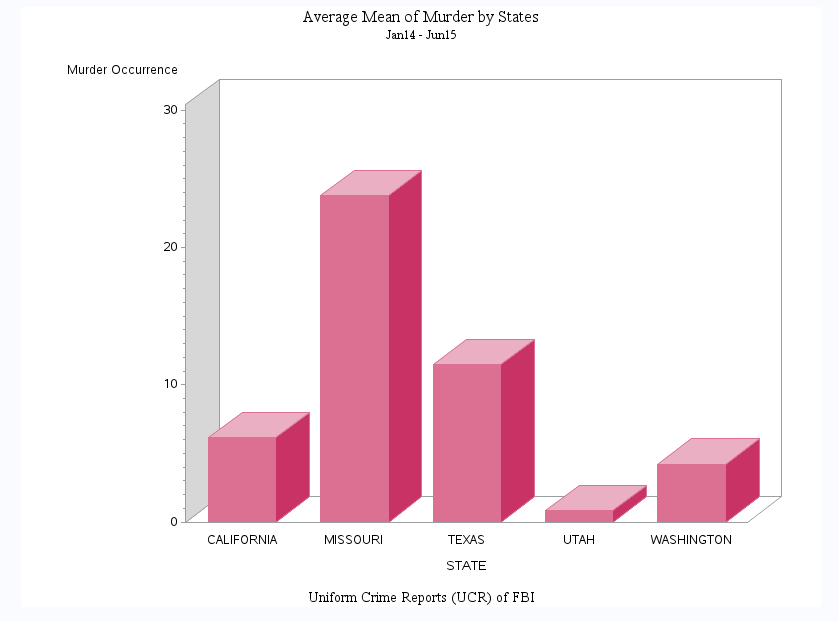


Figure 20: SAS Bar Chart for Average Mean of Murder

|  |  |
| --- | --- |
| Analysis: | This analysis could assist to identify factors that affect high murder activities in the states. |
| Discussion: | Missouri has the highest average mean of murder occurrence whereby it doubles the average mean in Texas. Based on external studies, St. Louis is known the ‘Most Dangerous City’ in the US due to its consistent increase in crime rate since 1960s.  St Louis is a metropolitan city. Studies suggests that factors that affect the rate and nature of crime are aspects such as degree of urbanization, economic changes and composition of age groups (Media, 2016) |

### 2.3.5 To analyse cities unsafe to travel alone

Comparison study was done for rape cases in both the year. About the figure below, Los Angeles and San Antonio has been the cities with highest rape crime consecutively.

