Crime Map of San Francisco characterized by Month and Incident Types

Sheikh Yasir Arafat

Introduction: San Francisco is not only the cultural, commercial, and financial center of the U.S. State of California but also the 17th most populous city in the United States [1]. Like all other densely populated cities, the crime rate of San Francisco is comparatively higher than the U.S. State of California's crime rate [2]. As a consequence, the police department of San Francisco needs to use more advanced and cautious techniques for fighting against crimes. In modern times, criminology has been widely studied to prevent crimes from happening in the future. It has also been used to analyze the various facets of criminal behavior. In addition to this, it has become more prevalent in recent years due to the increasing number of crimes that are likely to happen due to demographic locations. Ongoing rapid development and easy accessibility of information technology enable the police departments to perform crime analysis using visualization techniques. Automation and networking make it available to access massive amounts of crime data, typically in the form of crime statistics. In numerous fields, studies and research have shown that visualization techniques are valuable; in crime research, nevertheless, there is a general lack of its application. In order to efficiently and effectively process crime data, criminologists and law enforcement are in demand of a more powerful tool.

Mapping is one of the most commonly and widely used techniques of finding and exploring crime information. It is the most effective one as it has not only developed and designed to provide a platform for the public communities to share about the crime they encountered but also help the public to take precautions from being victims of criminal activities by raising their awareness. The seasonal fluctuation in crime is also an important factor. The US department of justice has reported that criminal seasonality has played an important role in observing and detecting crimes [3]. In this project, we aim to achieve the following objectives:

1. Sketch a map that will identify the locations of San Francisco where any specific type of incident has occurred in any mentioned month.

2. Examine how and whether different types of incidents in San Francisco have changed over the years.

This project applies the PROC SGMAP to construct the maps in SAS. This paper will discuss three different map types using SGMAP: "OpenStreetMap", "ERSI map" and "chord map". Choropleth maps are produced by using a combination of a map data set and a response data set. A map data set contains the information needed to draw map boundaries. A response data set contains the information that is to be displayed on the map such as crime location. In addition to x-y coordinates which are 'Latitude', and 'Longitude', map data sets also contain an identification variable by which we have identified San Francisco. However, for the "OpenStreetMap" and"ERSI map", we don't need any additional map data set. These processes automatically provides the maps information and only required the response data set. The next section contains the description of the crime data along with the data preparation and cleaning process. The techniques and strategies that we have used to overcome the challenges for developing maps are explained in the Strategy Employed section. The results section accommodates the visual display of the data and frequency analysis along with the findings of our project. Finally, the discussion section concludes the major findings and takeaway of this project.

Data Description: In this project, we are dealing with the data set named "SFPD Incident Report: 2018 to Present" which contains 514199 observations and 34 variables [4]. This data set compiles data from the department's Crime Data Warehouse (CDW) to provide information on incident reports files by the SFPD in CDW or filed by the public with the SFPD. Out of these 34 variables, we have 13 character variables and 21 numeric variables. The date and time variables are considered numeric variables. As our focus is to create maps that will detect the locations of San Francisco where any particular type of incident has occurred in any given month and explore the changes of incidents in San Francisco over the years, so we don't need to work on all the given variables. Out of all these variables, we will work on only five variables namely 'Incident_Category', 'Incident_Year', 'Incident_Date', 'Latitude', and 'Longitude'. The variable 'Incident_Category' contains different types of incidents reported in San Francisco and this data set has 48 types of incidents, including Larceny Theft, Assault, and Burglary. The variable 'Incident_Year' provides the information regarding the year in which year the incident has conducted. This data set contains four years of incident information from 2018 to 2021. The variables 'Latitude', and 'Longitude' provide the location where the incident has been conducted. Some incidents have been recorded without the location information and we have removed them from the data set and dealt with the incident type where the value of Latitude', and 'Longitude' are provided. The variable 'Incident_Date' provides us with the date on which the incident has conducted.

