Acknowledgement

"*Gratitude is the attitude of the heart!*"

I thank my family for their unending love and support to me. I am grateful to the almighty for throwing light on my path when everything seems bewildered. I am grateful to our knowledgeable professor *Dr. Kalai Anand Ratnam* for helping us all the way through Data Analytical Programing module and for bearing with us. Finally, I thank my dear friends and batch mates for always keeping things interesting in our journey of learning.

- Rakesh Jalla
  TP045722

Contents

INTRODUCTION

This document discusses about the population census of various states in the United States of America (USA) and details on number of property crimes and violent crimes that occurred in these states. We shall carry out a comprehensive analysis on the preliminary data sets provided by the Uniform Crime Reporting department of the Federal Bureau of Investigation (FBI). The preliminary data sets provided for the analysis comprise of crime reports for the first 6 months of the years 2014 and 2015 for the cities and counties with a population estimate of 100,000 and above. To get an insight of the preliminary data sets provided, we shall apply certain statistical models and well defined data analytical methods to tabulate and picturize the outcomes through which we shall be able to formulate our recommendations which would be used by the FBI, Police or any law enforcing departments to identify the key reasons for crime and implement an effective plan to prevent any expansion of crime rates in the United States of America.

PRILIMINARY REPORT

The preliminary data sets considered for the analysis is obtained from the Federal Bureau of Investigation department through Uniform Crime Reporting program. There are 4 tables provided which are described as below.

Table1: Table1 provides the percentage change in the various crimes known to law enforcement by population groups for the first six months of the years 2014 and 2015.

Table2: Table2 shows the percentage change in crimes known to law enforcement by regions for the first six months of the years 2014 and 2015.

Table3: This table provides the percentage change in offenses known to law enforcement for the nation for the first 6 months in consecutive years from 2011 through 2015.

Table4: This table provides the number of offenses known to law enforcement for the first 6 months of 2014 and 2015 in cities with populations of 100,000 and over.

The primary focus for our analysis will be on Table4. Table4 has crime details for the first 6 months of the years 2014 and 2015 for the cities with population estimates of more than 100,000. The crime details can be divided into three groups, Violent Crime, Property Crime and Arson. Even though Arson is also a property crime, this is considered differently since the data collection procedure for Arson was not by UCR but rather the data is collected by Toledo fire department. Crimes that categorize under Violent Crime in Table4 are Murder, Rape (Legacy Definition), Rape (Revised Definition) and Robbery. Crimes that categorize under Property Crime are Burglary, Larceny-theft and Motor Vehicle theft.

The population census for the year 2015 is entirely missing from the preliminary data Table4. Table4 has population and crime data for 43 states and cities of these states for which the population estimate is more than 100,000. There is about 5% of Arson data missing from the data set. Some basic points we notice from the data are as below considering the data only for the year 2014 since the population census for the year 2015 is not available yet in the data set.

- State of **California** estimates the highest population for the year 2014 from the data set with a population estimate of 17,822,767.
- The State of **Montana** estimates the lowest population from the data set with a population of 110,245.
- The cities **Tucson** from **Arizona** State, **Honolulu** from **Hawaii** State, **Tyler** from **Texas** State, **Provo** from **Utah** State and **West Valley** from **Utah** State have complete missing data from January through June for the year 2014.

- The cities **Thornton** from **Colorado** State and **Toledo** from **Ohio** State have their data collected by the agency not following the UCR program guide lines as determined by the FBI.
- The **Ohio** State, **Louisville Metro** from **Kentucky** and **Bellevue** from Washington State's data is not comparable to the previous year's data as the agency has changed the reporting practices as determined by the FBI.

The overall violent crime has increased for about 1.7% for the year 2015 in the first six months compared to the same period in the year 2014 making this one of our prime concerns or one of the major objectives in our analysis. We would try to find out which states contributed to the increased rate in crime and possible variables contributing to this change.

The overall property crime has decreased by 4.2% for the year 2015 in the first six months compared to the same period in the year 2014. We shall analyze which states made maximum contribution for the decrease in the crime rate and the possible reasons for the decrease.

From Table2 we can observe the percentage changes in the crime comparing the first six months of the years 2104 and 2015 by region. The Northeast region displays a comparable decrease in its crime rates standing a top and the West region shows quite a rise in the crime rate.

OBJECTIVES

- **Objective 1**: To find the top 2 states that contributed for rise in the violent crime rate for the first 6 months in the year 2015 compared to the year 2014.
- **Objective 2**: To find the top 2 states that contributed for the fall in the property crime rate for the first 6 months of the year 2015 compared to 2014.
- **Objective 3**: To summarize the violent crimes, property crimes and arson data for the selected sample states from the entire population.

- **Objective 4**: To find the safest and most dangerous cities in terms of violent crimes among the chosen states for the years 2014 and 2015.
- **Objective 5**: To find the safest and the most dangerous cities in terms of property crimes among the chosen states in the years 2014 and 2015.
- **Objective 6**: To categorize all the states in the Unites States of America by crime rate for the years 2014 and 2015.

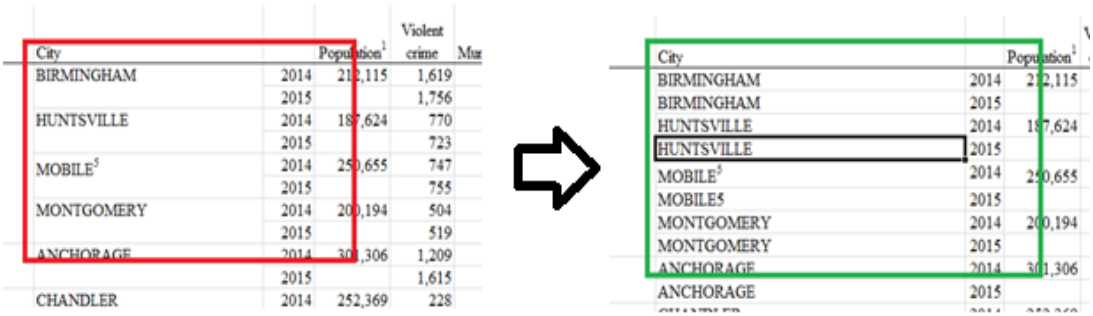
METHODOLOGY

Data Munging and Transformation

Before we consume the raw data for our analysis, we need to clean, re-format, re-arrange or re-structure the raw data to enable it to be analyzed using any analytical tool. For this report the preliminary data set **Table 4** will alone be considered for data munging.

Steps for Data Munging:

- Population census for the year 2015 is entirely missing from the provided preliminary data set. This data for the year 2015 is very essential for our analysis and thus needs to be made available. The population estimates for the first six months of the year 2015 is acquired from the U.S. census bureau site ” https://www.census.gov/data.html”.
- Un-merge the cells for the column ‘City’ and copy the City name to the new cells corresponding to the year 2015 as demonstrated below,



City		Population[1]	Violent crime	Mur
BIRMINGHAM	2014	212,115	1,619	
	2015		1,756	
HUNTSVILLE	2014	187,624	770	
	2015		723	
MOBILE[5]	2014	250,655	747	
	2015		755	
MONTGOMERY	2014	200,194	504	
	2015		519	
ANCHORAGE	2014	301,306	1,209	
	2015		1,615	
CHANDLER	2014	252,369	228	

City		Population[1]
BIRMINGHAM	2014	212,115
BIRMINGHAM	2015	
HUNTSVILLE	2014	187,624
HUNTSVILLE	2015	
MOBILE[5]	2014	250,655
MOBILE5	2015	
MONTGOMERY	2014	200,194
MONTGOMERY	2015	
ANCHORAGE	2014	301,306
ANCHORAGE	2015	

- City name must be copied in each row as when we read the data from our analytical tool (SAS) it reads as empty string if we did not perform the above step.
- Un-merge the cells for the column ‘State’ and copy the correct state name to its corresponding city as demonstrated below.

	State	City
6	ALABAMA	BIRMINGHAM
7		BIRMINGHAM
8		HUNTSVILLE
9		HUNTSVILLE
10		MOBILE[5]
11		MOBILE5
12		MONTGOMERY
13		MONTGOMERY
14	ALASKA	ANCHORAGE
15		ANCHORAGE

	State	City
6	ALABAMA	BIRMINGHAM
7	ALABAMA	BIRMINGHAM
8	ALABAMA	HUNTSVILLE
9	ALABAMA	HUNTSVILLE
10	ALABAMA	MOBILE[5]
11	ALABAMA	MOBILE5
12	ALABAMA	MONTGOMERY
13	ALABAMA	MONTGOMERY
14	ALASKA	ANCHORAGE
15	ALASKA	ANCHORAGE
16	ARIZONA	CHANDLER

This step is necessary as the tool reads the record as empty if we did not perform the above step.

- Column name for the column year is empty. Enter the column name ‘Year’.
- All the numerical data in the data set needs to be in general number format by removing separators.
- Insert a new column ‘State_Code’ between State and City columns. Values for states codes for the states are referred from “https://www.infoplease.com/state-abbreviations-and-state-postal-codes”. State Code is needed as this is an important parameter to generate reports for Objective 7.

- Un-employment rate by state for the years 2014 and 2015 data is also introduced in Table4 for the purpose of our analysis.
- 5% of Arson data is missing. We consider mean replacement of the missing data for Arson since the missing data percentage is less than 10%. Missing data is replaced with '29'.
- Rename the column names to suite the SAS variable type declarations. Re-named column names are as per the below table.

Current Column Name	New Column Name
Population1	Population
Violent Crime	Violent_Crime
Rape (revised definition)2	Rape_Revised
Rape (legacy definition)3	Rape_Legacy
Aggravated assault	Aggravated_Assult
Property Crime	Property_Crime
Larceny-theft	Larceny_Theft
Motor Vehicle Theft	Motor_Vehicle_Theft
Arson4	Arson

1.1 Data Upload

- Table4.xslx is uploaded to the cloud version of SAS where we perform all the required analysis on the data set.

