DATA ANALYTICAL PROGRAMMING (DAP)

CT050-3-M

Individual Assignment

SUBMITTED TO

Dr. Kalai Anand Ratnam

PREPARED BY:

NAME   : FOO YEONG JIN

STUDENT ID   : TP044538

INTAKE   : UCMP1701DSBA

DATE   : 15TH MAY 2017

Acknowledgement

I would like to express my deep gratitude to Dr. Kalai Anand Ratnam, my lecturer of this module (Data Analytical Programming), for his patient guidance and assistance from the very beginning of this assignment, without which the completion would not be possible. Apart from the guidance, Dr. Kalai also broadens my knowledge in the Data Science by giving a lot of example on how the Data Science works in the industry and prepare me for the bigger challenge.

I would also like to extend my thanks to fellow classmates, who contributed to this assignment directly or indirectly.

Finally, I wish to thank my wife for her support and encouragement throughout my study.

Introduction

The Uniform Crime Reporting (UCR) program was form in 1929 by the Internal Association of Chief of Police in order to collect reliable data for the crime that reported to the law enforcement in the United States. Since 1930, The Federal Bureau of Investigation (FBI) is tasked to gather the crime related statistic for the Uniform Crime Reporting (UCR) from the law enforcement agencies in United States.

Today, the data that gathered through this program has enabled the Federal Bureau of Investigation (FBI) to produce four annual publications which includes Crime in the United States, National Incident-Based Reporting System, Law Enforcement Officers Killed and Assaulted, and Hate Crime Statistics. These publications were produced from data received from over 18,000 city, university and college, county, state, tribal, and federal law enforcement agencies voluntarily participating in the program. In addition to the publications, the data collected through the Uniform Crime Reporting (UCR) also enabled the law enforcement analyst, students, researchers, media, and the public to analyze the crimes in the United States.

In this assignment, I am responsible to analyze the crime data of the United States together with the members of the data science team in Federal Bureau of Investigation (FBI) by using the data collected from the Uniform Crime Reporting (UCR) mentioned above.

For this assignment, our team would like to achieve the following objectives:

The objective:

1. Analyze and understand the overall crime rate across United States.
2. Top 3 types of crimes in United States.
3. Identify the top 3 states with highest crime rate in year 2014 and their crime rate in year 2015
4. Identify the top 3 states with lowest crime rate year 2014 and their crime rate in year 2015
5. Compare the Police Employee Ratio in the Top 3 States with lowest crime rate (Objective 4) and the top 3 states with highest crime rate (Objective 3) in Year 2014.
6. Compare the crime rate for the states close to Mexico border and the states close to Canada border.

Defining Target

In this assignment, crime rate will be used as the target instead of the absolute number of offences/crimes. This is because the number of offences is direct correlate to the population size (linear relationship), which means the bigger the population the higher the number of offences. Comparing the number of offences between 2 selected samples can be misleading.

State	Total Population	# Of Offences
CALIFORNIA	17,822,767	299,513
MISSOURI	1,186,439	43,189

*Data retrieve from table 4

Based on the example above, if we measure the target by number of offences, it seems like California is more than 5 times more dangerous than Missouri. However, this is not true because if we measure by the crime rate, it shows that the probability of experiencing an offence/crime is much higher in Missouri.

State	Total Population	# Of Offences	Crime Rate
CALIFORNIA	17,822,767	299,513	1.68%
MISSOURI	1,186,439	43,189	3.64%

*Data retrieve from table 4

The basis of crime rate is to measure the probability of experiencing an offence/crime in every 100 citizens and it calculated based on the formula below:

Crime Rate = # of Offences / Total Population

By using this approach, it can eliminate the population effect on the number of offences/crimes and reveal the actual security status of a state or city.

Data Acquisition

There are 4 tables provided for this assignment, detail as follow:

Table 1

This table provides the percent change in offenses known to law enforcement, by population group, for the first 6 months of 2014 and 2015.

Table 2

This table provides the percent change in offenses known to law enforcement, by region, for the first 6 months of 2014 and 2015.

Table 3

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

Table 4

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 data that used for this analysis is the table 4, which contain the detail of offences including the cities, states and population. Additional data also being acquired from the Uniform Crime Reporting (UCR) and United States Census Bureau in order to achieve the objectives. Below are the details of the additional data:

1. Law enforcement employee (Table 78)

   This data is retrieved from the Uniform Crime Reporting (UCR), it contains number of officers which is the police employee who has the arresting power and also the civilian employee by states and cities. In addition to the number of officers, this data also contain the population of each city.

2. State code, division and region data (Region)

   This data is acquired from United States Census Bureau, it enables the team to map the cities and states to respected division and regions.

Data Cleansing and Preparation

Data cleansing and preparation is a very important step in any kind of analysis. In fact, this step consume most of the time (~50% - 80%) in the whole project. Through this step, the analyst will have a detail understanding of the input data in term of the *Accuracy, Integrity, Cleanliness, Correctness, Completeness* and *Consistency* before we start the analysis in SAS.

In this project, we split the data cleansing and preparation in to two parts. The first part of data cleansing and preparation is done in Excel. The team prepares and cleans the table 4 and additional data acquired, then convert the tables into a structured table format to import to SAS. The details as follow:

The first step is to remove the unnecessary items and rename the columns in the table4, which includes:

Figure 1: Table 4 header description

Figure 2: Table 4 footnote

- Remove the description of the table (highlighted in **blue** in figure 1) and footer (highlighted in **orange** in figure 2). The description is not in the structure format and will not be used in the analysis.
- Remove the special characters and remarks in the columns header, i.e. bracket and numbering (highlighted in **green** in figure 1).
- Unmerge the merged cells (highlighted in **red**) as SAS is not able to convert the merge cell properly during the import process. Without removing the merge cell, the output table will be inaccurate.
- Rename the columns name by replacing the space to underscore ('_').