Strategy Employed: As we have a very large data set with a lot of variables, therefore the main challenge of this project is to read the data in a data step. Therefore, we have used "PROC IMPORT" to read our data. We have used the option "guessingrows" and "validvarname" in the proc. The "guessingrows" option allows us to read all the character values of observation and "validvarname" automatically defines the variable's category into SAS.

The next challenge, we have encountered is to create a variable that will give the name of the month in which the incident has occurred. We need this because based on our interest in this project, we need to find out the specific type of incident that has occurred in a given month. Unfortunately, we don't have any variable like that which will directly provide us the name of the month in which the incident has occurred. Therefore, we have created a variable called 'month_name' by using the "put" function where the argument is 'Incident_Date' variable which implies the name of the month in which the incident occurred. The created variable 'month_name' has some space in the character string which will create issues while creating maps. We have used the function named 'compress' to overcome this issue.

As we have mentioned earlier that, the variables 'Latitude', and 'Longitude' have some missing values. Therefore to overcome this, we have used the if statement for deleting the missing observations. By using the above-mentioned process, we have prepared our data for developing maps.

The "PROC FREQ" and "PROC SGPLOT" are adopted to perform the frequency analysis and create a bar graph respectively. These have helped us for extracting the most frequent incidents and enabled us to observe the changes of frequencies of these incidents with respect to time. Finally, we have used the "PROC SGMAP" for developing the maps. In this proc, we have to import the map and there are several ways that exist for importing maps. Each of these procedures has specific characteristics. We have used the "OpenStreetMap", "ERSI map" and "chord map". The scatter function is implemented for identifying the location of the incidents. The

"OpenStreetMap" and "ERSI map" will help us to visualize the location where the incident has been conducted and the "chord map" provides the boundaries of the city which helps us to understand which part of the city usually contains the most of the incidents.

Results: We have performed the frequency analysis to find out the most frequent incidents in San Francisco. In this data set, the most frequent type of incidents that are reported includes Larceny Theft, Malicious Mischief, Non-Criminal, and Assault. Larceny Theft has conducted almost 31%. The next frequent incident type contains Malicious Mischief, Non-Criminal, and Assault, and all of them have occurred around 6%. The bar graphs of these most frequent types of incidents show a decreasing crime rate for Larceny Theft, Non-Criminal, and Assault (Ref. Figure 1). Though we have only the information from 2018 to 2021 (October), the decreasing trend is promising. However, the rate of Malicious Mischief remains unchanged over the years (Ref. Figure 2). From the frequency analysis of the month, we have observed that over all the years, all the months contain around 8% to 9% of the incidents.

Now, we have developed several graphs characterized by month and incident types. For developing the graphs, we have chosen the most frequent type of incidents which includes Larceny Theft, Malicious Mischief, and Assault. Using the developed macros, we have sketched the map of San Francisco by detecting the location where the Larceny Theft has conduced in January. We have created two different types of maps for this circumstances by applying Open Street Map and Choropleth Map. The Open Street Map approach helps us to detect the crime location by visualizing the roads, whereas the Choropleth Map helps us to visualize the overall location of San Francisco. From Figure 4, we have observed that Larceny Theft has occurred all over san Francisco. However, the prevalence is more frequent in the North, North-East, and central north parts rater than central and west parts of San Francisco. Especially the northeast bay area is the most risk zone for Larceny Theft.

Then we have developed maps for assault in June by utilizing the ESRI and Choropleth approach. The ESRI provides more specific visuals of San Francisco such as the location of parks. We have also observed that assault is most frequently reported in the central north of San Francisco rather than any other city (Ref. Figure 5).

Similarly, we have created maps for Malicious Mischief. Figures 7 and 8 imply that the Malicious Mischief has frequently occurred in the North part of San Francisco in December. By using our developed macros, we can easily find out the exact locations of the incident by specifying the month name and incident type.