  Code to import the data from Table 4 into SAS data sets looks as below,

```
/** Import an XLSX file.  **/

PROC IMPORT DATAFILE="/home/tp0457220/DAP_Assignment_Code/Table_4"
            OUT=WORK.TABLE4 DBMS=XLS
            REPLACE;
            SHEET = "Table4"; /*Name of the sheet to import the data*/
            GETNAMES=YES; /*This parameter considers the data as column names*/
            NAMEROW=5;
            DATAROW=6;
            ENDROW=525;
            ENDCOL=Q;
RUN;
```

- The above code for importing the excel data into SAS data set is available in the file "***ImportData.sas***".
- PROC CONTENTS output on the imported data set as below,

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
11	Aggravated_Assult	Num	8	BEST17.		Aggravated_Assult
16	Arson	Num	8	BEST9.		Arson
13	Burglary	Num	8	BEST9.		Burglary
3	City	Char	44	$44.	$44.	City
14	Larceny_Theft	Num	8	BEST9.		Larceny_Theft
15	Motor_Vehicle_Theft	Num	8	BEST9.		Motor_Vehicle_Theft
7	Murder	Num	8	BEST9.		Murder
5	Population	Num	8	BEST11.		Population
12	Property_Crime	Num	8	BEST9.		Property_Crime
9	Rape_Legacy	Num	8	BEST13.		Rape_Legacy
8	Rape_Revised	Num	8	BEST13.		Rape_Revised
10	Robbery	Num	8	BEST9.		Robbery
1	State	Char	30	$30.	$30.	State
2	State_Code	Char	30	$30.	$30.	State_Code
17	Unemployment_Rate_By_State	Num	8	BEST10.		Unemployment_Rate_By_State
6	Violent_Crime	Num	8	BEST7.		Violent_Crime
4	Year	Num	8	BEST5.		Year

- Certain States and Cities are having numbers and special characters in their names. These names are corrected to remove the numbers and the special characters using the SAS COPRESS function.
- Code for the same can be referred from Data_Cleaning.sas

```
/*SAS code to remove numbers and special
characters from State and City names*/
DATA work.table4;
    set work.table4;
    State = compress(State,',1234567890');
    City = compress(City,',1234567890');
RUN;
```

- Rape_Revised and Rape_Legacy columns are combined as a single column 'Rape' to represent the Rape data. Code for this is available from Data_Cleaning.sas. We have concatenated the data in this case as it is evident from the data that Rape_Revised and Rape_Legacy did not co-present for even a single record, If in the latter case, we would need to replace the missing values to 0 and then add the rape details into a single column.

```
/*SAS Code to merge Rape_Legacy and Rape_Revised to
single variable rape*/
DATA work.Table4;
    set work.Table4;
    Rape = catt(Rape_Revised,Rape_Legacy);
    Rape = compress(Rape,'.');
    drop Rape_Revised Rape_Legacy;
RUN;
```

- After all the data cleaning and transformation is completed, we have exported the final data set into work folder as CRIME_DATA.xslx.
- New data set CRIME_DATA.xslx will be used for our analysis going ahead.

ANALYSIS

Data analytics plays a key role in getting insights on the data streams stored or collected. Understanding the data and recognizing the patterns in the data helps us to get an insight in the behavior of the object or topic in question. Thus knowing the behavior, we will be enabled to optimize processes, recognize process and business errors, introduce action or correction plans or make business predictions and also reduce business risks by a great extent. Performing a comprehensive analysis on the

preliminary data set provided to us regarding the crime data in the United States for the first six months of the years 2014 and 2015 will enable us to understand the crime patterns across different states and hence apply corrective measures to reduce the crime rates in the cities by allocating suitable budgets, appropriate educational reforms or reviews of certain policies depending on the crime to be targeted.

OBJECTIVE 1:

To find the top 2 states that contributed for rise in the violent crime rate for the first 6 months in the year 2015 compared to the year 2014.

Finding the top 2 states as mentioned above would help us recognize as to which states it would be necessary for the law enforcement agencies to focus and try reducing the violent crimes by identifying the probable reasons. For the purpose of our objective we will only find the top 2 states.

Methodology:

To find the states as mentioned in our Objective 1, we need to calculate the crime rate of violent crimes by state for the years 2014 and 2015. We need to filter all the states where there is a raise in the crime rate.

In the filtered data set, top 2 states with maximum difference in the crime rate between the years 2014 and 2015 represent the output of our objective 1.

SAS code references for calculating the crime rate grouped by state can be found in the below files,

1. Ob1VC2014.sas
2. Ob1VC2015.sas
3. Ob1final.sas

The results for the Objective1 can be viewed from the file Ob1final-results.pdf.

Report for the Objective 1 is as below,

United States
Violent Crime Report 2014-2015
****Crime Rate per 100,000**

Obs	State	Violent Crime Rate 2014	Violent Crime Rate 2015	% Increase
1	ALASKA	401.253	540.685	25.7880
2	ARIZONA	167.028	217.157	23.0842
3	IOWA	226.308	286.440	20.9927
4	MONTANA	176.879	211.313	16.2953
5	INDIANA	201.506	234.337	14.0100
6	KANSAS	286.885	332.027	13.5958
7	MISSOURI	534.878	615.121	13.0451
8	NEW MEXICO	423.530	481.291	12.0013
9	NORTH CAROLINA	242.057	273.105	11.3684
10	RHODE ISLAND	235.670	265.057	11.0871
11	CALIFORNIA	203.335	226.244	10.1254
12	NEVADA	217.299	241.758	10.1169
13	MARYLAND	639.762	679.265	5.8155
14	MISSISSIPPI	411.890	436.505	5.6390
15	SOUTH CAROLINA	211.925	222.487	4.7474
16	NEBRASKA	164.081	169.823	3.3812
17	ALABAMA	427.939	439.589	2.6501
18	ILLINOIS	361.060	368.239	1.9495
19	OKLAHOMA	357.180	361.920	1.3095
20	TEXAS	370.626	366.997	-0.4944

There are 22 states that has an increase in the number of violent crimes committed in the year 2015 compared to the year 2014.

The above report shows the percentage increase in the violent crime rate in the year 2015 compared to the year 2014 by state, and arranged in the descending order of percentage increase in the violent crime rate.

Hence from the above report it is very evident that the states "**Alaska**" and "**Arizona**" have maximum increase in the violent crime rate in the year 2015 with an increase of 25.788% and 23.08% respectively.

Note:

- The state of "**Utah**" and "**Hawaii**" are not considered for calculating the percentage increase in the violent crime rate as more than 50% of data is not available for this state in the year 2014.
- Violent crime rate is calculated per 100,000.

OBJECTIVE 2:

To find the top 2 states that contributed for the fall in the property crime rate for the first 6 months of the year 2015 compared to 2014.

Achieving this objective would help us to idealize the reasons for the fall of the property crime rate and try implementing action plans in other states of United States of America to achieve similar situations in the states with higher property crime rate.

To find the states as mentioned in our objective 2, we need to find out the property crime rate for the first six months of the year 2014 and compare it with property crime rate in the year 2015 to find out the percentage decrease. The top 2 states with the highest percentage decrease represent the outcome of our objective 2.

Methodology:

We shall find out the property crime rate of each state for the year 2014 and 2015. Since there is a decline in the property crime rate as evident from Table 1. Crime rates in 2014 would be more than in 2015. Hence we calculate percentage decrease in the violent crime rate and sort the data in the descending order to achieve our objective.

SAS Code for the above objective can be referred from the files below,

1. Ob2PC2014.sas
2. Ob2PC2015.sas

3. Ob2final.sas

Reports can be respectively viewed from the documents below,

1. Ob2PC2014-results.pdf
2. Ob2PC2015-results.pdf
3. Ob2final.pdf

Final report for objective 2 is as below,

United State
Property Crime 2014-2015
****Crime rate per 100,000**

Obs	State	Property crime rate 2014	Property crime rate 2015	% Decrease
1	MICHIGAN	1688.42	1323.86	21.5921
2	MISSISSIPPI	2806.08	2412.20	14.0365
3	RHODE ISLAND	1729.74	1493.25	13.6721
4	WISCONSIN	1670.24	1447.87	13.3135
5	WASHINGTON	2907.42	2549.06	12.3255
6	COLORADO	1669.56	1505.57	9.8226
7	MASSACHUSETTS	1372.21	1246.75	9.1430
8	OKLAHOMA	2132.56	1939.44	9.0562
9	VIRGINIA	1396.03	1272.61	8.8407
10	PENNSYLVANIA	1539.56	1413.52	8.1865
11	SOUTH CAROLINA	1798.60	1653.71	8.0557
12	KENTUCKY	1899.41	1749.54	7.8902
13	ILLINOIS	1393.81	1285.96	7.7376
14	OREGON	2069.83	1910.62	7.6922
15	INDIANA	1912.85	1771.12	7.4091
16	TEXAS	1895.70	1756.54	7.3408
17	CONNECTICUT	1456.50	1359.68	6.6477
18	NEW YORK	823.03	768.62	6.6103
19	ALABAMA	2452.54	2296.45	6.3643
20	FLORIDA	1875.81	1777.30	5.2515

There are a total of 29 states with a noticeable decrease in the property crime rate for the year 2015 compared to the year 2014 in which the state of "**Michigan**" stands a top with the highest percentage decrease in the property crime rate by 21.59% followed by "**Mississippi**" with a decrease of 14.03%.

Thus it can be reasoned that there is quite an impact of effective administrations by law enforcement groups that try to reduce the crime in the states of "**Michigan**" and "**Mississippi**".

Note:

- The states "**Utah**" and "**Hawaii**" are not consider for the analysis on objective as both the states have more that 50% of data missing.
- Property crime rate is calculated per 100,000.