After remove the unnecessary items and rename the columns, the team then follow the steps in excel (Refer to: Appendix 1) to tidy up the table4. After all the steps, now the gaps are filled in with the correct information as below (Figure 3). We follow the same steps for the additional data and turn those data into a structured table.
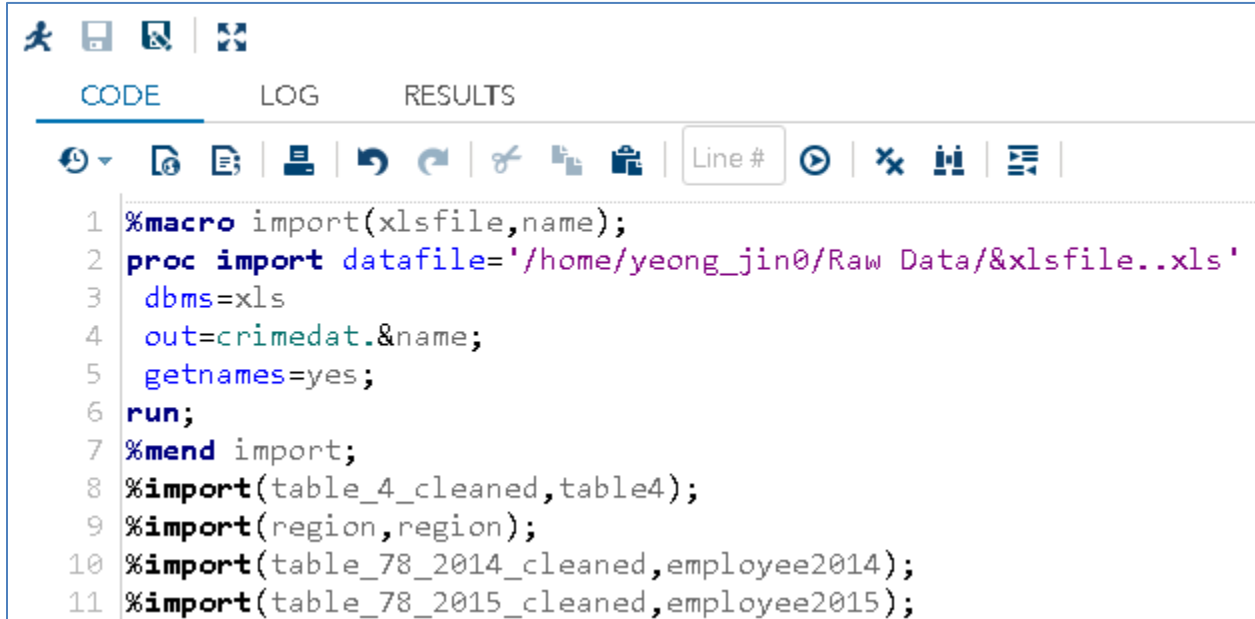
	State	City	Year	Population	Violent_Crime	Murder	Rape_Revised	Rape_Legacy	Rape_Total	Robbery	Aggravated_Assault	Property_Crime	Burglary	Larceny_Theft	Motor_Vehicle_Theft	Arson
2	ALABAMA	BIRMINGHAM	2014	212,115	1,619	23	83		83	454	1,059	6,536	1,716	4,169	711	76
3	ALABAMA	BIRMINGHAM	2015		1,756	30	77		77	507	1,142	6,246	1,446	4,120	680	
4	ALABAMA	HUNTSVILLE	2014	187,624	770	12	50		50	188	520	4,376	908	3,111	357	12
5	ALABAMA	HUNTSVILLE	2015		723	5	65		65	173	480	4,121	836	2,903	382	
6	ALABAMA	MOBILE	2014	250,655	747	17	67		67	203	460	5,747	1,461	4,039	247	54
7	ALABAMA	MOBILE	2015		755	9	63		63	186	431	5,210	1,032	3,886	292	
8	ALABAMA	MONTGOMERY	2014	200,194	504	14	21		21	218	251	4,142	1,153	2,689	300	
9	ALABAMA	MONTGOMERY	2015		519	20	14		14	191	294	4,029	1,236	2,442	351	
10	ALASKA	ANCHORAGE	2014	301,306	1,209	6	193		193	247	763	5,515	629	4,474	412	33
11	ALASKA	ANCHORAGE	2015		1,615	16	323		323	271	1,005	5,732	811	4,516	405	63
12	ARIZONA	CHANDLER	2014	252,363	228	1	24		24	55	148	2,941	518	2,315	108	49
13	ARIZONA	CHANDLER	2015		206	0	26		26	51	129	2,524	384	2,033	107	26
14	ARIZONA	GILBERT	2014	235,430	113	0	3		3	31	73	1,670	239	1,368	63	10
15	ARIZONA	GILBERT	2015		92	0	13		13	14	65	1,588	241	1,283	64	12
16	ARIZONA	GLENDALE	2014	236,780	431	8	59		59	170	194	6,055	969	4,584	502	23
17	ARIZONA	GLENDALE	2015		436	5	49		49	161	221	6,290	1,066	4,701	523	26
18	ARIZONA	PEORIA	2014	164,722	120	1	17		17	27	75	1,752	294	1,365	93	3
19	ARIZONA	PEORIA	2015		146	1	16		16	25	104	1,676	367	1,207	102	5
20	ARIZONA	PHOENIX	2014	1,529,852	4,234	50	486		486	1,407	2,291	28,145	6,328	17,672	3,545	173
21	ARIZONA	PHOENIX	2015		4,513	52	509		509	1,502	2,450	27,356	6,645	17,089	3,622	
22	ARIZONA	SCOTTSDALE	2014	229,325	173	0	38		38	39	96	2,629	449	2,076	104	5
23	ARIZONA	SCOTTSDALE	2015		189	3	48		48	53	85	2,580	338	2,031	91	4
24	ARIZONA	SURPRISE	2014	125,043	63	3	4		4	19	43	938	167	758	73	9
25	ARIZONA	SURPRISE	2015		85	1	6		6	20	58	1,088	231	807	50	10
26	ARIZONA	TEMPE	2014	169,812	429	1	48		48	93	287	3,878	536	3,061	221	24
27	ARIZONA	TEMPE	2015		348	2	47		47	83	216	3,731	531	3,043	157	30
28	ARIZONA	TUCSON	2014	527,328					0							
29	ARIZONA	TUCSON	2015		1,706	16	213		213	516	961	16,680	1,784	13,919	977	58
30	CALIFORNIA	ANAHEIM	2014	346,956	537	5		37	37	203	292	4,206	619	2,915	672	9
31	CALIFORNIA	ANAHEIM	2015		628	10	65		65	211	342	5,018	724	3,472	822	
32	CALIFORNIA	ANTIOCH	2014	108,223	419	3		30	30	148	238	2,131	701	921	509	23
33	CALIFORNIA	ANTIOCH	2015		387	2	31		31	135	219	1,931	496	895	540	18
34	CALIFORNIA	BAKERSFIELD	2014	367,406	861	7		8	8	334	512	7,609	2,147	4,354	1,108	203
35	CALIFORNIA	BAKERSFIELD	2015		879	12	38		38	324	505	7,748	1,885	4,661	1,202	204
36	CALIFORNIA	BERKELEY	2014	117,753	179	1	20		20	108	50	2,331	410	1,680	301	5
37	CALIFORNIA	BERKELEY	2015		269	0	20		20	165	84	2,903	587	1,971	351	9
38	CALIFORNIA	BURBANK	2014	105,041	70	1		4	4	27	38	1,264	161	1,011	92	9
39	CALIFORNIA	BURBANK	2015		69	0	5		5	25	39	1,282	143	1,047	92	4
40	CALIFORNIA	CARLSBAD	2014	112,297	101	1		8	8	22	70	894	188	655	51	2
41	CALIFORNIA	CARLSBAD	2015		88	0	15		15	18	55	1,028	188	782	58	5

Figure 3: Table 4 output

4 excel tables are created through this process:

1. Table_4_Cleaned.xls
2. Region.xls
3. Table_78_2014_Cleaned.xls
4. Table_78_2015_Cleaned.xls

The second part of the data cleansing and preparation is done in the SAS. In this project, we are using the SAS® OnDemand for Academics, a web-based SAS development environment. 4 tables created in the part one were uploaded into the SAS® OnDemand for Academics using the code below. SAS macro was used in this step as 4 tables are using the same code to import.
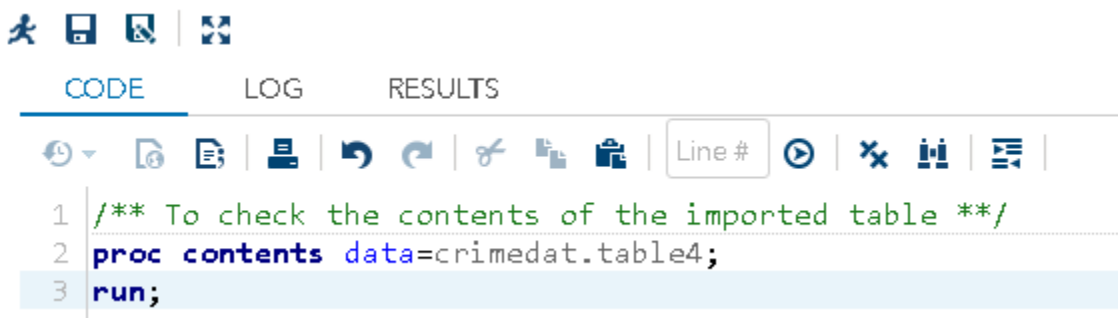

```
1  %macro import(xlsfile,name);
2  proc import datafile='/home/yeong_jin0/Raw Data/&xlsfile..xls'
3   dbms=xls
4   out=crimedat.&name;
5   getnames=yes;
6  run;
7  %mend import;
8  %import(table_4_cleaned,table4);
9  %import(region,region);
10 %import(table_78_2014_cleaned,employee2014);
11 %import(table_78_2015_cleaned,employee2015);
```

Figure 4: Import data

In this stage of cleansing and preparation, we zoom in into the data structure to ensure the data are imported correctly. The team is reviewing the result of the **Proc Contents** of the 4 tables to ensure the number of records, data type, length and formats are assigned correctly. (Refer to: Appendix 2)


```
1  /** To check the contents of the imported table **/
2  proc contents data=crimedat.table4;
3  run;
```

Figure 5: Sample code for proc contents

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	818
Obs in First Data Page	520
Number of Data Set Repairs	0
Filename	C:\Users\fyjn\Desktop\DAP\Data\Assignment\SASdata\table4.sas7bdat
Release Created	9.0401M3
Host Created	Linux

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
10	aggravated_assault	Num	8	BEST20.		aggravated_assault
15	arson	Num	8	BEST6.		arson
12	burglary	Num	8	BEST9.		burglary
2	city	Char	35	$35.	$35.	city
13	larceny_theft	Num	8	BEST15.		larceny_theft
14	motor_vehicle_theft	Num	8	BEST20.		motor_vehicle_theft
6	murder	Num	8	BEST7.		murder
4	population	Num	8	BEST11.		population
11	property_crime	Num	8	BEST16.		property_crime
8	rape_legacy	Num	8	BEST14.		rape_legacy
7	rape_revised	Num	8	BEST15.		rape_revised
9	robbery	Num	8	BEST9.		robbery
1	state	Char	16	$16.	$16.	state
5	violent_crime	Num	8	BEST14.		violent_crime
3	year	Num	8	BEST5.		year

Figure 6: Proc contents output for Table4

The team then zooms in into the values inside the table4, which is the main table for this analysis. The acceptance level of the missing value in this project is set as 5% for each crime, anything more than 5% will be removed as it might impact the accuracy of the final result. Proc Means was applied to produce the output.


```
1  /** To understand the data and identify the missing values vertically **/
2  proc means data=crimedat.table4 n nmiss;
3      class year;
4      var _numeric_;
5  run;
```

Figure 7: code for proc means

9

The next step is to access the missing value horizontally to review the missing crimes for each city. Simple transformation was used in this step to analyze the missing value for each city (Refer to appendix 4). In this assignment, anything more than 50% missing crime horizontally will be removed from the final analysis.

The last step is to verify the data consistency. At this stage, we've noticed that the population is missing for 2015. So, we will measure the consistency based on 2014 population (Refer to appendix 5).

Data Consistency

Obs	Year	Violent Crime	Murder	Robbery	Aggravated Assault	Property Crime	Burglary	Larceny Theft	Motor Vehicle Theft	Arson	Sexual Assault
1	2014	0.29%	0.00%	0.10%	0.17%	1.56%	0.31%	1.08%	0.18%	0.01%	0.02%
2	2015	0.30%	0.00%	0.10%	0.18%	1.58%	0.30%	1.10%	0.18%	0.01%	0.02%

These numbers are based on the data from FBI UCR

Figure 8: Consistency result

Based on the output above, the data is consistent across both years as there is no significant spike or drop in the percentage for every crime.

From the steps above, the team has summarized some observations which need to be addressed before we start the analysis.