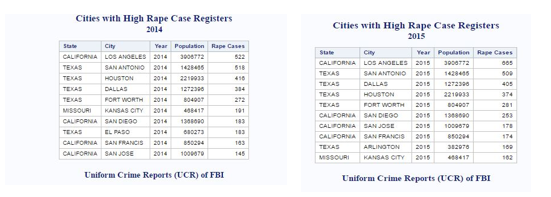


Figure 21: SAS Rape Summary

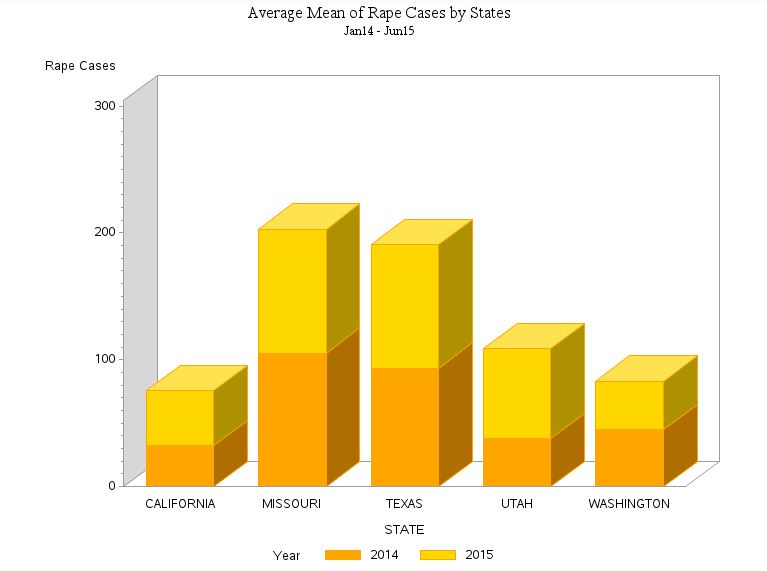


Figure 22: Average Mean of Rape Cases

|  |  |
| --- | --- |
| Analysis: | This analysis done is to counsel people to avoid travelling alone to these stated cities. |
| Discussion: | Based on the summary with 10 observations of rape occurrence in year 2014 and 2015, cities in Texas has cases with high occurrence of sexual assault. The findings are limited with justifications to identify the gender of the victims. However, based on the chart above, when average mean is calculated based on overall records of rape, Missouri and Texas has the highest records with minimal fraction of difference.  Based on external studies, in 2005, Texas enforced its law against sexual harassment. In spite enforcing strict law and order against sexual harassment, rape cases in Texas shows no decline in number of incidents reported. As a first world country, with globally available business opportunities, states like Texas should be enforced with law enforcement agencies and government bodies. Citizens and fellow immigrants should be protected from being the victim of sexual assault and of any crimes.  Next, as for Missouri, the state encompasses the highest number of crime occurrence for all crime types. Thus, Missouri is branded as state with top criminal activities. |

### 2.3.6 To analyse highly populated (over 1 million) cities and states over crime rates

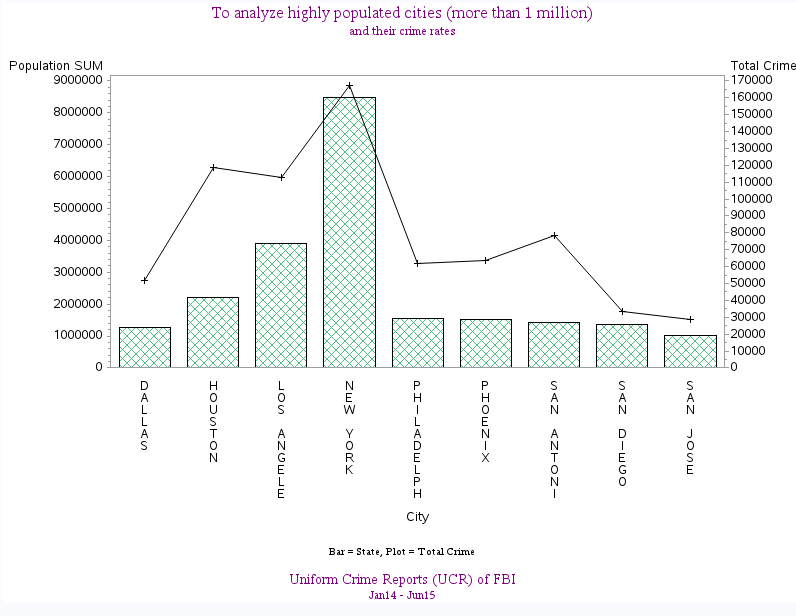


Figure 23: SAS City Population More than 1 Million

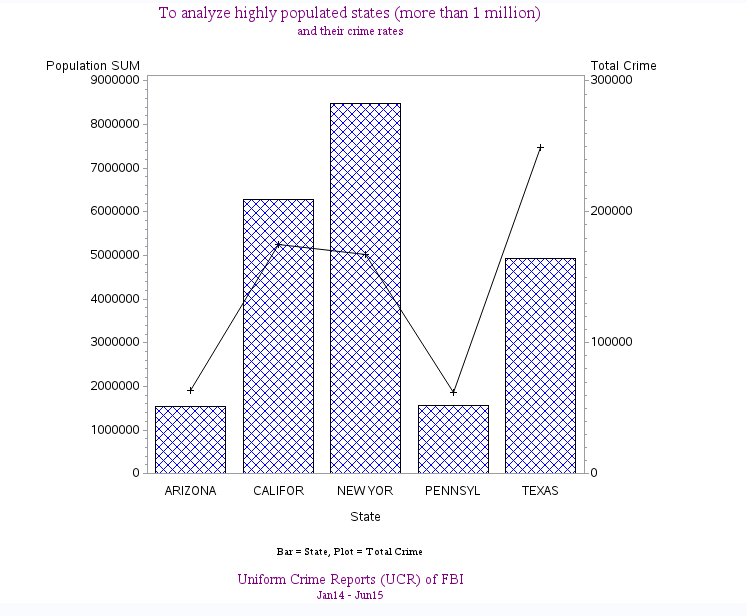


Figure 24: SAS State Population More than 1 Million

|  |  |
| --- | --- |
| Analysis: | This analysis done to see the relationship between the population of the cities and states with total crime rates respectively. |
| Discussion: | Based on the chart above, the total population in both cities and states as variable, does not affect the total crime rate occurrences. Therefore, the number of population is not a contributing factor to the crime activities in the US |

## 2.4 Overall Evaluation

Based on overall analysis of the study, it is arguable whether the crime rates will continue to decline in the upcoming year. This is because, the comparison was made based on only half year data of 2015 compared to full year data of 2014.