Discussion: In this project, the frequency analysis and bar graphs have assisted us to conclude that several types of incidents have changed over the years. Figures 1-2 provide support regarding this. The frequency analysis helps us to find out the most frequent incidents and we have sketched the maps for those incidents. Finally, the maps provide information regarding the places where the given type of incident frequently occurred in any specific month. We have developed macros that allow us to create maps for any kind of incident type for a specific month. This will help the police and administrator for developing maps and find out the exact locations. As a consequence, they can raise public awareness in a timely manner and the police can take the necessary actions such as increasing the patrolling and ensuring capturing the criminals.

References

[1] "Don't Call It 'Frisco': The History of San Francisco's Nicknames | by The Bold Italic Editors | The Bold Italic." [Online]. Available: https://thebolditalic.com/don-t-call-it-frisco-the-history-of-san-francisco-s-nicknames-the-bold-italic-san-francisco-5c14348d49c#.m605gfvq5. [Accessed: 18-Nov-2021].

[2] "Crime Dashboard | San Francisco Police Department." [Online]. Available: https://www.sanfranciscopolice.org/stay-safe/crime-data/crime-dashboard. [Accessed: 18-Nov-2021].

[3] C. R. Block, "Is Crime Seasonal ? Carolyn Rebecca Block Statistical Analysis Center Illinois Criminal Justice Information Authority William Gould , Chairman J . David Coldren , Executive Director Printed by authority of the State of Illinois January 198," no. January, 1984.

[4] "Police Department Incident Reports: Historical 2003 to May 2018 | DataSF | City and County of San Francisco." [Online]. Available: https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-Historical-2003/tmnf-yvry. [Accessed: 18-Nov-2021].

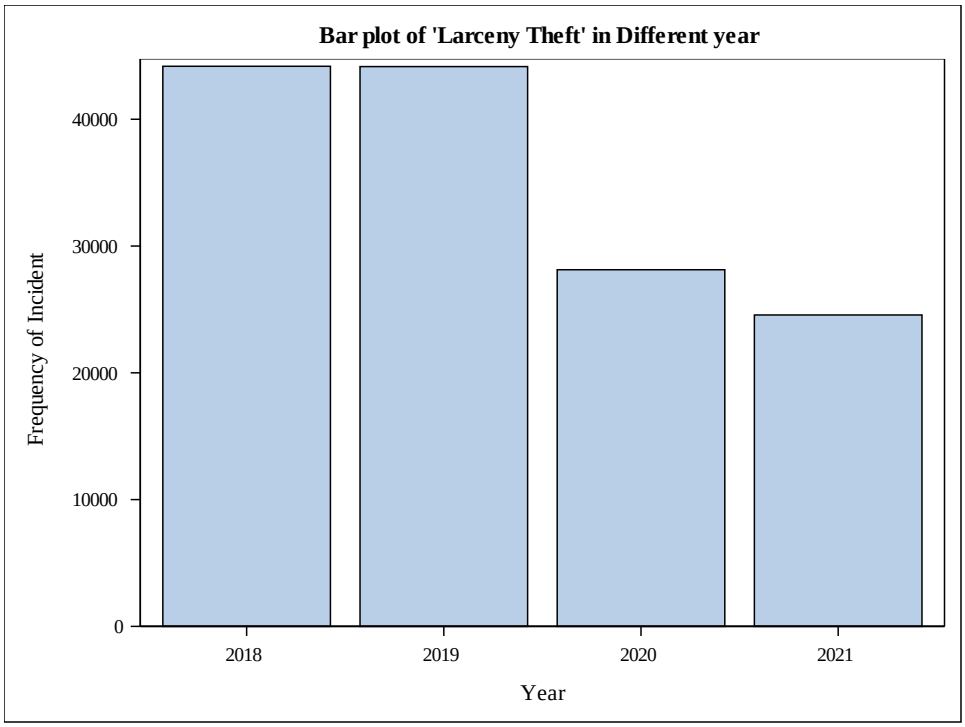

Figure 1: Bar plot of ‘Larceny Theft’ in different year.