SAMPLE SELECTION

It is not possible to analyze the entire population for finding results hence we apply our analytics to significant subset of the population and can infer the results to the entire population as a whole. From our preliminary data set, three states covering most number of cities or having maximum population would be considered significant. To find out the three states to match our requirement we shall use the SAS code below.

```
/*Create a data set for the year 2014 alone*/

DATA WORK.CRIME_DATA_2014;
    SET WORK.CRIME_DATA;
    WHERE YEAR=2014;
RUN;

PROC PRINT DATA=WORK.CRIME_DATA_2014;
RUN;
```

```
/*
 * Create a data set with state name and number of cities
 * and sort the data set with descending order
 * of number of cities
 */
PROC SQL;
    CREATE TABLE WORK.SAMPLE_2014 AS
    SELECT STATE,
           COUNT(STATE) AS NO_OF_CITIES
    FROM WORK.CRIME_DATA_2014
    GROUP BY STATE
    ORDER BY NO_OF_CITIES DESCENDING;
QUIT;


/*
 * print the first three observations as the selected
 * sample
 */
PROC PRINT DATA=WORK.SAMPLE_2014 (OBS=3) LABEL;
    LABEL NO_OF_CITIES='No of Cities';
RUN;
```

The output of the above code and our chosen sample for performing our analytics is as below,

Obs	State	No of Cities
1	CALIFORNIA	66
2	TEXAS	30
3	FLORIDA	21

Hence our chosen sample states for objectives 3-7 are "**California**", "**Texas**" and "**Florida**".

OBJECTIVE 3:

To summarize the violent crimes, property crimes and arson data for the selected sample states from the entire population.

The below table describes a brief details about population and various crime details for the states "**California**", "**Texas**" and "**Florida**" for the years 2014 and 2015.

US: California, Florida & Texas
Crime details over the years 2014 & 2015
****Crime rate per 100,000**

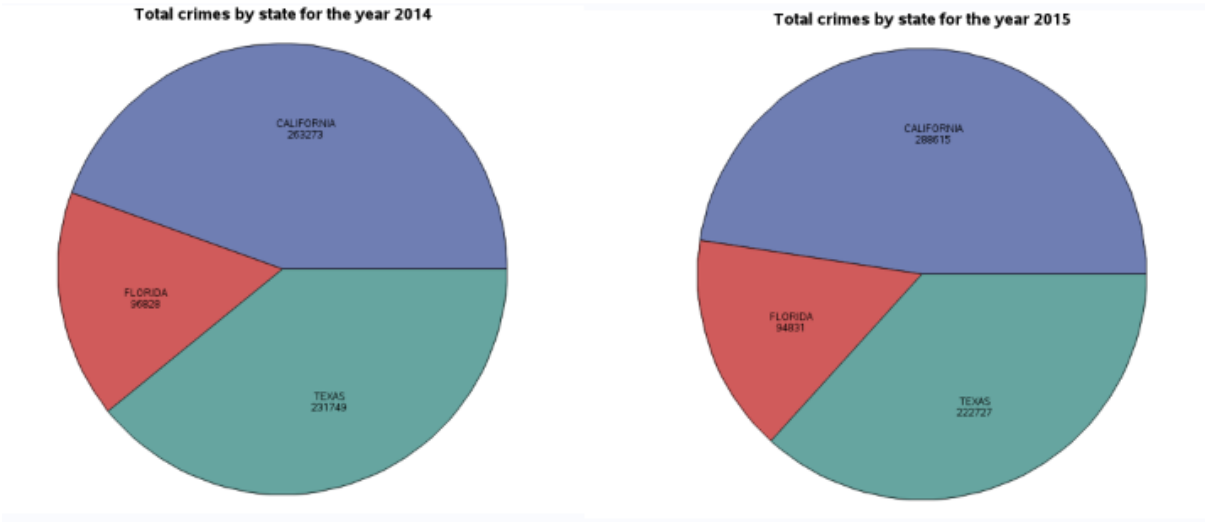
State	Year											
	2014						2015					
	Population	Violent Crimes	Property Crimes	Arson	Total Crimes	Crime Rate	Population	Violent Crimes	Property Crimes	Arson	Total Crimes	Crime Rate
CALIFORNIA	17,822,767	36,240	225,019	2,014	263,273	1477.17	18,086,261	40,919	245,852	1,844	288,615	1595.77
FLORIDA	4,422,308	13,642	82,954	232	96,828	2189.54	4,532,438	14,015	80,555	261	94,831	2092.27
TEXAS	10,605,950	29,657	201,057	1,064	231,749	2185.08	10,882,702	30,579	191,159	989	222,727	2046.61

SAS code for the above tabulate results can be referred from the file **Ob3Tabulate.sas** and the results can be referred from **Ob3Tabulate-results.pdf**

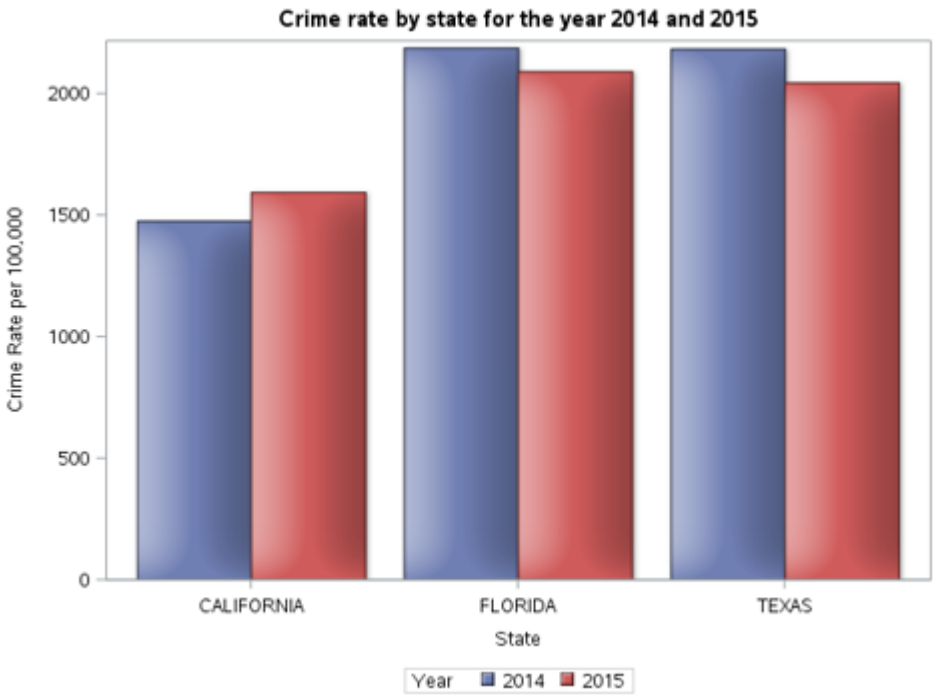
As per the above observations, total crime rate of the **California** State has increased in the year 2015 compared to the year 2014 and the States **Florida** and **Texas** has a reduction on the crime rates.

The same can be seen form the pi-chart below, it shows the sum of total number of crimes for the three states in the years 2014 and 2015.

Pi-Chart representing total crimes in the years 2014 and 2015



Below bar plot helps to get a better comparison of the crime rate in the years 2014 and 2015 for the selected states.



SAS code for the above charts can be referred from the file **Ob3TotalCharts.sas** and the results can be viewed from the document **Ob3TotalCharts-results.pdf**.

To have a better insight on the details of the crimes in the three states for the years 2014 and 2015, we shall divide the violent crime into murder, rape, robbery and aggravated assault. Similarly we can divide the property crimes into burglary, larceny theft and motor vehicle theft.

VIOLENT CRIME:

The detailed report on the violent crime for the three states by year is as below,

Detailed violent crime report : Year 2014

Obs	STATE	MURDER	ROBBERY	AGGRAVATED ASSAULT	RAPE
1	CALIFORNIA	411	13,203	20,506	2,120
2	FLORIDA	190	3,792	8,754	906
3	TEXAS	304	10,004	16,548	2,801

Detailed violent crime report : Year 2015

Obs	STATE	MURDER	ROBBERY	AGGRAVATED ASSAULT	RAPE
1	CALIFORNIA	399	14,799	22,834	2,887
2	FLORIDA	184	3,722	9,148	961
3	TEXAS	378	10,232	16,864	2,891

From the above report, we can infer the details,

1. The state of California has an increased number of crimes in robbery, aggravated assault and rape; however the number of murders decreased.
2. The state of Florida has only a rise in the aggravated assault and rape.
3. The state of Texas shows an increase in all the violent crimes however by only a very small percent.