1. Additional columns accidentally created due to the excel formatting during the proc import. These columns will be removed when the team prepare the final data for the analysis

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat	Label
7	G	Char	9	$9.	$9.	G
8	H	Char	9	$9.	$9.	H
9	I	Char	9	$9.	$9.	I
10	J	Char	9	$9.	$9.	J
11	K	Char	9	$9.	$9.	K
12	L	Char	9	$9.	$9.	L
2	city	Char	44	$44.	$44.	city
3	population	Num	8	BEST20.		population
1	state	Char	32	$32.	$32.	state
6	total_civilians	Num	8	BEST20.		total_civilians
4	total_employees	Num	8	BEST20.		total_employees
5	total_officers	Num	8	BEST20.		total_officers

Figure 9: Proc contents output for table employee2015

2. Missing Population for all the cities in year 2015

The MEANS Procedure

year	N Obs	Variable	Label	N	N Miss
2014	260	year	year	260	0
		population	population	260	0
		violent_crime	violent_crime	255	5
		murder	murder	255	5
		rape_revised	rape_revised	157	103
		rape_legacy	rape_legacy	98	162
		rape_total	rape_total	260	0
		robbery	robbery	255	5
		aggravated_assault	aggravated_assault	255	5
		property_crime	property_crime	254	6
		burglary	burglary	255	5
		larceny_theft	larceny_theft	254	6
		motor_vehicle_theft	motor_vehicle_theft	255	5
		arson	arson	249	11
2015	260	year	year	260	0
		population	population	0	260
		violent_crime	violent_crime	260	0
		murder	murder	260	0
		rape_revised	rape_revised	243	17
		rape_legacy	rape_legacy	17	243
		rape_total	rape_total	260	0
		robbery	robbery	260	0
		aggravated_assault	aggravated_assault	260	0
		property_crime	property_crime	258	2
		burglary	burglary	259	1
		larceny_theft	larceny_theft	260	0
		motor_vehicle_theft	motor_vehicle_theft	259	1
		arson	arson	244	16

Figure 10: Proc means output for table 4

In table4, the populations are missing for all the cities. To estimate the population for year 2015, the team is using the information from United States Census Bureau below:

2015

Region	Population	Percentage	Growth Rate
Northeast	56,184,737	17.50%	0.12%
Midwest	67,838,387	21.10%	0.17%
West	75,834,288	23.60%	1.08%
South	121,039,206	37.70%	1.12%

2014

Region	Population	Percentage	Growth Rate
Northeast	56,116,791	17.60%	0.23%
Midwest	67,726,368	21.30%	0.27%
West	75,023,986	23.60%	1.05%
South	119,696,311	37.60%	1.07%

Figure 11: Based on the data publish in United States Census Bureau.

3. Both rape 2 types of rape in the report (legacy and revised)
   Based on the output in the proc means above, the team identified high missing values in both rape categories. However, this is mainly due to the change of the rape definition in 2013. The team then decided to combine both rape categories in the transformation stage.

4. In year 2014, several cities are having missing records for all the crimes.

Missing Crime Info (> 50%)

Obs	year	state	city	Missing Crime %
1	2014	ARIZONA	TUCSON	100.00%
2	2014	HAWAII	HONOLULU	100.00%
3	2014	TEXAS	TYLER	100.00%
4	2014	UTAH	PROVO	100.00%
5	2014	UTAH	WEST VALLEY	100.00%

These numbers are based on the data from FBI UCR

Figure 12: Proc print output for missing crimes cities

The team decided to remove the cities with all missing crimes in 2014 as it will underestimate the state's crime rate. To ensure the consistency, the same cities will also be removed from 2015.

5. During the process, the team found out that the property Crime in the table4 is the summation of burglary, larceny theft and motor-vehicle theft. However, 3 cities are having a blank value in the property crime. To solve this issue, we will drop the property crime in the table4 and re-create the property crime by sum up all the 3 types of crimes.

6. To create a total crime columns by sum up all the crime type.

Data Transformation

After the data review, the team proceeds to compile the final data for the analysis. The purpose is to prepare a dataset that incorporate all the changes on the finding above. In this stage, we also merge in the region data. Please refer to the appendix 6 for the detail.

Analysis

1. **Analyze and Understand the Overall Crime Rate (Excluding Arson) Across the United States.**
   (Excluding Tucson, Honolulu, Tyler, Provo and West Valley)

US Crime Stats : Overall Crimes by Year

	Estimated Population	# of Crimes	% of Crimes
2014	77,046,537	1,691,607	2.20
2015	77,652,372	1,684,426	2.17

These numbers are based on the data from FBI UCR

Overall crime rate in United States is around 2.10%-2.20% per population. In year 2014, the overall crime rate is at 2.20% with the estimated population of 77.05 million. Compare to 2014, the overall crime rate for the year 2015 is further improved; it dropped from 2.20% to 2.17%. The improvement is not only come from the increase of the population, the number of crimes is also dropped compare to year 2014.

US Crime Stats : Overall Crimes by Year & Region

		Estimated Population	# of Crimes	% of Crimes
2014	MIDWEST	11,682,716	299,630	2.56
	NORTHEAST	13,555,174	221,876	1.64
	SOUTH	25,435,359	664,595	2.61
	WEST	26,373,288	505,506	1.92
2015	MIDWEST	11,702,575	293,303	2.51
	NORTHEAST	13,571,439	211,739	1.56
	SOUTH	25,720,235	645,587	2.51
	WEST	26,658,123	533,797	2.00

These numbers are based on the data from FBI UCR

Although the overall crime rate is decreased in year 2015, the team is notice that there is a slight increase in the crime rate in the west side of United States (Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming, Alaska, California, Hawaii, Oregon, Washington). The crime rate of the West side of United States is increased from 1.92% in 2014 to 2.00% in 2015.

US Crime Stats : Overall Crimes by Year, Region & Division

			Estimated Population	# of Crimes	% of Crimes
2014	MIDWEST	EAST NORTH CENTRAL	8,757,516	217,061	2.48
		WEST NORTH CENTRAL	2,925,200	82,569	2.82
	NORTHEAST	MIDDLE ATLANTIC	12,201,374	193,169	1.58
		NEW ENGLAND	1,353,800	28,707	2.12
	SOUTH	EAST SOUTH CENTRAL	3,938,391	125,966	3.20
		SOUTH ATLANTIC	8,928,058	217,336	2.43
		WEST SOUTH CENTRAL	12,568,910	321,293	2.56
	WEST	MOUNTAIN	6,312,263	137,393	2.18
		PACIFIC	20,061,025	368,113	1.83
2015	MIDWEST	EAST NORTH CENTRAL	8,772,403	207,809	2.37
		WEST NORTH CENTRAL	2,930,172	85,494	2.92
	NORTHEAST	MIDDLE ATLANTIC	12,216,014	184,465	1.51
		NEW ENGLAND	1,355,425	27,274	2.01
	SOUTH	EAST SOUTH CENTRAL	3,982,501	122,251	3.07
		SOUTH ATLANTIC	9,028,054	215,275	2.38
		WEST SOUTH CENTRAL	12,709,680	308,061	2.42
	WEST	MOUNTAIN	6,380,438	139,830	2.19
		PACIFIC	20,277,685	393,967	1.94

These numbers are based on the data from FBI UCR

The team then further segment the data by the division and notice that the crime rate of 3 divisions across 2 regions are slightly increase compare the year 2014. The 3 divisions are West North Central in Midwest and all the divisions in the West.

*Please refer to appendix 7 for the full code.

2. Top 3 Types of Crime (Excluding Arson) in United States.

US Crime Stats : Overall Crimes (Excluding Arson)

year=2014 | year=2015

2014: 73.02, 0.17, 0.98, 4.48, 7.85, 13.49

2015: 72.04, 0.19, 1.06, 4.56, 8.18, 13.98

Legend: Aggravated Assault | Murder | Property Crime | Robbery | Sexual Assault | Violent Crime