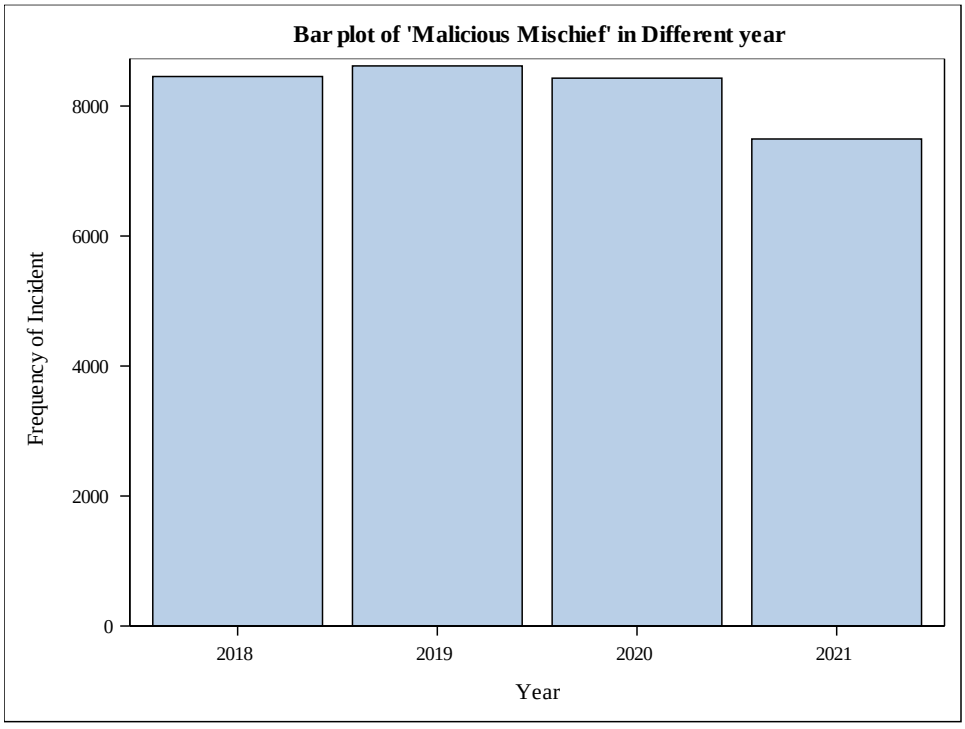

Figure 2: Bar plot of ‘Malicious Mischief’ in different year.

Figure 3: Map of San Francisco in January for Larceny Theft (using Open Street Map).