We can break down the violent crimes into more details to find out the number of crimes that categorize under violent crimes using the transpose procedure. Sample Code is as following,

```
/*transpose funciton to get detailed violent crime report*/
PROC TRANSPOSE DATA=WORK.VIOLENT_CRIME_DATA_2014
    OUT=WORK.VC_2014_DETAILS
    NAME=VIOLENT_CRIMES;
    ID STATE;
RUN;

PROC TRANSPOSE DATA=WORK.VIOLENT_CRIME_DATA_2015
    OUT=WORK.VC_2015_DETAILS
    NAME=VIOLENT_CRIMES;
    ID STATE;
RUN;
```

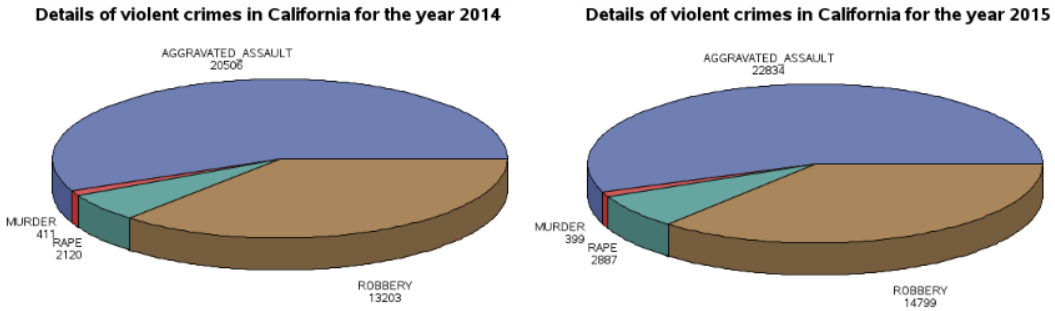
Details of the violent crimes in the states for the years 2014 and 2015 are as following,

Obs	VIOLENT CRIME	CALIFORNIA	FLORIDA	TEXAS
1	MURDER	411	190	304
2	ROBBERY	13,203	3,792	10,004
3	AGGRAVATED_ASSAULT	20,506	8,754	16,548
4	RAPE	2,120	906	2,801

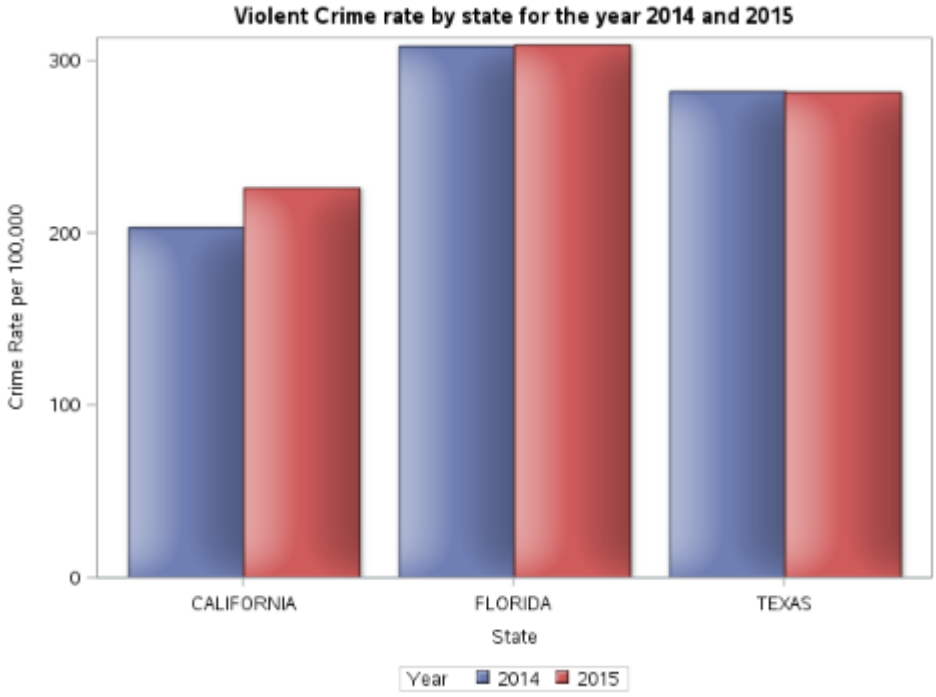
Details of violent crime by state. Year 2015

Obs	VIOLENT CRIME	CALIFORNIA	FLORIDA	TEXAS
1	MURDER	399	184	378
2	ROBBERY	14,799	3,722	10,232
3	AGGRAVATED_ASSAULT	22,834	9,148	16,864
4	RAPE	2,887	961	2,891

A more detailed representation of violent crimes in the state of California for the years 2014 and 2015 can be understood from the pie-chart below,



The below bar plot represents the variation of violent crime rate of the selected states in the years 2014 and 2015.



PROPERTY CRIME:

Detailed report on property crime for the selected states for the years 2014 and 2015 are as below,

Detailed property crime report : Year 2014

Obs	STATE	BURGLARY	LARCENY THEFT	MOTOR VEHICLE THEFT
1	CALIFORNIA	44,509	142,979	37,531
2	FLORIDA	16,503	60,452	5,999
3	TEXAS	38,678	142,506	19,873

Detailed property crime report : Year 2015

Obs	STATE	BURGLARY	LARCENY THEFT	MOTOR VEHICLE THEFT
1	CALIFORNIA	44,443	160,090	41,319
2	FLORIDA	14,502	59,390	6,663
3	TEXAS	34,717	135,399	19,154

We can infer the below points from the property crime report above,

1. Number of burglaries for the states of California has increased in the year 2015, whereas larceny and motor vehicle thefts have decreased.
2. The Florida state has a remarkable decrease in the number of burglaries. Number of larceny thefts also decreased by a very small number and there is an increase in the number of motor vehicle thefts.
3. The state of Texas has all the crimes reduced that categorizes under property crimes.

We can use transpose procedure to get more details of the crimes that categorize under the property crimes. Sample code as below,

```
/*transpose proc to get details of crimes
that categorize under property crimes*/
PROC TRANSPOSE DATA=WORK.PROPERTY_CRIME_DATA_2014
    OUT=WORK.PC_2014_DETAILS
    NAME=PROPERTY_CRIMES;
    ID STATE;
RUN;

PROC TRANSPOSE DATA=WORK.PROPERTY_CRIME_DATA_2015
    OUT=WORK.PC_2015_DETAILS
    NAME=PROPERTY_CRIMES;
    ID STATE;
RUN;
```

Details of the crimes that categorize under property crimes are as following,

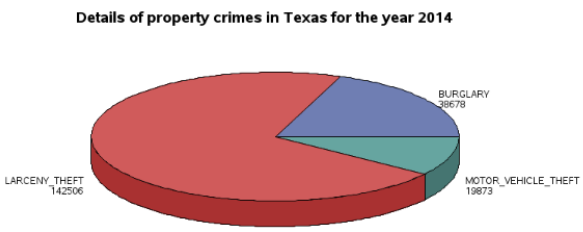
Details of property crimes by state. Year 2014

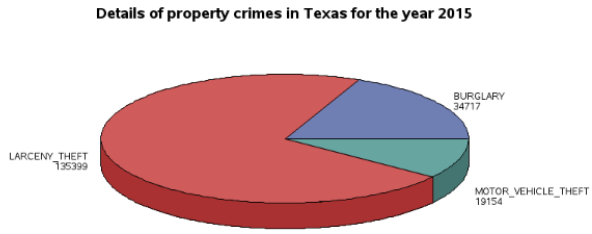
Obs	PROPERTY CRIME	CALIFORNIA	FLORIDA	TEXAS
1	BURGLARY	44,509	16,503	38,678
2	LARCENY_THEFT	142,979	60,452	142,506
3	MOTOR_VEHICLE_THEFT	37,531	5,999	19,873

Details of property crimes by state. Year 2015

Obs	PROPERTY CRIME	CALIFORNIA	FLORIDA	TEXAS
1	BURGLARY	44,443	14,502	34,717
2	LARCENY_THEFT	160,090	59,390	135,399
3	MOTOR_VEHICLE_THEFT	41,319	6,663	19,154

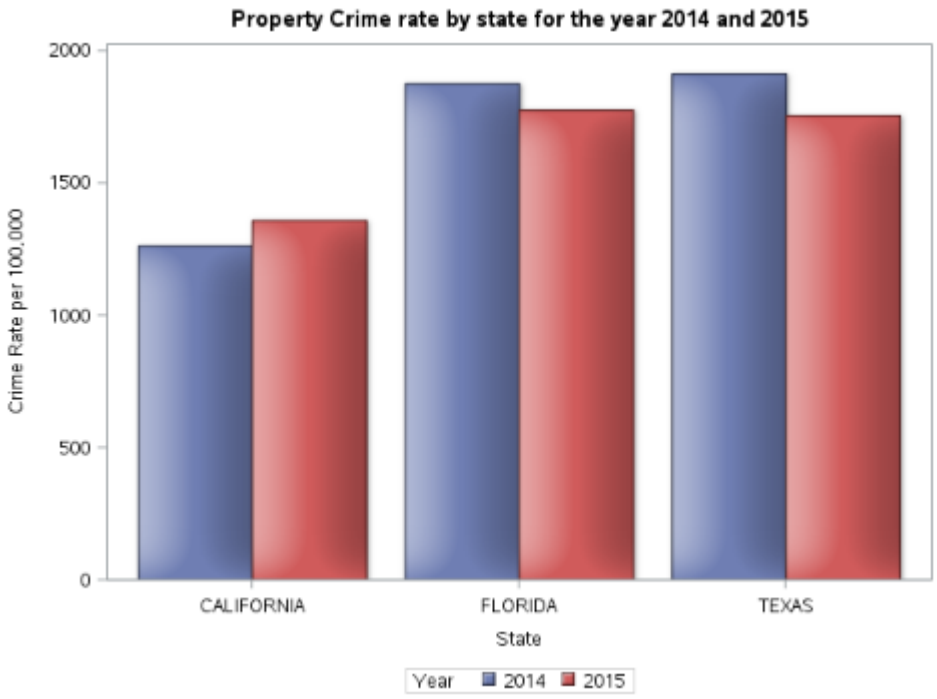
The below pie-chart represents the composition of property crimes in the state of Texas for the years 2014 and 2015,





More detailed break-down of the property crimes can be referred from the file **Ob3PCDetailsCharts-result.sas.**

Below bar chart represent the variation of the property crime rate in the years 2014 and 2015,



So by understanding the readings from the bar plots, we can conclude that the crime rate increase in the state **California** is contributed by the increased rates of both violent crime and the property crime. For the states of **Florida** and **Texas**, even though there is not much difference in the violent crime rates, the property crime rates have been reduced by a sound number.

The state of **California** stands number one in terms of settlement of immigrant population in the United States of America. About **27%** of the total population in **California** is immigrant population in the year 2015. The percentage of immigrant population in the state has been slowing down since the 1990s and has been constant around 27% from 2010, Reference:
"http://www.ppic.org/main/publication_show.asp?i=258"

However, the crime rate has been in a constant increase in the state over the years.

This rise in the crime rate for the state of **California** in the year 2015 is mainly because of the rise in the violent crimes in the state as quoted by the state's attorney general, the information says there is a rise in homicide cases by 9.7% and whereas robbery and aggravated assault rates have climbed more than 8%. Most of the cases

involved religious bias, anti-Muslim and anti-Jewish incidents and also targeting the Latinos and thus this is a clear statement that the crime rates have be not been affected much by the fall in the un employment rate from 8.1% in 2014 to 6.8 % in 2015.