These numbers are based on the data from FBI UCR
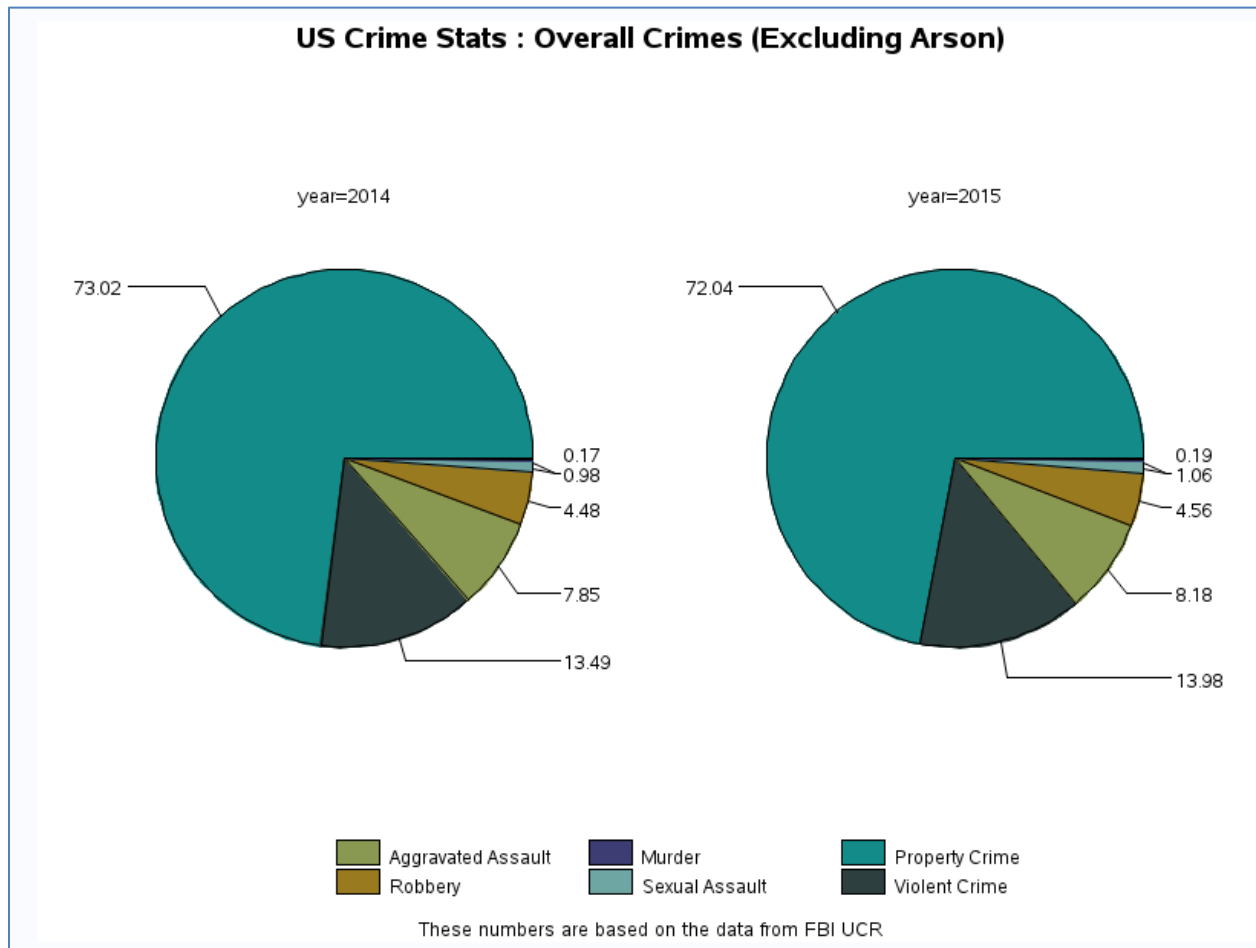
Based on the data in table4, the top 3 type of crime in the United States are Property Crime, Violent Crime and Aggravated Assault. The distributions are 73.02%, 13.49% and 7.85% respectively in 2014. In 2015, the top 3 are remaining the same with a slight improvement in property crime (73.02% in 2014 vs. 72.04% in 2015).

US Crime Stats : Property Crimes

year=2014

68.70

11.25

20.04

year=2015

69.51

11.67

18.82

Burglary   Larceny Theft   Motor Vehicle Theft

These numbers are based on the data from FBI UCR
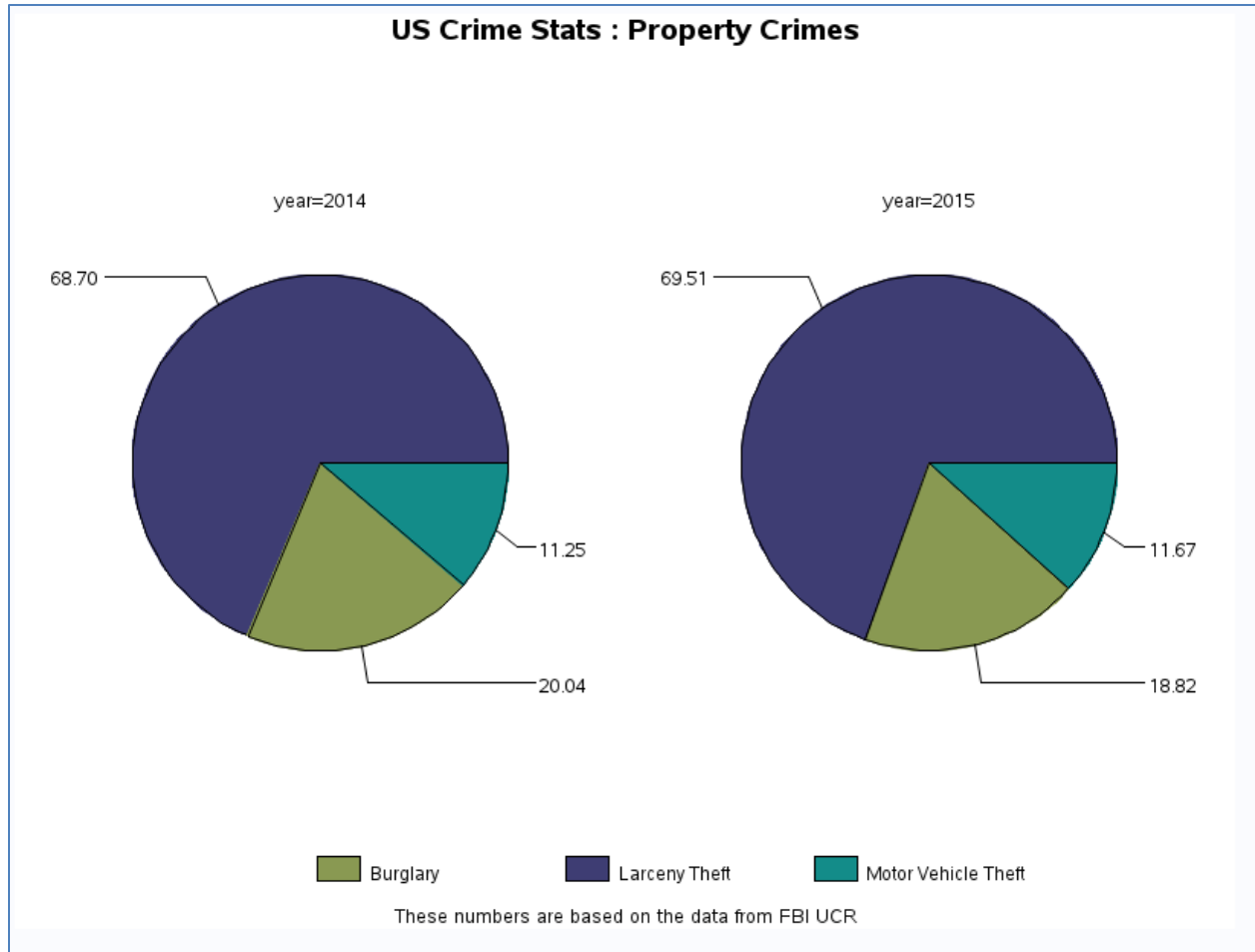
As the property crime consists of burglary, larceny theft and motor vehicle theft, the team decided to zoom in to the property crime to see which type of sub-crime is causing improvement. From the chart above, the team is notice that the improvement in the property crime is mainly due to decrease in burglary.

US Crime Stats : Property Crimes

These numbers are based on the data from FBI UCR
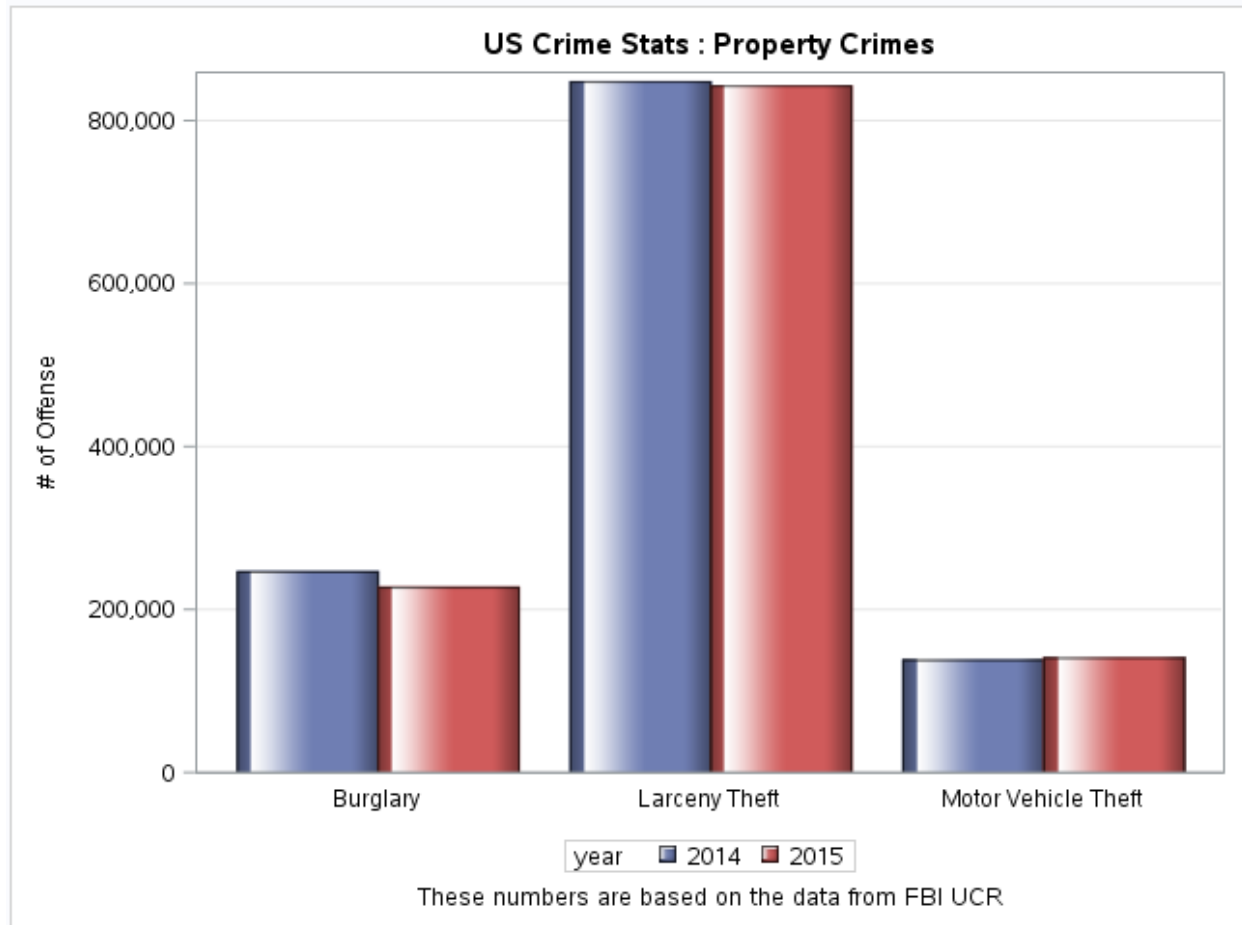
The improvement in burglary is not only on the distribution, the chart above shows that the number of offence is also lower compares to year 2014.