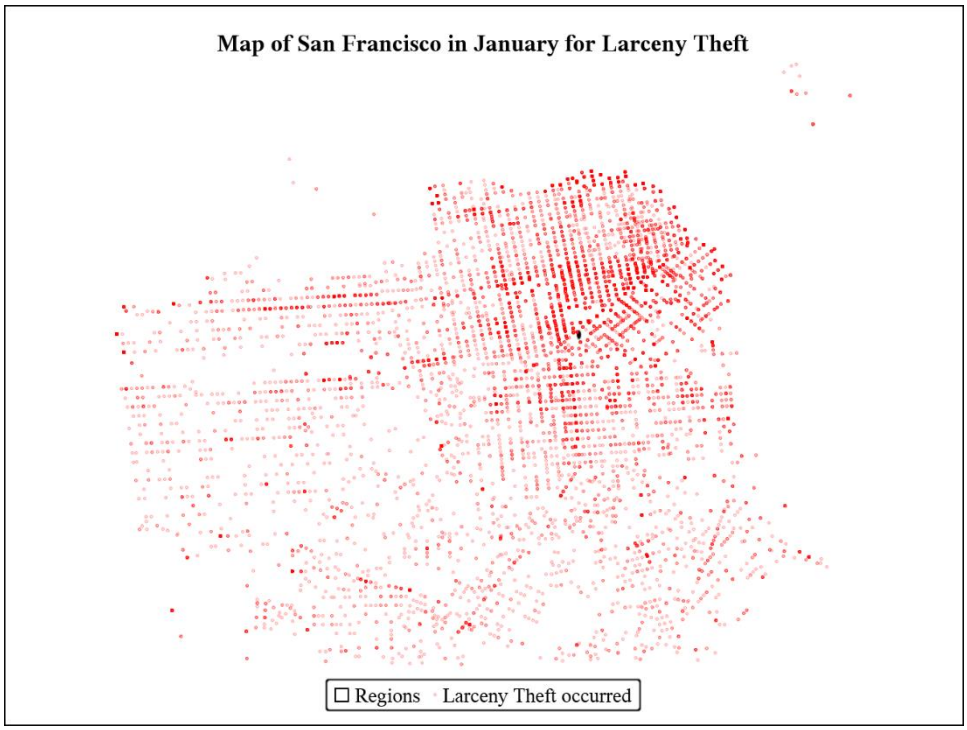

Figure 4: Map of San Francisco in January for Larceny Theft (using Choropleth Map).

Figure 5: Map of San Francisco in June for Assault (using ESRI Map).

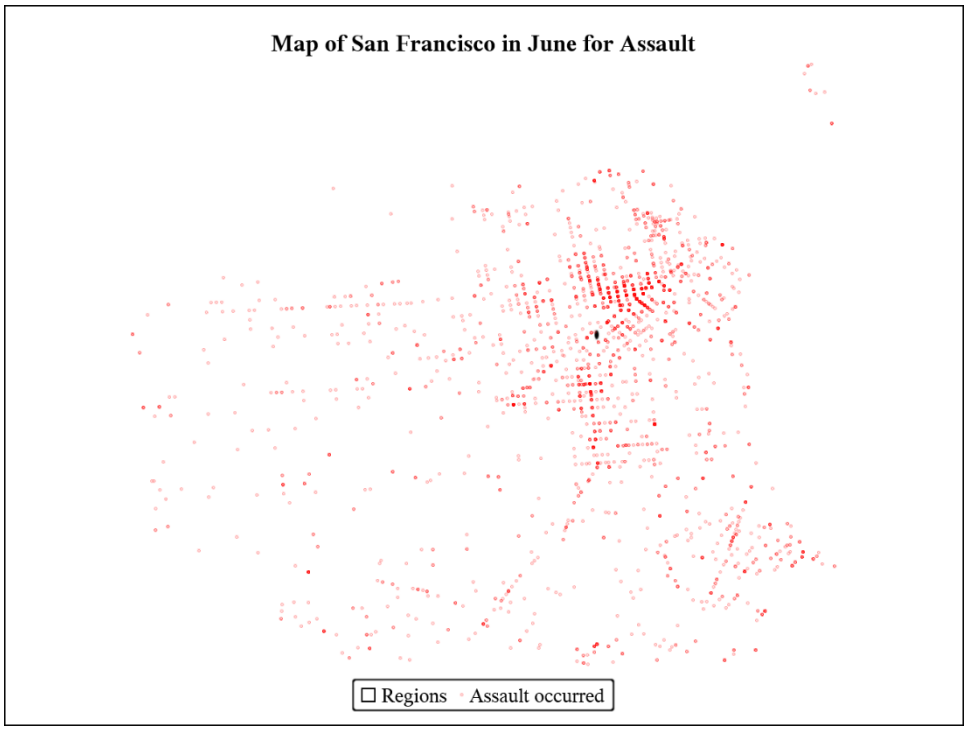

Figure 6: Map of San Francisco in June for Assault (using Choropleth Map).