Though, the changes are noticeable where the total crime rates for both years has immense amount of difference in spite only having half a year record of 2015.

The cities in states with highest crime ratio should be monitored more meticulously to drive more changes in crime occurrence for forthcoming years. Based on mere observation of the data, these states with highest ratio of crime are the ones with average overall population. For instance, New York with the highest population rate was not listed in the top 15 cities with highest crime ratio. This explains that, population is not the responding variable to the crime rates in US and evidently proven based on analysis done.

And, Missouri as the state with high crime rate should be monitored more than the urban areas. The famous metropolitan city should be enforced with more officials, such as Sheriffs, police and other government bodies to monitor the concurrent crime activities in its state. Exclusively, the government should focus in cities like St. Louis to strategically reduce its crime rate.

Next, 5 states are chosen based on cities with high crime rates to further study factors that might influence the crime rate of these city. However, the findings are less meaningful where other details such as gender, immigrants, occupation and other details are not tabulated in the FBI records. Thus, the criminal activities that takes place in these cities could not be justified on the root cause of these crimes. Different environmental factors could be a contributing element towards the crime rate fluctuations such as political influence and economic stability.

Nevertheless, the highlighted cities and states are to be taken into consideration on how crime rates can be condensed with knowledge and information of these analysed data set.

As a suggestion for preventive efforts to diminish the crime rates, to the government, more CCTV in remote area should be placed to monitor the activities happening around those areas.

In addition, more policemen or active law bodies should drive frequent visits to areas with less human interaction and remotely away from the city. Policemen should also govern criminal gangs and their actions to monitor criminal acts by the members of the gang,

As an individual with social responsibility, if unusual behaviours or events befalls around the housing zones, the public should inform the officers as a preventive measure before any illegal crime occurs.

The public are also responsible of their own and property safety. Triggering alarms could be set to alert the presence of burglary or property crime. Besides, items like pepper spray can be kept as a self-defence measure during hazardous times. Early nurture and teachings of morale values are to be provided to children since young. Finally, martial arts skills would be aided benefit for young children to protect themselves from being victimised.

# Data Flow Diagram (DFD)

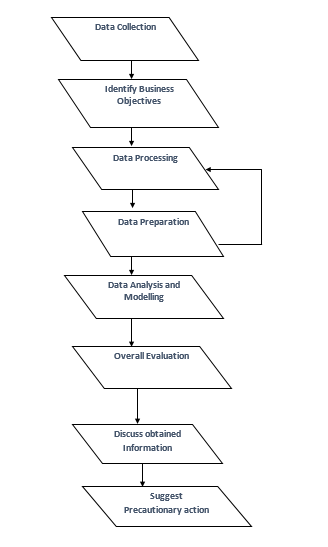


Figure 25: Simple DFD to illustrate the flow of process

# Conclusion

The aim and objectives of this research is met whereby based on the analysis, precautionary actions could be taken to not become the victim of these criminal acts. The data are prepared and visualized using Microsoft Excel and SAS to transform the raw data to meaningful information. Furthermore, with additional intense study on the overall environment and economic state of the country, factors influencing the crime rates could be identified.

Besides, available online statistics from previous years has showed decline in crime rates in the US. However, several cities still experience high violent crimes and property crimes whereby the crime occurrence varies across the nation. Thus, no sufficient evidence concludes these crime rates will persist in upcoming years.

.

# Reference

Ajimotokin, S. and Haskins, A. (2015). The Effects of Unemployment on Crime Rates in the U.S. 1, pp.4-10.

Federal Bureau of Investigation. (2016). FBI Releases 2015 Crime Statistics. [online] Available at: https://www.fbi.gov/news/pressrel/press-releases/fbi-releases-2015-crime-statistics [Accessed 3 May 2017].

Gelb, A. (2016). National Imprisonment and Crime Rates Continue to Fall. [online] Pewtrusts.org. Available at: http://www.pewtrusts.org/en/research-and-analysis/fact-sheets/2016/12/national-imprisonment-and-crime-rates-continue-to-fall [Accessed 2 May 2017].

Grawert, A. and Cullen, J. (2016). Crime in 2015: A Final Analysis. BRENNAN CENTER FOR JUSTICE, pp.4-6.

James, N. (2015). Is Violent Crime in the United States Increasing?. Congressional Research Service, pp.3-5.