*Please refer to appendix 8 for the full code.

3. **Identify the Top 3 States with Highest Crime Rate in Year 2014 and Their Crime Rate in Year 2015**
    (Excluding Tucson, Honolulu, Tyler, Provo and West Valley)

US Crime Stats : Top 3 States with Highest Crime Rate (2014)

Obs	Year	State	Estimated Population	# of Crime	% of Crime
1	2014	UTAH	192,368	8,687	4.52%
2	2014	MISSISSIPPI	172,376	6,257	3.63%
3	2014	MISSOURI	1,186,439	42,940	3.62%

These numbers are based on the data from FBI UCR

The top 3 states with highest crime rate in year 2014 are Utah, Mississippi and Missouri. The crime rates are 4.52%, 3.63% and 3.62% respectively. The highest crime rate state, Utah, is more than 2X higher than the average crime rate in United States.

US Crime Stats : Top 3 States with Highest Crime Rate in 2014
Crime Rate in 2015

Obs	Year	State	Estimated Population	# of Crime	% of Crime
1	2015	UTAH	194,446	9,809	5.04%
2	2015	MISSISSIPPI	174,307	5,607	3.22%
3	2015	MISSOURI	1,188,456	44,497	3.74%

These numbers are based on the data from FBI UCR

In year 2015, the crime rate for Utah is further increase to 5.04% even the overall crime rate in United States is improving. Mississippi is the only city in the top 3 that having an improvement in the crime rate. The crime rate for Mississippi improved from 3.63% in 2014 to 3.22% in 2015.

*Please refer to appendix 9 for the full code.

4. Identify the Top 3 States with Lowest Crime Rate in Year 2014 and Their Crime Rate in Year 2015
(Excluding Tucson, Honolulu, Tyler, Provo and West Valley)

US Crime Stats : Top 3 States with Lowest Crime Rate (2014)

Obs	Year	State	Estimated Population	# of Crime	% of Crime
1	2014	IDAHO	216,260	2,742	1.27%
2	2014	NEW YORK	9,406,722	132,498	1.41%
3	2014	NORTH DAKOTA	115,686	1,677	1.45%

These numbers are based on the data from FBI UCR

The top 3 states with lowest crime rate in year 2014 are Idaho, New York and North Dakota. The crime rates are 1.27%, 1.41% and 1.45% respectively. The lowest crime rate state, Idaho, is more than 3.5X lower than the crime rate of the highest crime rate state, Utah.

US Crime Stats : Top 3 States with Lowest Crime Rate in 2014
Crime Rate in 2015

Obs	Year	State	Estimated Population	# of Crime	% of Crime
1	2015	IDAHO	218,596	2,708	1.24%
2	2015	NEW YORK	9,418,010	125,744	1.34%
3	2015	NORTH DAKOTA	115,883	1,866	1.61%

These numbers are based on the data from FBI UCR

In year 2015, the crime rate for Idaho is further improved to 1.24%. New York is also further improved from 1.41% to 1.34%. North Dakota is the only city in the top 3 that having a deterioration in the crime rate. The crime rate for North Dakota increased from 1.45 % in 2014 to 1.61 % in 2015.

*Please refer to appendix 10 for the full code.

5. **Compare the Police Employee Ratio of the Top 3 States with Lowest Crime Rate and the Top 3 States with Highest Crime Rate in Year 2014.**

The law enforcement employee data is retrieve from the Uniform Crime Reporting (UCR). In the raw data, the law enforcement employee is categorized into 2 which are:

1. Officer

   Individuals who ordinarily carry a firearm and a badge, have full arrest powers, and are paid from governmental funds set aside specifically to pay sworn law enforcement.

2. Civilian

   Include full-time agency personnel such as clerks, radio dispatchers, meter attendants, stenographers, jailers, correctional officers, and mechanics.

In the following analysis, only officer is considered as the civilian will not have any direct impact to the crimes. In addition, the state will be removed if the police employee data is missing or not available.

6 states below were selected for this analysis:

Top 3 highest crime rate states   : Utah, Missouri and Mississippi

Top 3 lowest crime rate states   : Idaho, New York and North Dakota

In the final report, Mississippi is excluded as the police employee data for the city Jackson is missing for year 2015 and this is the only city in Mississippi exceeds 100K population.

US Crime Stats : Police Employee Ratio

		2014	2015	2014	2015	2014	2015	2014	2015
		# of Crimes	# of Crimes	# of Police Employee	# of Police Employee	% of Crimes	% of Crimes	% of Police Employee	% of Police Employee
1	UTAH	16,042	18,150	428	433	8.34	9.33	0.22	0.22
3	MISSOURI	73,188	74,302	3,423	3,269	6.17	6.25	0.29	0.28
4	NORTH DAKOTA	2,980	3,394	144	156	2.58	2.93	0.12	0.13
5	NEW YORK	209,918	198,676	37,245	38,017	2.23	2.11	0.40	0.40
6	IDAHO	4,820	4,830	259	275	2.23	2.21	0.12	0.13

These numbers are based on the data from FBI UCR

Based on the proc tabulate above, it shows that the police employee ratio for the highest crime rate states are generally more than the lowest crime rate states except for New York. New York states has highest police employee ratio within the 6 states.

*Please refer to appendix 11 for the full code.

6. Compare the Crime Rate for The States Close to Mexico Border and The States Close to Canada Border.
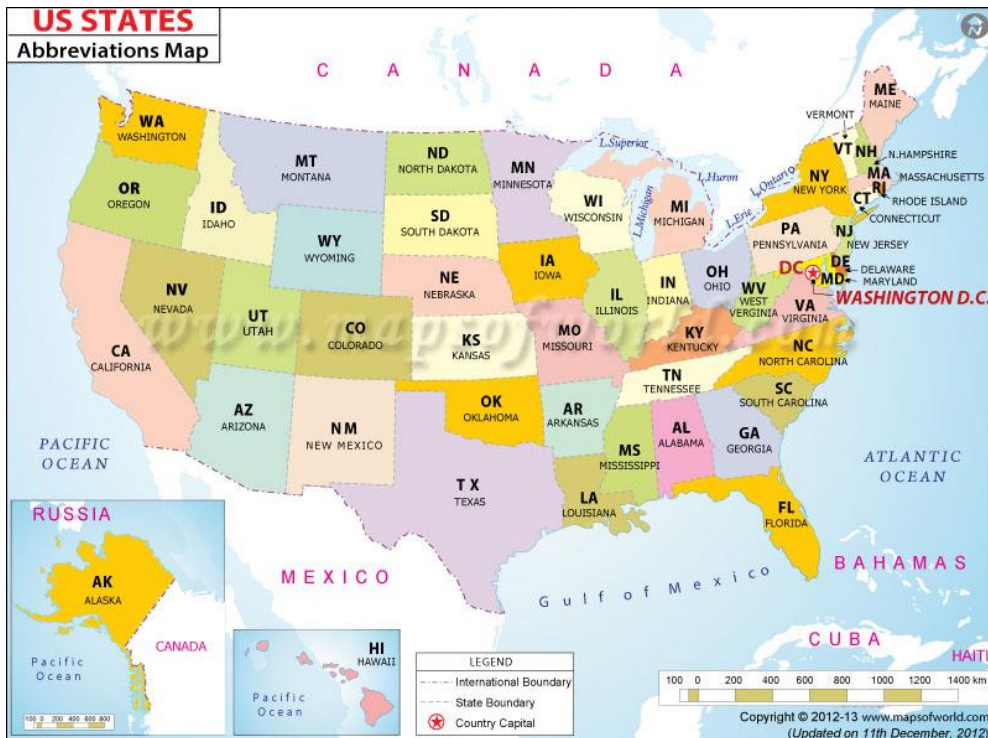
In this objective, 6 states are chosen for the comparison:

States close to Canada border : Washington, Montana and North Dakota

States close to Mexico border : Arizona, New Mexico and Texas

US Crime Stats : Mexico Border Vs. Canada Border

	2014	2015	2014	2015
	# of Crimes	# of Crimes	% of Crimes	% of Crimes
CANADA	59,334	54,771	3.22	2.94
MEXICO	339,831	330,098	2.43	2.33

These numbers are based on the data from FBI UCR

Surprisingly, the crime rate of the states close to Mexico border is actually lower than the states that close to the Canada border. The overall crime rate of the states close to Mexico is 2.33% while the states close to Canada is 2.97%.

US Crime Stats : Mexico Border Vs. Canada Border (States)

		2014	2015	2014	2015
		# of Crimes	# of Crimes	% of Crimes	% of Crimes
CANADA	MONTANA	2,735	2,835	2.48	2.54
	NORTH DAKOTA	1,677	1,866	1.45	1.61
	WASHINGTON	54,922	50,070	3.40	3.07
MEXICO	ARIZONA	59,662	58,863	2.03	1.98
	NEW MEXICO	19,798	21,235	3.54	3.76
	TEXAS	260,371	250,000	2.48	2.35

These numbers are based on the data from FBI UCR

However, when the team zooms into the states, we found that New Mexico is having the highest crime rate among the 6 states. The crime rate of New Mexico is 3.54% in year 2014 and increase to 3.76% in year 2015. The second highest crime rate state in this analysis is Washington. The crime rate for Washington is 3.40% in year 2014, and improved to 3.07% in year 2015.