The total property crimes climbed up by 8% driven by the significant increase in the motor vehicle theft. Reference:" http://www.latimes.com/local/lanow/la-me-violent-crimes-20160701-snap-story.html".

OBJECTIVE 4:

To find the most dangerous and safest cities in terms of violent crime for the selected states in the years 2014 and 2015.

Methodology:

To find the most dangerous and the safest cities, we need to calculate the crime rate in terms of violent crime for each city in the selected states.

- The city with the lowest crime rate will be the safest city.
- The city with the highest crime rate will be the most dangerous city.

We shall carry out our analysis grouped by the year and compare the results for the years 2014 and 2015. We can then analyze each crime in the resulted cities for the years 2014 and 2015.

SAS code to calculate the crime rates for violent by city is as follows,

```
/*data set to select the records from
the states CA, FL and TX and add new variable
violent crime rate*/

DATA WORK.VC_RATE_CITY_2014;
    SET WORK.CRIME_DATA;
    WHERE YEAR=2014 AND
    STATE_CODE IN('CA','FL','TX') AND
    CITY NE 'TYLER';/*Tyler no considered for a
    nalysis die to unavailability of data*/
    VIOLENT_CRIME_RATE=(VIOLENT_CRIME/POPULATION)*100000;
    KEEP STATE CITY POPULATION
        VIOLENT_CRIME MURDER RAPE
        ROBBERY AGGRAVATED_ASSULT VIOLENT_CRIME_RATE;
RUN;
```

We have sorted the data set by the descending order of crime rates calculated for all the cities in the selected states for the year 2014 and the result looks as following,

US: Californai, Florida, Texas
Violent Crime Details
Year 2014

Obs	State	City	Population	Violent Crime	Murder	Rape	Robbery	Aggravated Assault	Violent Crime Rate
1	CALIFORNIA	STOCKTON	299,519	1,972	21	61	532	1,358	658.389
2	CALIFORNIA	SAN BERNARDINO	214,588	1,058	26	47	400	585	493.038
3	FLORIDA	MIAMI	421,996	2,080	43	35	768	1,234	492.896
4	TEXAS	HOUSTON	2,219,933	10,401	100	416	4,717	5,168	468.528
5	FLORIDA	TALLAHASSEE	187,573	840	5	89	150	596	447.826

.....

112	CALIFORNIA	TEMECULA	108,308	44	1	4	26	13	40.625
113	TEXAS	FRISCO	141,940	57	1	10	6	40	40.158
114	CALIFORNIA	SIMI VALLEY	126,604	50	0	9	10	31	39.493
115	CALIFORNIA	IRVINE	242,971	54	0	9	17	28	22.225
116	CALIFORNIA	MURRIETA	108,376	21	0	2	7	12	19.377

*Crime rate per 100,000

The above tables are fragments of the resultant data set showing the cities with the least crime rate and the highest crime rate in the year 2014.

➔ So for the year **2014**, the most dangerous city in terms of violent crime is **Stockton** from the state of California with a crime rate of **658.389 per 100,000**.

➔ Safest city in terms of violent crime is **Murrieta** from the state of California with a crime rate of **19.377 per 100,000**.

Similarly the crime rates are calculated for each city in the year 2015 and the sorted in the descending order of the violent crime rate. The resulting data set is as follows,

US: Californai, Florida, Texas
Violent Crime Details
Year 2015

Obs	State	City	Population	Violent Crime	Murder	Rape	Robbery	Aggravated Assault	Violent Crime Rate
1	CALIFORNIA	STOCKTON	305,658	2,019	18	65	570	1,366	660.542
2	CALIFORNIA	SAN BERNARDINO	216,108	1,223	16	50	419	738	565.921
3	FLORIDA	TALLAHASSEE	189,907	990	8	119	166	697	521.308
4	FLORIDA	MIAMI	441,003	2,226	42	33	829	1,322	504.758
5	TEXAS	ODESSA	118,968	589	6	32	79	472	495.091
.....									
113	CALIFORNIA	TEMECULA	112,011	51	0	5	20	26	45.531
114	CALIFORNIA	MURRIETA	109,830	42	0	7	17	18	38.241
115	TEXAS	FRISCO	154,407	49	1	12	14	22	31.734
116	CALIFORNIA	IRVINE	256,927	56	1	11	20	24	21.796

*Crime rate per 100,000

From the above fragments of data set we can derive the following,

➔ In the year **2015**, the most dangerous city is **Stockton** from California State with a violent crime rate of **660.542 per 100,000**.

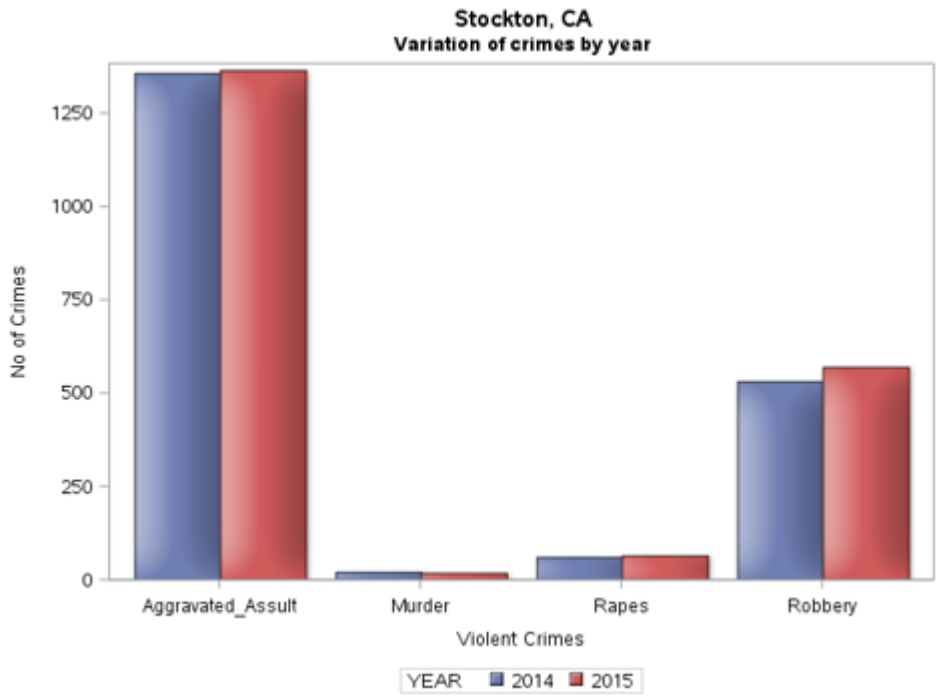
➔ Safest city for the year **2015** is **Irvine** form the California State with a violent crime rate of **21.796 per 100,000**.

The above mentioned results and the related SAS codes can be referred from Ob4Data-result.pdf and Ob4Data.sas respectively.

Comparison of violent crimes in the resultant cities:

Stockton, CA:

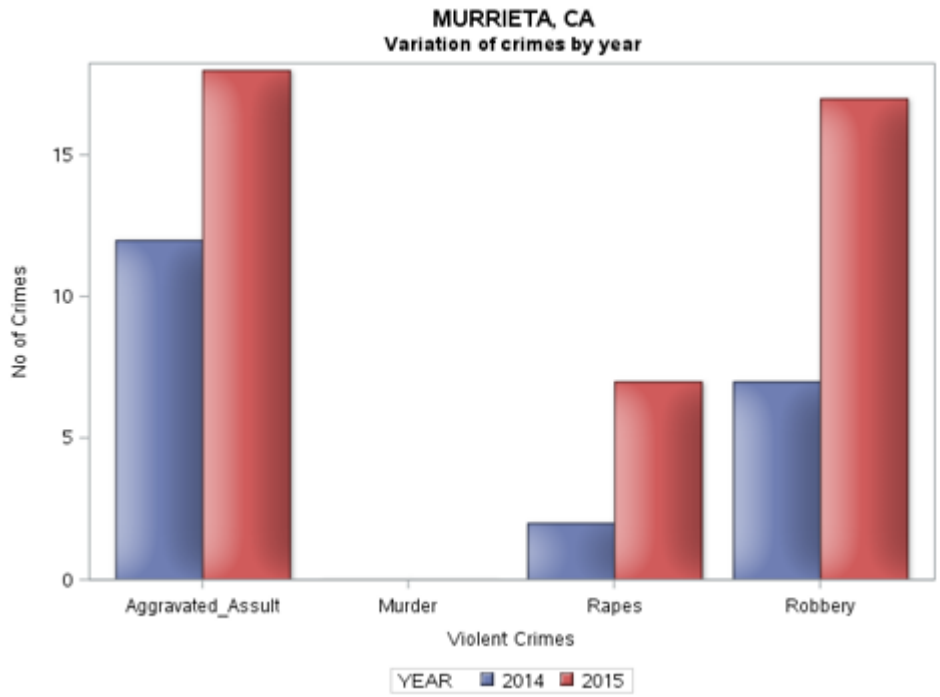
Refer the below bar chart to understand the variation in number of crimes in the city of **Stockton** for the years 2014 and 2015,



We can observe only the number of murders have reduced in the year 2015 compared to the year 2014 and the crime that has increased more is robbery.