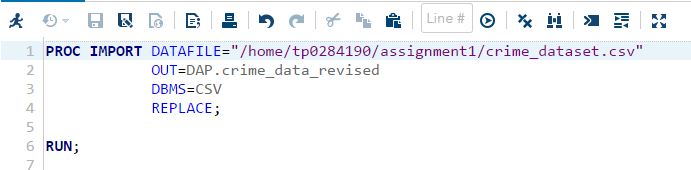
Media, T. (2016). St. Louis named murder capital of the U.S.. [online] FOX2now.com. Available at: http://fox2now.com/2016/02/29/the-30-murder-capitals-in-the-u-s/ [Accessed 8 May 2017].

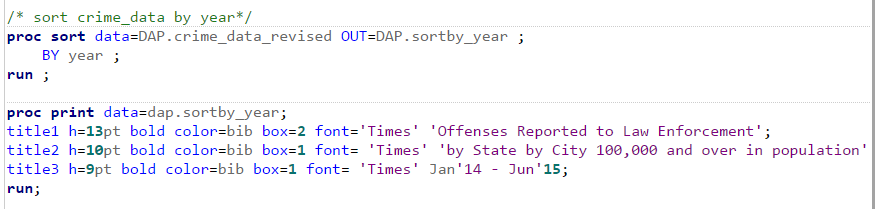
Nevin, R. (2007). Understanding international crime trends: The legacy of preschool lead exposure. Environmental Research, 104(3), pp.315-336.

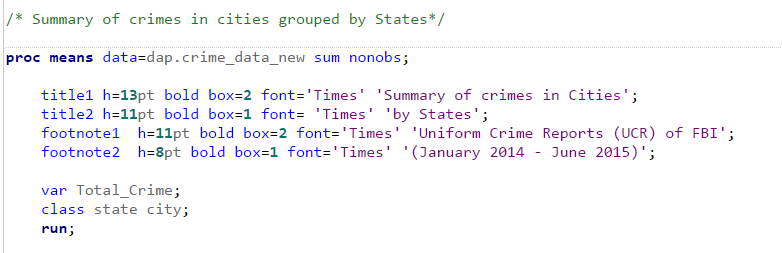
# Appendices

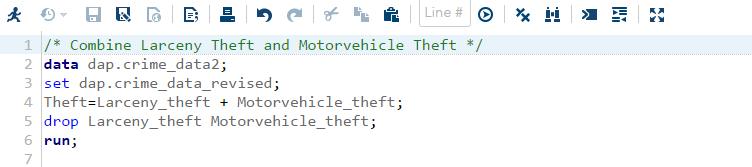
## SAS PROGRAM CODES

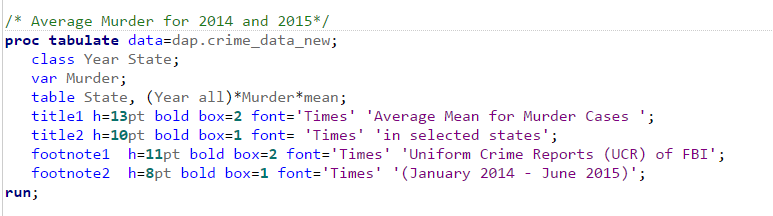
Few Sample codes of SAS Program

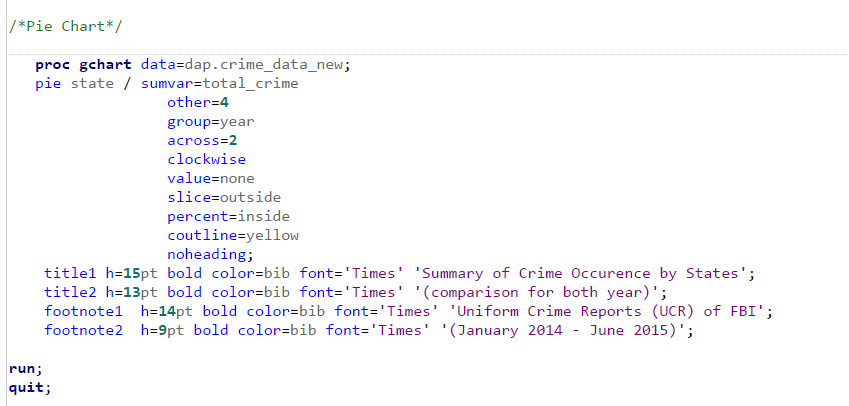












## SAS LISTINGS/ OUTPUTS