*Please refer to appendix 12 for the full code.

Conclusion

Through this analysis, we understand that the overall crime rate of United States is in an improving trend. Overall crime rate has been reduced from 2.20% to 2.17%. This means that the current crime fighting strategy is performing well.

Property crime remains as the highest offence among all the crime types. However, we saw an improvement in burglary. We should identify the drivers that causing the improvement and try to implement for the larceny theft and motor vehicle theft to further reduce the property crime distribution.

As for the police employee, the data is not conclusive. However, New York is having a high police employee ratio (0.40%) and the crime rate is significant lower than the high crime rate states. In order to improve the crime rate for Utah, Missouri and Mississippi, local government can increase the police employee ratio to New York level if the states' financial is allowed. The increase in police employee will surely help in reducing the crime rate.

Crime rate of the states close to Mexico border is lower than the states close to Canada border. However, the highest crime rate among the 6 states that were chosen for this analysis is New Mexico (Mexico Border). In this case, building a barrier across the states close to the Mexico border might not be necessary. Focus on the New Mexico might be a good alternative to the barrier.

References

SAS 9.4 Programming Documentation. Available at:
http://documentation.sas.com/?cdcId=pgmmvacdc&cdcVersion=9.4&docsetId=basewn&docsetTarget=helpcenterfeedback.htm&locale=en

SAS Support Communities . Available at:https://communities.sas.com/

FBI: Uniform Crime Reporting. Available at: https://ucr.fbi.gov/

United States Census Bureau. Available at: https://www.census.gov/

Appendix

Appendix 1: Excel Preparation

Step 1

On the **Home** tab, in the **Editing** group, choose the **Find & Select** drop-down list and then click **Go To Special...**: and select the **Blanks** option and click **OK.**

Step 2

On the **Formula bar**, type an equal sign (=) followed by the address of the first cell with an entry in the column (=A6) and press **Ctrl + Enter**.

Step 3

Reselect the original range and press **Ctrl + C** to copy the selection. On the **Home** tab, in the **Clipboard** group, choose the **Paste** drop-down list and then click **Paste Values** to convert the formulas to values.

Appendix 2: Proc Contents Codes

```
/** To check the contents of the imported table **/
proc contents data=crimedat.table4;
run;

proc contents data=crimedat.region;
run;

proc contents data=crimedat.employee2014;
run;

proc contents data=crimedat.employee2015;
run;
```

Appendix 3: Proc Contents Output

Output for table Region

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	2143
Obs in First Data Page	51
Number of Data Set Repairs	0
Filename	/home/yeong_jin0/CrimeAnalysis/region.sas7bdat
Release Created	9.0401M3
Host Created	Linux
Inode Number	43651318
Access Permission	rw-r--r--
Owner Name	yeong_jin0
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
2	division	Char	20	$20.	$20.	division
3	region	Char	11	$11.	$11.	region
4	statecode	Char	9	$9.	$9.	statecode
1	states	Char	21	$21.	$21.	states

Output for table Employee2014

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	8
First Data Page	1
Max Obs per Page	1168
Obs in First Data Page	1145
Number of Data Set Repairs	0
Filename	/home/yeong_jin0/CrimeAnalysis/employee2014.sas7bdat
Release Created	9.0401M3
Host Created	Linux
Inode Number	43651319
Access Permission	rw-r--r--
Owner Name	yeong_jin0
File Size	1MB
File Size (bytes)	1179648

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
2	city	Char	45	$45.	$45.	city
3	population	Num	8	BEST11.		population
1	state	Char	30	$30.	$30.	state
6	total_civilians	Num	8	BEST15.		total_civilians
4	total_employees	Num	8	BEST20.		total_employees
5	total_officers	Num	8	BEST13.		total_officers

Output for table Employee2015

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	21
First Data Page	1
Max Obs per Page	430
Obs in First Data Page	419
Number of Data Set Repairs	0
Filename	/home/yeong_jin0/CrimeAnalysis/employee2015.sas7bdat
Release Created	9.0401M3
Host Created	Linux
Inode Number	43651320
Access Permission	rw-r--r--
Owner Name	yeong_jin0
File Size	3MB
File Size (bytes)	2883584

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
7	G	Char	32	$32.	$32.	G
8	H	Char	32	$32.	$32.	H
9	I	Char	32	$32.	$32.	I
10	J	Char	32	$32.	$32.	J
11	K	Char	32	$32.	$32.	K
12	L	Char	32	$32.	$32.	L
2	city	Char	44	$44.	$44.	city
3	population	Num	8	BEST20.		population
1	state	Char	32	$32.	$32.	state
6	total_civilians	Num	8	BEST20.		total_civilians
4	total_employees	Num	8	BEST20.		total_employees
5	total_officers	Num	8	BEST20.		total_officers

Appendix 4: Identify the Missing Values Horizontally

```sas
/** To understand the data and identify the missing values horizontally **/
data miss_crimes;

    set crimedat.table4;

    /** To combine 2 types of rape definition avoid the inaccurate missing
        value due to the change of definition **/
    if rape_revised=. and rape_legacy=. then sexual_assault=.;
    else

    sexual_assault=sum(rape_revised,rape_legacy);

    drop rape_revised rape_legacy;

    /** To calcultate number of missing in each crime **/
    all_crimes=sum(violent_crime=.,murder=.,robbery=.,aggravated_assault=.,
                   property_crime=.,burglary=.,larceny_theft=.,
                   motor_vehicle_theft=.,arson=.,sexual_assault=.);

    /** To derive the missing % **/
    crimes_miss_rate=all_crimes/10;

    /** To output the row with missing %>=50% **/
    if crimes_miss_rate>=0.5 then output;
run;

/** To print the row with missing %>=50% **/
proc print data=miss_crimes label;
    format crimes_miss_rate percent10.2;
    label crimes_miss_rate='Missing Crime %';
    var year state city crimes_miss_rate;
run;
```

Appendix 5: Consistency Check

```
/** Summarized the data by year and crime **/
proc summary data=work.all_crimes nway;
      class year;
      var population violent_crime murder robbery aggravated_assault
          property_crime burglary larceny_theft motor_vehicle_theft arson
          sexual_assault;
      output out=work.summary_all_crimes(drop=_freq_ _type_) sum=;
run;
/** To retain 2014 population and calculate the crime type **/
data work.crime_consistency;
      length new_population 8;
      set work.summary_all_crimes;
      retain new_population;
      if population ne . then new_population=population;
      /** To calculate the % of each crime type **/
      p_violent_crime          =     violent_crime/new_population;
      p_murder                 =     murder/new_population;
      p_robbery                =     robbery/new_population;
      p_aggravated_assault     =     aggravated_assault/new_population;
      p_property_crime         =     property_crime/new_population;
      p_burglary               =     burglary/new_population;
      p_larceny_theft          =     larceny_theft/new_population;
      p_motor_vehicle_theft    =     motor_vehicle_theft/new_population;
      p_arson                  =     arson/new_population;
      p_sexual_assault         =     sexual_assault/new_population;
run;
/** To print the consistency result **/
title 'Data Consistency';
proc print data=work.crime_consistency label;
      label    year                     = 'Year'
               p_violent_crime          = 'Violent Crime'
               p_murder                 = 'Murder'
               p_robbery                = 'Robbery'
               p_aggravated_assault     = 'Aggravated assault'
               p_property_crime         = 'Property Crime'
               p_burglary               = 'Burglary'
               p_larceny_theft          = 'Larceny Theft'
               p_motor_vehicle_theft    = 'Motor Vehicle Theft'
               p_arson                  = 'Arson'
               p_sexual_assault         = 'Sexual Assault';
      format   p_violent_crime          percent10.2
               p_murder                 percent10.2
               p_robbery                percent10.2
               p_aggravated_assault     percent10.2
               p_property_crime         percent10.2
               p_burglary               percent10.2
               p_larceny_theft          percent10.2
               p_motor_vehicle_theft    percent10.2
               p_arson                  percent10.2
               p_sexual_assault         percent10.2;
      var year p_violent_crime p_murder p_robbery p_aggravated_assault
          p_property_crime p_burglary p_larceny_theft p_motor_vehicle_theft
          p_arson p_sexual_assault;
      footnote "These numbers are based on the data from FBI UCR" ;
run;
```