Figure 7: Map of San Francisco in December for Malicious Mischief (using ESRI Map).

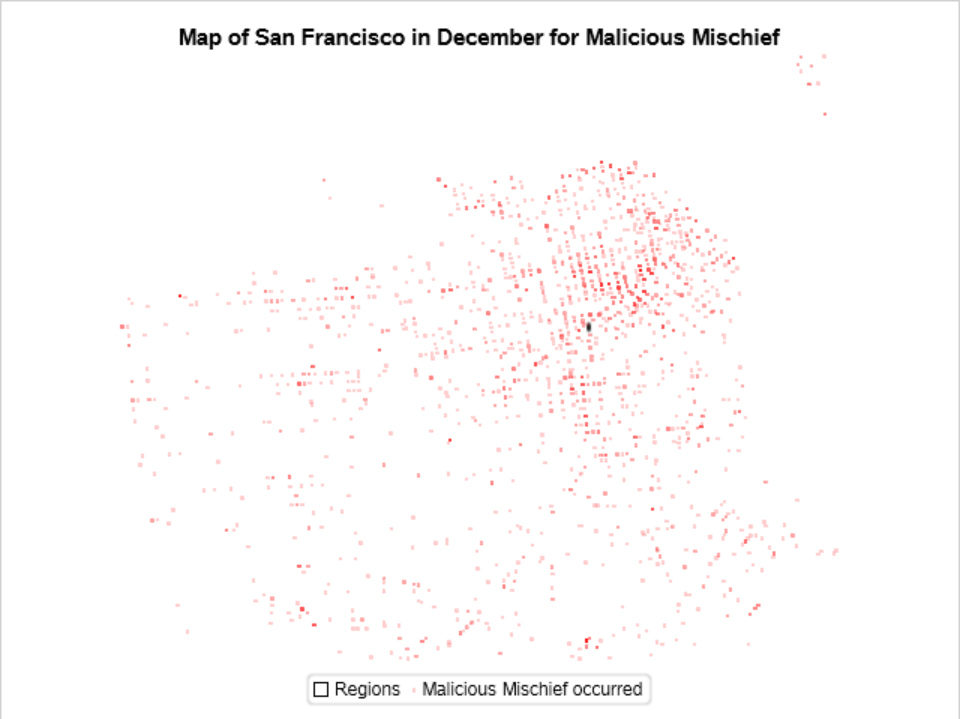

Figure 8: Map of San Francisco in December for Malicious Mischief (using Choropleth Map).