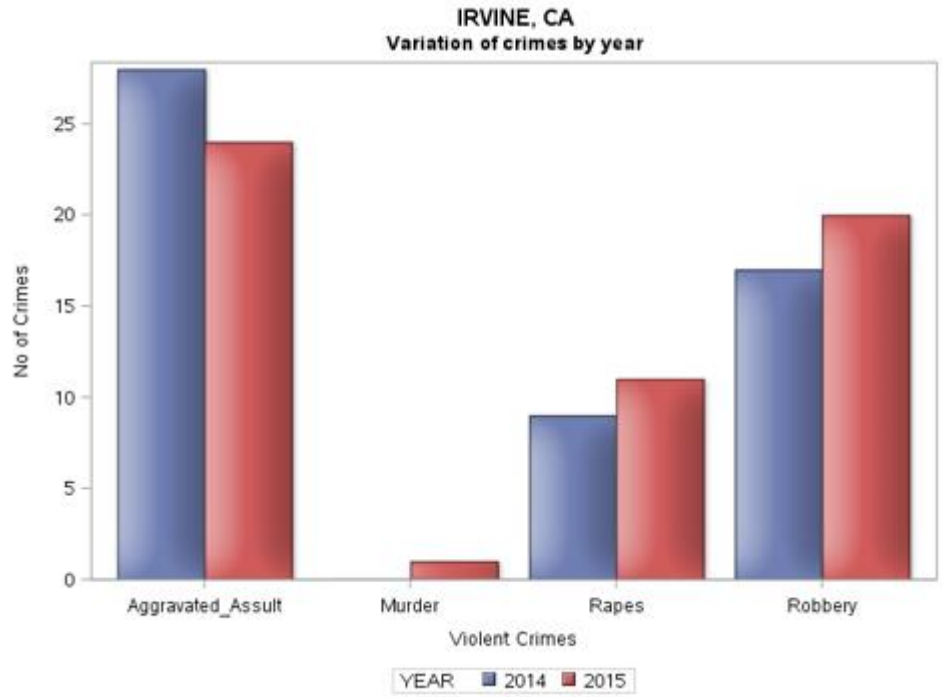
Murrieta, CA:

The below bar chart shows the number of violent crimes for the city Murrieta in the years 2014 and 2015,



Number of reported murders remain zero for both the years, however rest of the violent crimes have slightly increased thus losing its name as the safest city in the year 2015.

Irvine, CA:



Irvine has the lowest crime rate of all the cities in the year 2015. The above chart represents the number of violent crimes occurred in Irvine in the years 2014 and 2015. However crime rate of Irvine has only reduced by 1% by the year 2015.

OBJECTIVE 5:

To find the safest and the most dangerous cities in terms of property crime among the chosen states for the years 2014 and 2015.

Methodology

To find the most dangerous and the safest cities in terms property crime for the selected states,

- The city with the lowest crime rate will be the safest city.
- City with the highest crime rate in the year 2014 and 2015 will be the most dangerous city in terms of property crime.

To find this, we shall calculate the crime rate for property crimes for each city and add it as a variable in the resulting data set. Then we arrange the data set in the descending order of the crime rate. First observation in the resulting set will be our most dangerous city in terms of property crime and the last observation will be our safest city in terms of property crime.

We shall perform the analysis separately for the years 2014 and 2015 and then compare the variation of each crime for the safest and the dangerous cities.

Below SAS code will result in the data set with the property crime rate calculated,

```
/*data set to select the records from
the states CA, FL and TX and add new variable
property crime rate for the year 2014*/
DATA WORK.PC_RATE_CITY_2014;
        SET WORK.CRIME_DATA;
        WHERE YEAR=2014 AND
        STATE_CODE IN('CA','FL','TX') AND
        CITY NE 'TYLER';/*Tyler not considered for
        analysis die to unavailability of data*/
        PROPERTY_CRIME_RATE=(PROPERTY_CRIME/POPULATION)*100000;
RUN;
```

The below data set fragments gives us the most dangerous and the safest cities for the year 2014 in the three states,

US: Californai, Florida, Texas
Property Crime Details
Year 2014

Obs	State	City	Population	Property Crime	Larceny Theft	Burglary	Motor Vehicle Theft	Property Crime Rate
1	FLORIDA	ORLANDO	259,675	8,311	6,044	1,801	466	3,200.539
2	FLORIDA	ST. PETERSBURG	250,772	6,830	4,832	1,358	640	2,723.590
3	TEXAS	SAN ANTONIO	1,428,465	38,179	28,292	6,230	3,657	2,672.729
4	FLORIDA	FORT LAUDERDALE	174,056	4,612	3,311	972	329	2,649.722
5	FLORIDA	LAKELAND	101,428	2,598	1,993	474	131	2,561.423
.....								
113	CALIFORNIA	MURRIETA	108,376	735	488	140	107	678.194
114	CALIFORNIA	IRVINE	242,971	1,495	1,167	266	62	615.300
115	CALIFORNIA	THOUSAND OAKS	129,175	794	582	162	50	614.670
116	CALIFORNIA	SIMI VALLEY	126,604	778	598	134	48	614.515

*Crime rate per 100,000

Observations for the year 2014

- ➔ **Orlando** from Florida State is the most dangerous city in terms of property crime with a crime rate of **3,200.539 per 100,000**.
- ➔ **Simi Valley** from California State is the safest city in terms of property crime with a crime rate of **614.515 per 100,000.**.

Similarly property crime report for the year 2015 is as below,

US: Californai, Florida, Texas
Property Crime Details
Year 2015

Obs	State	City	Population	Property Crime	Larceny Theft	Burglary	motor Vehicle Theft	Property Crime Rate
1	CALIFORNIA	SAN FRANCISCO	864,816	27,001	20,854	2,653	3,494	3,122.167
2	FLORIDA	ORLANDO	270,934	7,874	5,753	1,588	533	2,906.243
3	FLORIDA	ST. PETERSBURG	257,083	7,156	5,263	1,127	766	2,783.537
4	FLORIDA	FORT LAUDERDALE	178,590	4,532	3,266	877	389	2,537.656
.....								
113	TEXAS	FRISCO	154,407	1,088	885	178	25	704.631
114	FLORIDA	PORT ST. LUCIE	179,413	1,164	924	204	36	648.782
115	CALIFORNIA	THOUSAND OAKS	129,339	830	638	131	61	641.724
116	CALIFORNIA	SIMI VALLEY	126,788	801	613	132	56	631.763

*Crime rate per 100,000

Observations for the year 2015,

- ➔ **San Francisco** from California State records the highest property crime rate making it as the most dangerous city with a crime rate of **3,122,167 per 100,000**.
- ➔ **Simi Valley** from California State remained as the safest city for the year 2015 with a crime rate of **631.763 per 100,000.**

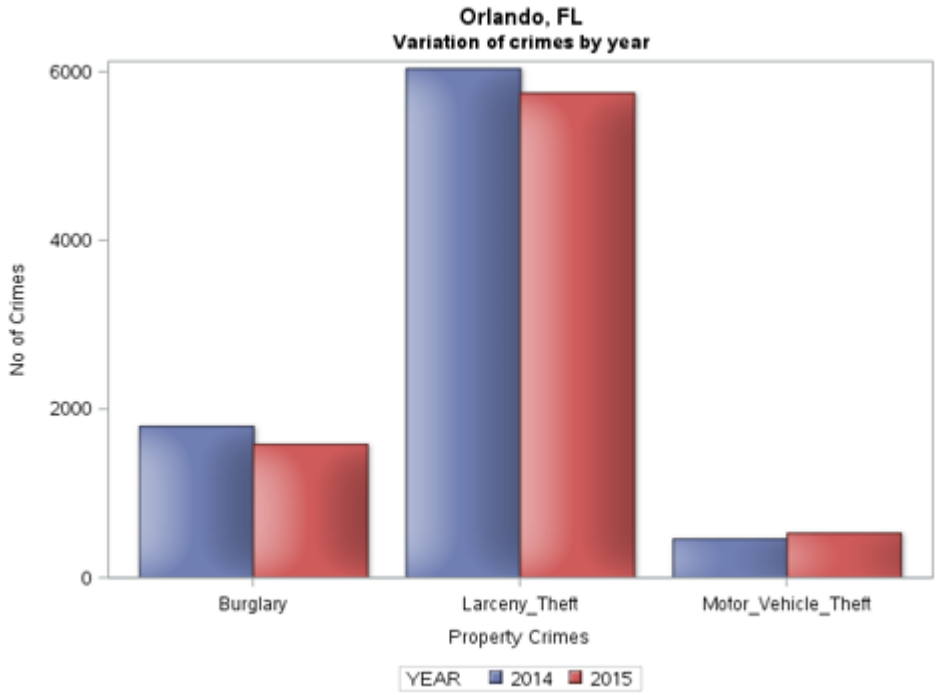
The above displayed results and the related SAS code can be referred from O5Data-result.pdf and Ob5Data.sas respectively.

Comparison of property crimes for the resultant cities:

Orlando, FL:

Orlando records a crime rate of 3,200.539 per 100,000 in the year 2014. Its crime rate has reduced to 2906.234 by the year 2015 which is remarkable as the crime rate has nearly reduced by 7.5%.

The below bar chart represents the variation of different property crimes for Orlando city in the year 2014 and 2015.