Appendix 6: Compile Data for Analysis

```
/** Finalized the final data set for analysis **/
data work.table4_tmp_v1;
      set crimedat.table4(drop=property_crime);
      /** To remove the cities with >=50% missing crimes**/
      if    (state='ARIZONA' and city='TUCSON')      or
            (state='HAWAII' and city='HONOLULU')     or
            (state='TEXAS' and city='TYLER')         or
            (state='UTAH' and city='PROVO')          or
            (state='UTAH' and city='WEST VALLEY') then delete;

      /** to combine 2 types of rape definition **/
      if rape_revised=. and rape_legacy=. then sexual_assault=.;
      else  sexual_assault=sum(rape_revised,rape_legacy);

      property_crime=sum(burglary,larceny_theft,motor_vehicle_theft);

      /* To calculate the total crimes **/
      total_crimes=sum(violent_crime,murder,robbery,aggravated_assault,
                       burglary,larceny_theft,motor_vehicle_theft,
                       sexual_assault);

      drop rape_revised rape_legacy arson;
run;

/** Merge the table4 with the divion, region and state code **/
proc sort data=work.table4_tmp_v1;
      by state;
run;

proc sort data=crimedat.region;
      by states;
run;

data work.table4_tmp_v2;
      merge work.table4_tmp_v1(in=a)
              /** Rename the column states to state for merging **/
          crimedat.region(in=b rename=(states=state));
      by state;
      if a;
run;

/** Assign the estimated population to year 2015 **/
proc sort data=work.table4_tmp_v2;
      by state city year;
run;
```

```
data crimedat.table4_final1;
    set work.table4_tmp_v2;
    by state city;

    /** Retain the population for year 2014 for the calculation **/
    retain new_population;

    /** Assign the 2014 population to the new column new_population **/
    if first.city then new_population=population;

    /** Based on the 2014 population and region to estimate the 2015
        population **/
    else do;
        if      region='NORTHEAST'    then
        new_population=round(new_population+(new_population*0.0012));
        else if region='MIDWEST'      then
        new_population=round(new_population+(new_population*0.0017));
        else if region='WEST'         then
        new_population=round(new_population+(new_population*0.0108));
        else if region='SOUTH'        then
        new_population=round(new_population+(new_population*0.0112));
    end;

    label state                   = 'State Name'
          city                    = 'City'
          year                    = 'Year'
          violent_crime           = 'Violent Crime'
          murder                  = 'Murder'
          robbery                 = 'Robbery'
          aggravated_assault      = 'Aggravated Assault'
          property_crime          = 'Property Crime'
          burglary                = 'Burglary'
          larceny_theft           = 'Larceny Theft'
          motor_vehicle_theft     = 'Motor Vehicle Theft'
          sexual_assault          = 'Sexual Assault'
          total_crimes            = 'Total Crimes';

    /** New_population will be used for the following analysis and drop the
        original column **/
    drop population;
run;
```

Appendix 7: Objective 1

```
/** Objective 1 **/
ods html body='/home/yeong_jin0/ODS Output/ Overall United States.htm'
style=HTMLBlue;

title 'US Crime Stats : Overall Crimes by Year';
proc tabulate data=crimedat.table4_final1 S=[foreground=black cellwidth=200
just=c];
     class year;
     var  new_population total_crimes;
     table year='', new_population='Estimated Population'*sum=''*f=comma14.
          total crimes='# of Crimes'*sum=''*f=comma14. total_crimes='% of
          Crimes'*pctsum<new_population>=''*f=8.2;
     footnote "These numbers are based on the data from FBI UCR" ;
run;

title 'US Crime Stats : Overall Crimes by Year & Region';
proc tabulate data=crimedat.table4_final1 S=[foreground=black cellwidth=200
just=c];
     class year region;
     var  new_population total_crimes;
     table year=''*region='', new_population='Estimated
          Population'*sum=''*f=comma14. total_crimes='# of
          Crimes'*sum=''*f=comma14. total crimes='% of
          Crimes'*pctsum<new_population>=''*f=8.2;
     footnote "These numbers are based on the data from FBI UCR" ;
run;

title 'US Crime Stats : Overall Crimes by Year, Region & Division';
proc tabulate data=crimedat.table4_final1 S=[foreground=black cellwidth=200
just=c];
     class year region division;
     var  new_population total_crimes;
     table year=''*region=''*division='', new_population='Estimated
          Population'*sum=''*f=comma14. total_crimes='# of
          Crimes'*sum=''*f=comma14. total crimes='% of
          Crimes'*pctsum<new_population>=''*f=8.2;
     footnote "These numbers are based on the data from FBI UCR" ;
run;
ods html close;
```

Appendix 8: Objective 2

```sas
/** Objective 2 **/
/** To summarized the crime type **/
proc summary data=crimedat.table4_final1 nway;
      class year;
      var violent_crime murder robbery aggravated_assault property_crime
          burglary larceny_theft motor_vehicle_theft sexual_assault
          total_crimes;
      output out=summary_year(drop=_freq_ _type_) sum=;
run;


/** To transpose the data **/
proc transpose data=summary_year out=table4_2014_trans;
      where year=2014;
run;


proc transpose data=summary_year out=table4_2015_trans;
      where year=2015;
run;


/** To rename the transposed data **/
data table4_2014_trans_v2;
      set table4_2014_trans;
      where _label_ not in ('Year','Total Crimes');
      year=2014;
      rename _label_    = crime_type
             col1       = no_of_offence;
      label _label_     = 'Crime Type'
            col1        = '# Offence';
      drop _name_;
run;


data table4_2015_trans_v2;
      set table4_2015_trans;
      where _label_ not in ('Year','Total Crimes');
      year=2015;
      rename _label_    = crime_type
             col1       = no_of_offence;
      label _label_     = 'Crime Type'
            col1        = '# Offence';
      drop _name_;
run;

/** To calculate the crime type distribution **/
proc freq data=table4_2014_trans_v2 noprint;
   where crime_type not in ('Burglary','Larceny Theft','Motor Vehicle
                            Theft');
   tables year*crime_type/out=table4_2014_trans_v3;
   weight no_of_offence;
run;
```

```sas
proc freq data=table4_2015_trans_v2 noprint;
    where crime_type not in ('Burglary','Larceny Theft','Motor Vehicle
                             Theft');
    tables year*crime_type/out=table4_2015_trans_v3;
    weight no_of_offence;
run;

/** To combine table 2014 and 2015 **/
data table4_1415_trans_v3;
  set table4_2014_trans_v3
      table4_2015_trans_v3;
run;


ods html path='/home/yeong_jin0/ODS Output/'
         body='Overall Crimes.html';
/* Define the titles */
title 'US Crime Stats : Overall Crimes (Excluding Arson)';

/* Define pattern color for each crime type    */
pattern1 color=day;        /* Aggravated assault */
pattern2 color=mob;        /* Murder             */
pattern3 color=vibg;       /* Property Crime     */
pattern4 color=deoy;       /* Robbery            */
pattern5 color=libg;       /* Violent Crime      */
pattern6 color=dabg;       /* Sexual assault     */

legend1 label=none
        position=(bottom)
        offset=(4,)
        across=3
        order=("Aggravated Assault" "Murder" "Property Crime" "Robbery"
"Sexual Assault" "Violent Crime" )
        value=(color=black)
        shape=bar(4,1.5);

/* Create the pie chart */
proc gchart data=table4_1415_trans_v3;
   pie crime_type / sumvar=percent
                    descending
                           other=0
                    legend=legend1
                           value=none
                           across=2
                    value=arrow
                    coutline=black
                    noheading
                           group=Year;
 footnote "These numbers are based on the data from FBI UCR" ;
run;
quit;
ods html close;
```

```
/* To calculate the % within property crime */
proc freq data=table4_2014_trans_v2 noprint;
    where crime_type in ('Burglary','Larceny Theft','Motor Vehicle Theft');
    tables year*crime_type/out=table4_2014_trans_propc;
    weight no_of_offence;
run;

proc freq data=table4_2015_trans_v2 noprint;
    where crime_type in ('Burglary','Larceny Theft','Motor Vehicle Theft');
    tables year*crime_type/out=table4_2015_trans_propc;
    weight no_of_offence;
run;

data table4_1415_trans_propc;
  set table4_2014_trans_propc
      table4_2015_trans_propc;
run;

ods html path='/home/yeong_jin0/ODS Output/'
         body='Property Crimes.html';
/* Define the titles */
title 'US Crime Stats : Property Crimes';

/* Define pattern color for each crime type   */
pattern1 color=day;        /* Burglary   */
pattern2 color=mob;        /* Larceny Theft */
pattern3 color=vibg;       /* Motor Vehicle Theft          */

legend1 label=none
        position=(bottom)
        offset=(4,)
        across=3
        order=("Burglary" "Larceny Theft" "Motor Vehicle Theft" )
        value=(color=black)
        shape=bar(4,1.5);

/* Create the pie chart */
proc gchart data=table4_1415_trans_propc;
   pie crime_type / sumvar=percent
                    descending
                    other=0
                    legend=legend1
                    value=none
                    across=2
                    value=arrow
                    coutline=black
                    noheading
                         group=Year;
footnote "These numbers are based on the data from FBI UCR" ;
run;
quit;
```

```sas
/** To create bar chart **/
title 'US Crime Stats : Property Crimes';
proc sgplot data=work.Table4_1415_trans_propc;
  format COUNT comma10.;
  label  COUNT = '# of Offense';
  vbar crime_type / response=COUNT group=year groupdisplay=cluster
    stat=sum dataskin=gloss;
  xaxis display=(nolabel noticks);
  yaxis grid;
  footnote "These numbers are based on the data from FBI UCR" ;
run;
ods html close;
```