The SAS code that we have used for this project is provided below:

```
/*
Project Title: Crime Map of San Francisco characterized by Month and Incident
Types.
Student Name: Sheikh Yasir Arafat

In this program, i will create different types of map for the san Francisco
crime data set. These maps are characterized by Month and Incident type. I
have read the SFPD Incident Report: 2018 to Present data set which is
downloaded from "https://data.sfgov.org/Public-Safety/
Police-Department-Incident-Reports-Historical-2003/tmnf-yvry".

This data set contains 514199 observations and 34 variables. The most
frequent type incidents are Larceny Theft, Assault, Malicious Mischief and
Non-Criminal.

This program has four phases.

Phase 1: Read the data by Proc Import and clean and prepared the data in a
data step. The final data named as "Police_clean".

Phase 2: Use proc freq to extract information from the cleaned data set and
develop bar charts. I have created a macro named as "%Incident_year_bar ()"
for creating the bar charts by which we can see the frequency of incidents
according to year. The examples are given below:

%Incident_year_bar (Incident_type= 'Larceny Theft') *used for Figure 1;

%Incident_year_bar (Incident_type= 'Malicious Mischief') *used for Figure 2;

%Incident_year_bar (Incident_type= 'Assault')

%Incident_year_bar (Incident_type= 'Non-Criminal')


Phase 3: In this stage, we have created the maps. We have used the SGMAP
to develop maps. There are several ways available for importing map into
SGMAP. We have used three different approach and created three macros for
that. The macro named as "%map_open()" used openstreetmap for importing map,
similarly "%map_choro()" used choromap and "%map_esri()" used esrimap. Each
of these macros has two arguments, month and Incident. The examples of these
macros are given below:

%map_open(month=January,Incident=Larceny Theft) *used for Figure 3;
%map_open(month=June,Incident=Assault)

%map_choro(month=January,Incident=Larceny Theft) *used for Figure 4;
%map_choro(month=June,Incident=Assault) *used for Figure 6;
%map_choro(month=December,Incident=Malicious Mischief) *used for Figure 8;

%map_esri(month=November,Incident=Larceny Theft)
%map_esri(month=June,Incident=Assault) *used for Figure 5;
%map_esri(month=December,Incident=Malicious Mischief) *used for Figure 7;

Phase 4: Here, we have developed a single macro that combines all the three
```

```
macros in Phase 3 and provide all the three types of maps at the same time.
This macro is named as "%map_san()". Similar to the above macros, this also
has two arguments, month and Incident. The examples of this macros are given
below:

%map_san(month=December,Incident=Larceny Theft)
%map_san(month=February,Incident=Assault)

*/


/* Phase 1*/

/* Here we have read our data set by using proc import and the step cleaned
and prepared the data. Here, we have deleted the missing observations as
well.
Keeping only the required variables out of all 34 variables */

/* user should define the location of the data set*/
%let path=M:\STA 502A\Project;

options validvarname=v7; *create valid variable name;

/* proc import is used to read the data*/
      /*used the same data file name as downloaded from the google drive*/
proc import
datafile="&path\Police_Department_Incident_Reports__2018_to_Present.csv"
      out=Police /*output dataset name*/
      dbms=csv;
      guessingrows=max; /*consider the all the information without
trancation*/
run;

/* data cleaning and processing*/
data Police_clean;
      set Police; *select the original data;
      month_name=put(Incident_Date,monname.); *create a variable for month;
      if Latitude= '' then delete; *delete missing obs;
      if Longitude= '' then delete; *delete missing obs;
      keep Latitude Longitude Incident_Category month_name Incident_Year;
      /* ask to keep only the selected variables*/
run;


/* Phase 2 */

/* Here we have performed the freuency analysis and create the bar graphs
based on year. For enhancing the flexibility, we have created a macro for
developing bar graphs*/

/* using the following proc freq to create the frequency table of incidents*/
title "Frequency table of Different types of Incidents";
proc freq data=Police_clean  order = FREQ noprint;
            /*ask to sort by frequency*/
tables Incident_Category /nocum; /*ask not to provide the cumulative*/
run;
```

```
/* using the following proc freq to create the frequency table of months*/
title "Frequency table for the Months of Incidents";
proc freq data=Police_clean  order = FREQ noprint;
            /*ask to sort by frequency*/
tables month_name /nocum; /*ask not to provide the cumulative*/
run;

/* The following macro is used to create bar graphs*/
%macro Incident_year_bar (Incident_type=,);