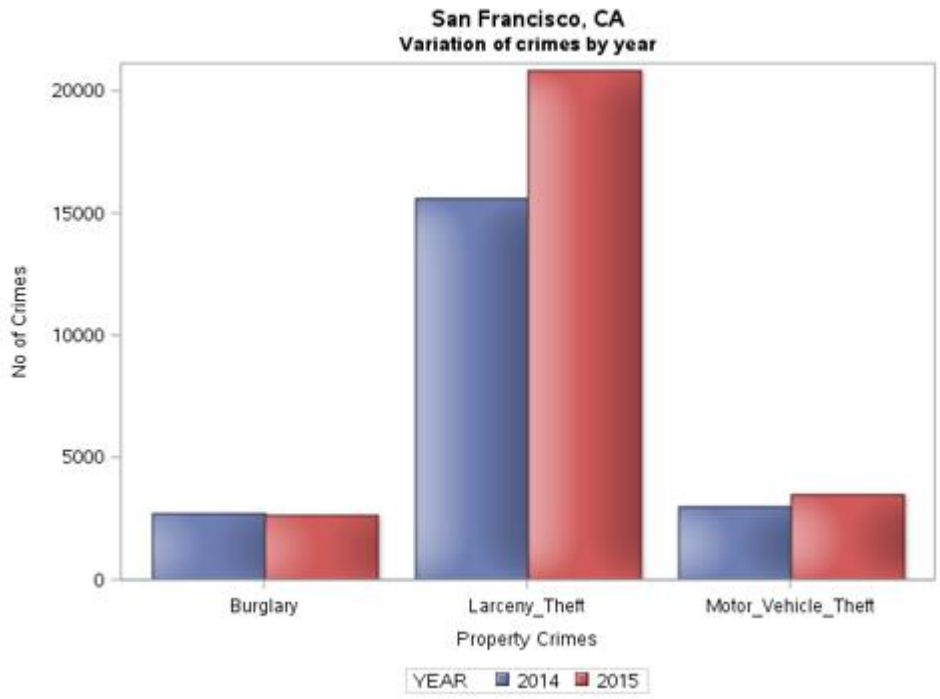


We can observe the number of larceny thefts and burglaries reduced by the year 2015 but the number of motor vehicle thefts has increased. Due to this decrease in the number of crimes, Orlando is not the most dangerous city in terms of property crimes in the year 2015. Orlando has dropped to the second most dangerous city in terms of property crime by the year 2015.

San Francisco, CA:

In the year 2014, San Francisco ranked sixth in the highest property crime rate with about 2,508.544 per 100,000. By the year 2015 its crime rate has increased to 3,122.167 per 100,000. Crime rate in San Francisco has increased about 19.65% which is alarming.

Below bar chart represents the variation of property crimes in San Francisco in the year 2014 and 2015,

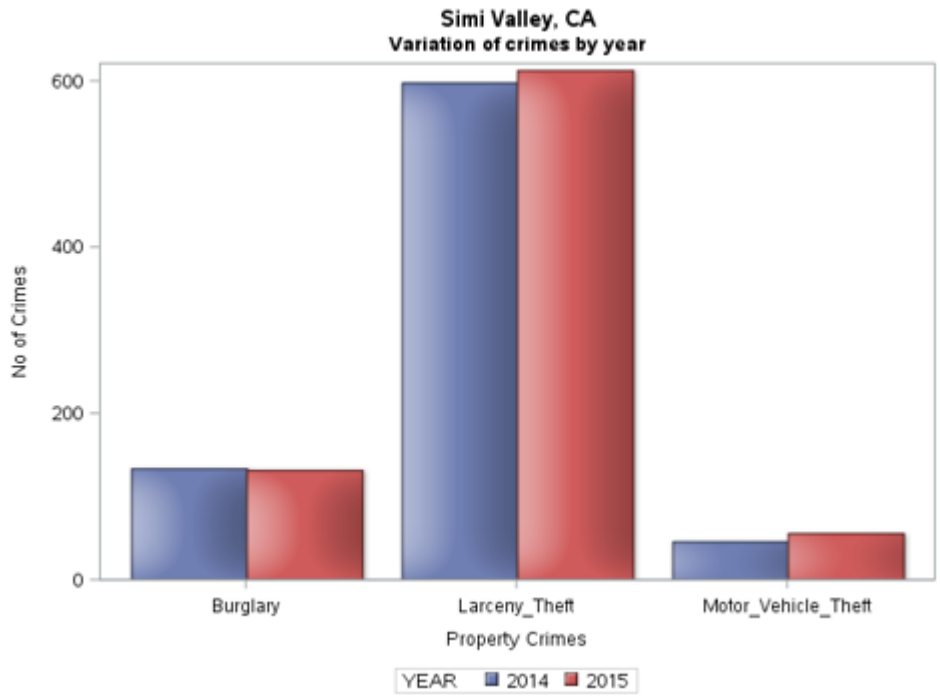


From the bar chart we can clearly observe that the increase in the crime rate is majorly due to number of larceny thefts that occurred in San Francisco in the year 2015.

Simi Valley, CA:

Simi Valley has remained mostly consistent in suppressing the crimes and maintained its crime rate at the lowest for both the years. It has an increase in the crime rate by 2.8 %. It recorded a crime rate of 614.515 per 100,000 in the year 2014 and in the year 2015 its crime rate has increased to 631.763 per 100,000.

Below bar chart shows the variation of property crimes for Simi Valley in the year 2014 and 2015.



We can clearly observe that the variation of the property crimes is very minimal for Simi Valley in the year 2015 compared to the year 2014.

OBJECTIVE 6:

To categorize all the states in the Unites States of America by crime rate for the years 2014 and 2015.

We can visualize the crime rate across all the states in the US using PROC GMAP by categorizing the crime rates into certain ranges and color coding the ranges on the US map. Analysis for the years 2014 and 2015 will be done separately.

Total crimes for the state of Hawaii is not available for the year 2014, hence we substitute the total crimes with same as in the 2015 for our analysis.

Data set for the year 2014 for visualization is as following,

United States: Year 2014
Crime Rate by State
****Crime rate per 100,000**

Obs	State	State Code	Population	Total Crimes	Crime Rate
1	ALABAMA	AL	850,588	24,501	2,880.478
2	ALASKA	AK	301,306	6,724	2,231.618
3	ARIZONA	AZ	3,470,667	53,865	1,552.007
4	CALIFORNIA	CA	17,822,767	261,259	1,465.872
5	COLORADO	CO	1,552,565	28,908	1,861.951
6	CONNECTICUT	CT	509,645	9,023	1,770.448
7	FLORIDA	FL	4,422,308	96,596	2,184.289
8	GEORGIA	GA	1,120,454	28,640	2,556.107
9	HAWAII	HI	994,034	14,994	1,508.399
10	IDAHO	ID	216,260	2,410	1,114.399

Similarly the data set for the year 2015 is as following,

United States: Year 2015
Crime Rate by State
****Crime rate per 100,000**

Obs	State	State Code	Population	Total Crimes	Crime Rate
1	ALABAMA	AL	853,752	23,359	2,736.040
2	ALASKA	AK	298,695	7,347	2,459.700
3	ARIZONA	AZ	3,555,486	71,234	2,003.495
4	CALIFORNIA	CA	18,086,261	286,771	1,585.574
5	COLORADO	CO	1,593,089	26,952	1,691.808
6	CONNECTICUT	CT	509,311	8,440	1,657.141
7	FLORIDA	FL	4,532,438	94,570	2,086.515
8	GEORGIA	GA	1,179,347	29,055	2,463.651
9	HAWAII	HI	998,714	14,994	1,501.331
10	IDAHO	ID	218,281	2,415	1,106.372

SAS code for the above data sets can be found from the file **Ob6DataSets.sas** and the result can be referred from **Ob6DataSets-result.pdf.**

State codes are used in the data sets to map our data set to the **MAPS.US** to get the state number for all the states in our data set. I.e. State code will be used as the field for inner join condition. Sample SAS code is as below,

```
/*Data set to have state code and state number*/
PROC SQL;
        CREATE TABLE WORK.STATE_CODES_US AS
        SELECT STATE,STATECODE
        FROM MAPS.US
        GROUP BY STATE,STATECODE;
QUIT;

/*proc sort to sort the data set by state number
and remove duplicates */
PROC SORT DATA=WORK.STATE_CODES_US NODUPKEY;
BY STATE;
RUN;

PROC SQL;
        CREATE TABLE WORK.MAP_2014_NO AS
        SELECT STATE_CODE,CRIME_RATE,B.STATE
        FROM WORK.MAP_2014 AS A,WORK.STATE_CODES_US AS B
        WHERE A.STATE_CODE=B.STATECODE;
QUIT;
```

Above sample code can be referred from **Ob6Gmap2014.sas.**

To categorize the states based on crime rate, we shall consider dividing the entire range of the crime rate into 5 categories as below,

Crime Rate ranging from 1000 - 1500 as Low,

Crime Rate ranging from 1501 - 2000 as Medium,

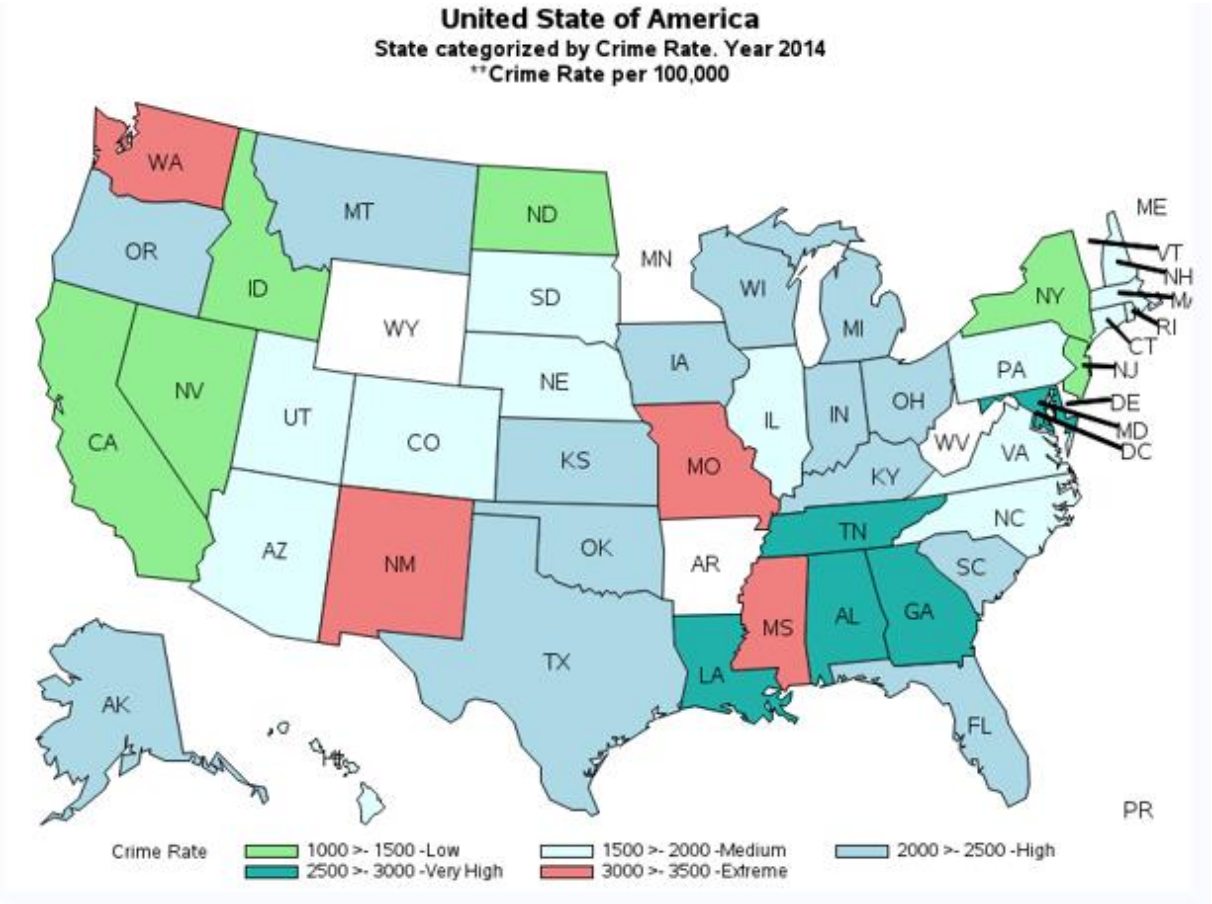
Crime Rate ranging from 2001 - 2500 as High,