Appendix 9: Objective 3

```sas
/** Objective 3 **/
/** To summarize the population and total crimes by year and states**/
proc summary data=crimedat.table4_final1 nway;
     class year state;
     var new_population total_crimes;
     output out=summary_crime_population(drop=_freq_ _type_) sum=;
run;

/** To select only 2014 **/
data crimehigh_population_2014;
  set summary_crime_population;
  where year=2014;
  crime_rate=total_crimes/new_population;
run;

proc sort data=crimehigh_population_2014;
  by descending crime_rate;
run;

ods html path='/home/yeong_jin0/ODS Output/'
        body='Top 3 Highest Crimes.html';
/** To print the result for the Top 3 **/
title 'US Crime Stats : Top 3 States with Highest Crime Rate (2014)';
proc print data=crimehigh_population_2014(obs=3) label style(header)={just=c
foreground=black}
  style(table)={width=100%};
  format new_population      comma14.
         total_crimes        comma14.
         crime_rate          percent7.2;
  label  year               = 'Year'
         state              = 'State'
         new_population      = 'Estimated Population'
         total_crimes        = '# of Crime'
            crime_rate        = '% of Crime';
  var year                    / style(data)={just=c};
  var state                   / style(data)={just=c};
  var new_population          / style(data)={just=c};
  var total_crimes           / style(data)={just=c};
  var crime_rate             / style(data)={just=c};
  footnote "These numbers are based on the data from FBI UCR" ;
run;
```

```sas
/** To select only the top 3 2014 in the 2015 data **/
/** To summarize the population and total crimes by year and states**/
data crimehigh_population_2015;
  set summary crime population;
  where year=2015 and state in ('UTAH','MISSISSIPPI','MISSOURI');

  if        state='UTAH'           then rank=1;
  else if   state='MISSISSIPPI'    then rank=2;
  else                                  rank=3;
  crime_rate=total_crimes/new_population;
run;


proc sort data=crimehigh_population_2015;
  by rank;
run;

/** To print the result for the Top 3 **/
title1 'US Crime Stats : Top 3 States with Highest Crime Rate in 2014';
title2 'Crime Rate in 2015';
proc print data=crimehigh_population_2015(obs=3) label style(header)={just=c
foreground=black}
  style(table)={width=100%};
  format new_population       comma14.
         total_crimes         comma14.
         crime_rate           percent7.2;
  label  year                 = 'Year'
         state                = 'State'
         new_population       = 'Estimated Population'
         total_crimes         = '# of Crime'
         crime_rate           = '% of Crime';
  var year                     / style(data)={just=c};
  var state                    / style(data)={just=c};
  var new_population           / style(data)={just=c};
  var total_crimes            / style(data)={just=c};
  var crime_rate              / style(data)={just=c};
  footnote "These numbers are based on the data from FBI UCR" ;
run;
ods html close;
```

Appendix 10: Objective 4

```
/** Objective 4 **/
proc summary data=crimedat.table4_final1 nway;
      class year state;
      var new_population total_crimes;
      output out=summary_crime_population(drop=_freq_ _type_) sum=;
run;

/** To select only 2014 **/
data crimelow_population_2014;
  set summary_crime_population;
  where year=2014;
  crime_rate=total_crimes/new_population;
run;

proc sort data=crimelow_population_2014;
  by crime_rate;
run;

ods html path='/home/yeong_jin0/ODS Output/'
         body='Top 3 Lowest Crimes.html';
/** To print the result for the Top 3 **/
title 'US Crime Stats : Top 3 States with Lowest Crime Rate (2014)';
proc print data=crimelow_population_2014(obs=3) label style(header)={just=c
foreground=black}
  style(table)={width=100%};
  format new_population        comma14.
         total_crimes          comma14.
         crime_rate            percent7.2;
  label  year                  = 'Year'
         state                 = 'State'
         new_population        = 'Estimated Population'
         total_crimes          = '# of Crime'
         crime_rate            = '% of Crime';
  var year                     / style(data)={just=c};
  var state                    / style(data)={just=c};
  var new_population           / style(data)={just=c};
  var total_crimes             / style(data)={just=c};
  var crime_rate               / style(data)={just=c};
  footnote "These numbers are based on the data from FBI UCR" ;
run;

/** To select only the top 3 2014 in the 2015 data **/
data crimelow_population_2015;
  set summary_crime_population;
  where year=2015 and state in ('IDAHO','NEW YORK','NORTH DAKOTA');

  if       state='IDAHO'          then rank=1;
  else if  state='NEW YORK'       then rank=2;
  else                                 rank=3;
  crime_rate=total_crimes/new_population;
run;
```

```
proc sort data=crimelow_population_2015;
   by rank;
run;

/** To print the result for the Top 3 **/
title1 'US Crime Stats : Top 3 States with Lowest Crime Rate in 2014';
title2 'Crime Rate in 2015';
proc print data=crimelow_population_2015(obs=3) label style(header)={just=c
foreground=black}
   style(table)={width=100%};
   format new_population        comma14.
          total_crimes          comma14.
          crime_rate            percent7.2;
   label  year                  = 'Year'
          state                 = 'State'
          new_population        = 'Estimated Population'
          total_crimes          = '# of Crime'
          crime_rate            = '% of Crime';
   var year                     / style(data)={just=c};
   var state                    / style(data)={just=c};
   var new_population           / style(data)={just=c};
   var total_crimes             / style(data)={just=c};
   var crime_rate               / style(data)={just=c};
   footnote "These numbers are based on the data from FBI UCR" ;
run;
ods html close;
```

Appendix 11: Objective 5

```sas
/** Objective 5 **/
/** Merge the law enforcement data to the dataset **/
data employee2014;
     set crimedat.employee2014;

     /** the city in table4 is upper case **/
     city=upcase(city);

     /** create a year for the merging **/
     year=2014;

     /** drop the duplicate or additional columns **/
     drop population;
run;
data employee2015;
     set crimedat.employee2015;

     /** the data in table4 is upper case **/
     city=upcase(city);

     /** create a year for the merging **/
     year=2015;

     /** drop the duplicate or additional columns **/
     drop population g h i j k l;
run;

proc sort data=employee2014 out=employee2014_tmp(keep=state city);
     by state city;
run;

proc sort data=employee2015 out=employee2015_tmp(keep=state city);
     by state city;
run;

/** Create a lookup table and only select the cities that have 2 years of
records **/
data employee_lookup;
     merge employee2014_tmp(in=a)
       employee2015_tmp(in=b);
     by state city;
     if a and b;
run;

/** Combine 2014 and 2015 law enforcement table **/
data total_employee;
     set employee2014
       employee2015;
run;

proc sort data=total_employee;
     by state city;
run;
```

```sas
/** Select the cities that have 2 years of records based on the lookup table
**/
data total_employee_final;
      merge total_employee(in=a)
          employee_lookup(in=b);
      by state city;
      if a and b;
run;


proc sort data=total_employee_final;
      by state city year;
run;


proc sort data=crimedat.table4_final1 out=table4_final1_tmp;
      by state city year;
run;

/** Merge with the transformed table4 and create the final data for analysis
**/
data crimedat.table4_final2;
      merge table4_final1_tmp(in=a)
          total_employee_final(in=b);
      by state city year;
      if a and b;
run;

/** Summarized the population total officers total crimes by year and state
**/
proc summary data=crimedat.table4_final2 nway;
      class year state;
      var new_population total_officers total_crimes;
      output out=summary_police_population(drop=_freq_ _type_) sum=;
run;

/** Select the highest top 3 and lowest top 3 states**/
data police_population;
  set summary_police_population;
  if state in ('UTAH','MISSISSIPPI','MISSOURI','IDAHO','NEW YORK','NORTH
DAKOTA');
  if     state='UTAH'         then rank=1;
  else if state='MISSISSIPPI'  then rank=2;
  else if state='MISSOURI'     then rank=3;
  else if state='NORTH DAKOTA' then rank=4;
  else if state='NEW YORK'     then rank=5;
  else                              rank=6;
run;


proc sort data=police_population;
  by rank year;
run;
```

```
ods html path='/home/yeong_jin0/ODS Output/'
        body='Poloce.html';
/** Print the result **/
title 'US Crime Stats : Police Employee Ratio';
proc tabulate data=police_population S=[foreground=black cellwidth=200
just=c];
      class state rank year;
      var  new_population total_crimes total_officers;
      table rank=''*state='',year=''*total_crimes='# of
Crimes'*sum=''*f=comma14. year=''*total_officers='# of Police
Employee'*sum=''*f=comma14.  year=''*total crimes='% of
Crimes'*pctsum<new_population>=''*f=8.2 year=''*total_officers='% of Police
Employee'*pctsum<new_population>=''*f=8.2;
      footnote "These numbers are based on the data from FBI UCR" ;
run;

ods html close;
```

Appendix 12: Objective 6

```
/** Objective 6 **/
/** To summarize the population and total crimes by year and states**/
proc summary data=crimedat.table4_final1 nway;
    where state in ('WASHINGTON', 'MONTANA', 'NORTH DAKOTA', 'ARIZONA', 'NEW
MEXICO', 'TEXAS');
        class year state;
        var new_population total_crimes;
        output out=summary_crime_border(drop=_freq_ _type_) sum=;
run;

/** Assign the border to each states**/
data crime_border_v2;
  set summary_crime_border;
  length border $6.;
  if state in ('WASHINGTON', 'MONTANA', 'NORTH DAKOTA') then border='CANADA';
  else                                                 border='MEXICO';
run;

ods html path='/home/yeong_jin0/ODS Output/'
        body='Border.html';
/** Print the result **/
title 'US Crime Stats : Mexico Border Vs. Canada Border';
proc tabulate data=crime_border_v2 S=[foreground=black cellwidth=200 just=c];
    class  year border;
    var  new_population total_crimes;
    table border='',year=''*total crimes='# of Crimes'*sum=''*f=comma14.
year=''*total_crimes='% of Crimes'*pctsum<new_population>=''*f=8.2 ;
    footnote "These numbers are based on the data from FBI UCR" ;
run;

/** Print the result **/
title 'US Crime Stats : Mexico Border Vs. Canada Border (States)';
proc tabulate data=crime_border_v2 S=[foreground=black cellwidth=200 just=c];
    class  year border state;
    var  new_population total_crimes;
    table border=''*state='',year=''*total_crimes='# of
Crimes'*sum=''*f=comma14.  year=''*total crimes='% of
Crimes'*pctsum<new_population>=''*f=8.2 ;
    footnote "These numbers are based on the data from FBI UCR" ;
run;
ods html close;
```