title "Bar plot of &Incident_type in Different year";
proc sgplot data=Police_clean (where =(Incident_Category = &Incident_type));
      /* select the incident types*/
 vbar Incident_Year; /*ask to create bar plot*/
 xaxis label="Year";
 yaxis label='Frequency of Incident';
run;
%mend Incident_year_bar;


/* Phase 3 */

/* Here, we have created the macros for developing the maps.We have imported
three different types of maps for creating the final maps. Each of the macro
has two arguments namely month and Incident. By specifying these into the
macros,
we can easily create the maps*/

/* The following is used to create maps*/
/* The following macro used openstreetmap*/
%macro map_open(month=,Incident=); /*select month and incident type*/
data data_out;
      set Police_clean; /* select the data*/
      if (compress(month_name)="&month" and Incident_Category="&Incident")
then output;
run;
title "Map of San Francisco in &month for &Incident";
proc sgmap plotdata=data_out;
      openstreetmap; /*import map*/
    scatter x=Longitude y=Latitude / /*specify the locations*/
      markerattrs=(color=red size=2)
    legendlabel = "&Incident occurred";
run;
quit;
%mend map_open;



/* The following is used to create maps*/
/*Using choromap to create the maps*/
%macro map_choro(month=,Incident=); /*select month and incident type*/
data San_Francisco;
set mapsgfk.uscity (where=(STATECODE="CA" and City="San Francisco"));
      /*import the san Francisco map area border*/
run;
data data_out;
      set Police_clean; /*select the data*/
```

```
      if (compress(month_name)="&month" and Incident_Category="&Incident")
then output;
run;
title "Map of San Francisco in &month for &Incident";
proc sgmap mapdata=San_Francisco plotdata=data_out;
      choromap / mapid=ID; /*create border*/
    scatter x=Longitude y=Latitude /
      legendlabel = "&Incident occurred"
      markerattrs=(color=red size=2);
run;
quit;
%mend map_choro;


/* The following is used to create maps*/
/*Using ESRI MAPS to create the maps*/

%macro map_esri(month=,Incident=); /*select month and incident type*/
data data_out;
      set Police_clean; /*select the data*/
      if (compress(month_name)="&month" and Incident_Category="&Incident")
then output;
run;
title "Map of San Francisco in &month for &Incident";
proc sgmap plotdata=data_out;
      esrimap
url='http://services.arcgisonline.com/arcgis/rest/services/USA_Topo_Maps';
            /*import ersi map*/
   scatter x=Longitude y=Latitude /
      legendlabel = "&Incident occurred"
      markerattrs=(color=red size=2);
run;
quit;
%mend map_esri;



/* Phase 4 */

/* Here, we have created a single macro that have used all the three
different imported maps and provide three different maps at a time. This
macro also has two arguments namely month and Incident. By specifying these
into the macros, we can easily create all the three maps at a time*/

/* The following is used to create maps*/

%macro map_san(month=,Incident=); /*select month and incident type*/
data data_out;
      set Police_clean; /* select the data*/
      if (compress(month_name)="&month" and Incident_Category="&Incident")
then output;
run;
title "Map of San Francisco in &month for &Incident using Openstreetmap";
/* This will provide the map using openstreetmap*/
proc sgmap plotdata=data_out;
      openstreetmap; /*import map*/
    scatter x=Longitude y=Latitude / /*specify the locations*/
```

```
      markerattrs=(color=red size=2)
    legendlabel = "&Incident occurred";
run;
/* The following data step has san Francisco map area border which
will require for the choromap*/
data San_Francisco;
set mapsgfk.uscity (where=(STATECODE="CA" and City="San Francisco"));
      /*import the san Francisco map area border*/
run;
title "Map of San Francisco in &month for &Incident using Choromap";

/* This will provide the map using choromap*/
proc sgmap mapdata=San_Francisco plotdata=data_out;
      choromap / mapid=ID; /*create border*/
    scatter x=Longitude y=Latitude /
      legendlabel = "&Incident occurred"
      markerattrs=(color=red size=2);
run;
title "Map of San Francisco in &month for &Incident using Esrimap";
/* This will provide the map using esrimap*/
proc sgmap plotdata=data_out;
      esrimap
url='http://services.arcgisonline.com/arcgis/rest/services/USA_Topo_Maps';
            /*import ersi map*/
   scatter x=Longitude y=Latitude /
      legendlabel = "&Incident occurred"
      markerattrs=(color=red size=2);
run;
quit;
%mend map_san;


/* Discussion: Here, we have developed maps based on the month and incident
type. We have also create the bar graphs for visualizing the changes of
crimes with respect to time. The frequency analysis helped us to find out the
most frequent type of incidents. Finally, the macros enhance the flexibility
of developing maps. */
```