Crime Rate ranging from 2501 - 3000 as Very High,

Crime Rate ranging from 3001 - 3500 as Extreme.

All the 43 states in our data set would fall into each of the above mentioned categories. Rest of the states in which the document did not cover any analysis would be marked white in the map.

The below map displays the US states categorized by crime rate for the year 2014.



SAS Code sample for the above map is as below,

```
/*Define the colorramp to represent
each category with a color*/
PROC TEMPLATE;
        DEFINE STYLE STYLES.COLORRAMP;
        PATTERN1 c=LIGHTGREEN;
        PATTERN2 c=LIGHTCYAN;
        PATTERN3 c=LIGHTBLUE;
        PATTERN4 C=LIGHTSEAGREEN;
        PATTERN5 c=LIGHTCORAL;
END;
RUN;
```

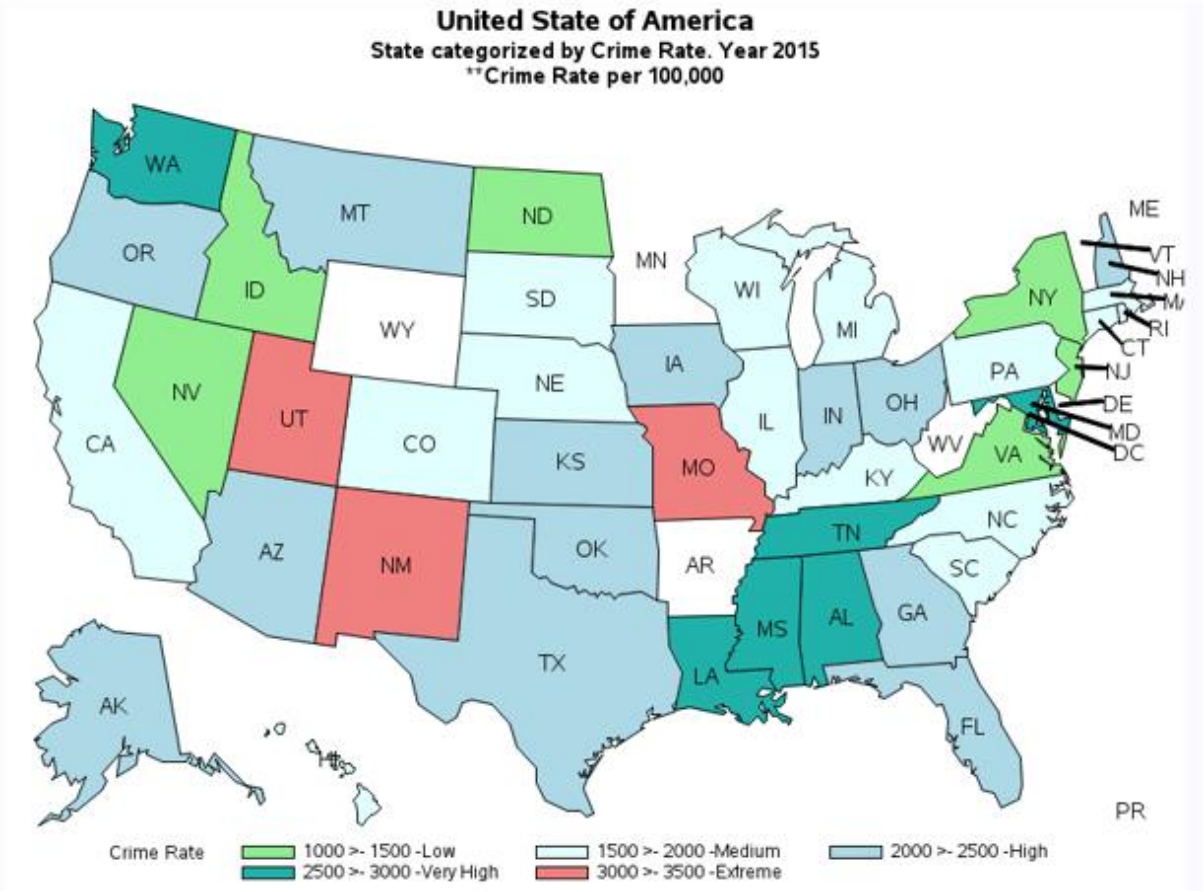
```
/*Define the categories*/
PROC FORMAT;
 VALUE RATE 1000 - 1500 = '1000 >- 1500 -Low'
            1501 - 2000 = '1500 >- 2000 -Medium'
            2001 - 2500 = '2000 >- 2500 -High'
            2501 - 3000 = '2500 >- 3000 -Very High'
            3001 - 3500 = '3000 >- 3500 -Extreme';
 RUN;
```

```
ODS LISTING STYLE=STYLES.COLORRAMP;
ODS HTML STYLE=STYLES.COLORRAMP;

/*proc gmap to generate the map, apply the category format on the
crime_rate field of our data set*/
PROC GMAP DATA=WORK.MAP_2014_NO MAP=MAPS.US;
       FORMAT CRIME_RATE RATE.;
       LABEL CRIME_RATE='Crime Rate';
       ID STATE;
       CHORO CRIME_RATE / DISCRETE COUTLINE=BLACK annotate=maplabel ;
TITLE1 'United State of America';
TITLE2 'State categorized by Crime Rate. Year 2014';
TITLE3 '**Crime Rate per 100,000';
RUN;
```

Complete SAS code can be referred from the file Ob6Gmap2014.sas and the result can be viewed from the file Ob6Gmap2014-result.pdf.

Similarly, we also categorized the states based on the defined range of crime rates and the map for the year 2015 is as below,



SAS Code for generating this map can be referred from the file **Ob6Gmap2015.sas** and the result can be viewed from the file **Ob6Gmap2015-result.pdf**.

Comparing the maps for the years 2014 and 2015, it's simple to decide the variation in the crime rate for the states. For example, the state of Washington has a reduced crime rate in the year 2015 and the state of California has an increase.

RECOMMENDATIONS

Referring the analysis, I would like to recommend the following,

- **Improvements in Infrastructure:** Alaska State has recorded the highest increase in the violent crime rate, the government and the law enforcement agencies would need to improve the administration and security infrastructure to provide the security for the citizens. Youngsters and students would bear the future of any country, hence it is really necessary that they be educated with social ethics and responsibilities towards the society. Coming to which government should mainly focus on increasing the literacy rate to enable competency among people and reduce the un-employment rate.
- **Increasing awareness in society:** According to our analysis and references, we observed that the rise in crime rate in the state of California is due to racism. In short this is due to cross cultural hatred. California records the highest settlement of foreign population among all the states in US. This kind of social issues are sensitive and need to be handled carefully as any hasty decisions could actually contribute to increase in the problem. One way to solve this is to conduct town hall meetings with people representing their cultures and to be counselled to understand the problems between the cultures and gain acceptance among them.
- **Ethical Responsibility towards society**: Parents, Schools and Universities would need to play a major role to build a safer nation. Educational institutions should implement social ethics, social responsibilities and equality of human rights as a fundamental practice which in time would turn

into a habit. Doing so is a responsibility of this generation as this would give a better society to the coming generations.

CONCLUSION

The prime purpose of this assignment is to analyze various crimes that occurred in the United States of America for the first six months of the years 2014 and 2015. This document discussed the variation of crimes and crime rates among different states for the years, crimes including violent crimes and property crimes.

We have found that the aggravated assault is the main driver for the increase of violent crime rate in the year 2015 compared to the year 2014. In the category of property crimes, the main driver is larceny thefts that have contributed to any increase in the property crimes. However the overall property crime rate has reduced by a remarkable percentage for all the states taken as a whole.

Code reference for all the objectives can be viewed either from the CD-ROM attached to the hard copy of this document or can be viewed from the GitHub repository in the link: https://github.com/RakeshJalla27/DS-DAP-ASSIGNMENT. Soft copy of this document can also be found in the same GitHub location.

REFERENCES

1. Population Estimates for the year 2015: https://www.census.gov/data.html
2. State Codes for the United States of America: https://www.infoplease.com/state-abbreviations-and-state-postal-codes .
3. Immigrant population details in California http://www.ppic.org/main/publication_show.asp?i=258.

4. State attorney General's report on the increase of crime rate in the state of California for the year 2015, http://www.latimes.com/local/lanow/la-me-violent-crimes-20160701-snap-story.html.
5. SAS code reference: http://support.sas.com/documentation/94/.