Trends and Predictions of Crimes & Crime Rates

Module: Introduction to Data Science

Module Code: INF6027

Student Number: 200226686

Word Count: 3150 words

Abstract

Background: Predictive policing refers to the use of law enforcement by statistical, predictive modeling, and other scientific approaches to detect suspected illegal behavior. It has been proven an effective technique to reduce the crimes from happening in the future to some extent.

Aims: To find if there are any patterns to crimes happening over different months and subsequently getting to know the factors contributing towards increase or decrease in crimes. Also comparing crimes of two counties (Derbyshire and South Yorkshire) from UK based on five different crimes (Burglary, Criminal damage and arson, Other theft, Violence and sexual offences, Vehicle crime) happened over the period of 3 years (2017-2020).

Methods: Publicly available UK Police data from 2017-2020 was used as a basis for this report. The crime data was retrieved for the counties of Derbyshire and South Yorkshire from the used Police Datasets. A combination of Line graph and Choropleth Maps were used to acquire the required result.

Results: Overall observations made from all the statistical graphs and maps suggests that the trends in crimes can be predicted based on past data and used as helping hand to safe-guard future possibilities of such incidents. And for the second part of the aim, apart from some regions in both counties, the crime rate and the crime type are similar. Also, Violence and sexual offences was the crime type that happened the most in both the regions considered for this study. But in the county of Derbyshire, the crime “other theft” was much more as compared to that of South Yorkshire.

Conclusions: The ability to forecast the locations of potential criminal incidents can, both from situational and logistical viewpoints, serve as a powerful source of intelligence for law enforcement. This study, of the possible applications of predictive mapping is by no means detailed, but it is intended to demonstrate the useful contributions that predictive models could provide and highlight the relevance of this field.

Contents

1. Introduction

The effort to eliminate and discourage violence and offenders is known as crime prevention. It is specifically applied with policymakers' attempts to prevent violence, preserve the rules, and enforce criminal justice. In order to address and implement the crime prevention, many techniques have been in development. Crime reduction has become an increasingly important aspect of many public safety and protection policies in the region. Hence, most of the public safety organizations and policymakers are investing huge sums on developing a reliable strategy to predict and prevent crimes. Many countries have been successful in implementing such strategies and methods in order to reduce unnecessary work load. By using these new techniques and in turn following the trends from the past crime data, police forces can now determine potential criminals and the probable locations of future events of disturbances. There are also a wide range of factors and circumstances that may affect or influence the predictions, for example: demographics, economic or financial instability, weather, specific time of the year (festivals or major events), environmental disasters, failing to report a crime.

The advent of big data has helped to capture greater awareness and generate clearer perspectives from the data through the application of new technologies. By using the data collected over the years and putting it to good use with the help of other tools such as Machine Learning and Artificial Intelligence could help in crime prevention and also reduce the money spent on unnecessary law enforcement.

As stated by (Bachner, J,2013), there are three different categories of analytical techniques that law enforcement departments use to predict crime:

- **Analysis of Space:** The plotting of criminal hot-spots, including regions in which there is a higher risk of crime than in the surrounding areas, is one of the initial applications of crime mapping.

- **Analysis of Time:** The plotting of criminal hot-spots based on past occurrences at a given point of time may help to predict the type of crimes in a specific area in the future.

- **Analysis of Social Network:** This method is primarily used to detect an individual as an interest instead of geographical location. It may include certain gangs or drug networks. But this may also contain some profiling of individuals which is done by biased algorithmic analysis.

1.1 Objective

The objectives of this study are:

- Using data processing methods to analyze the gathered data to assess the accuracy of predicting potential offenses based on different attributes.
- And, using Data Visualization techniques to compare occurrences of major type of crime in different regions of two counties of the UK.
- Finally, as we have past UK crime data as well as recent, we can use it to analyse and determine how the pandemic affected the occurrences of crimes.

1.2 Motivation

There are assumptions that, many factors contribute in the trends of crime rate. To find relevant graphical evidence to support these hypotheses, I tested the crimes happening in two counties in the UK which are namely; South Yorkshire and Derbyshire, based on their crime type and number of crimes over the years (2017-2020). Also, changes in the trends of a crime were worth analysing after the outbreak of the global pandemic.

1.3 Literature Review

1.3.1 Data Analysis

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data mining is particular analysis technique that focuses on statistical modelling. Chen, Chung, Xu, Wang, Qin & Chau (2004) in their paper stated that Traditional Data Mining techniques can be used without raising any issues, but when it comes to mining crime data, it raises privacy concerns. One of the first person-based predictive police systems in the United States was initiated by the Chicago Police Department. The initiative, called the "heat list" or "strategic subjects list," was first piloted in 2012, which created a list of individuals who were deemed most likely to commit or be a victim of gun abuse. Chicago police frequently touted the program as key to their strategy for combating violent crime (Tim Lau ,2020).

1.3.2 Time of the year and crime rates

Time of the year suggests that different factors like weather, economic, health, social and environmental conditions contribute towards increase or decrease in crime rate. Weather is one of the major factors that highly alters the crime rate. Studies found that high temperatures correspond with higher crime rate and low temperatures with lower crime rates. Similarly, by including these factors during predictive policing we also understand that types of crimes also change according to changing conditions.

1.3.3 Predictive Policing

Predictive policing refers to the usage of mathematical, predictive analytics, and other analytical techniques in law enforcement to identify potential criminal activity. This can be used in four ways:

- For predicting crimes
- For predicting offenders
- For predicting perpetrators
- For identifying victims

In order to make the best use of resources or to provide the highest chance of deterring or stopping potential incidents, predictive policing use evidence on the time, locations and extent of previous crimes to provide police strategists with information on when and at what times police patrols can patrol or maintain a presence.

An example of one of the first predictive policing used by the UK police was in Kent. The Kent police used machine-learning algorithms to forecast potential crimes in terms of place and time using crime type, location and time (Sarah Marsh,2019).

Another example is, A number of predictive policing programs have been implemented by the LAPD, including LASER, which detects places where gun violence is believed likely to occur, and PredPol, which measures "hot spots" for a high risk of property-related crimes (Tim Lau, 2020).

2. Methodology

A Crime Detection & Surveillance Architecture Focused on Spatial Analysis is applied for efficient analysis and prediction of crime events. Various visualization methods are used in this framework to evaluate the data in a better way. To create these visual representation, Rstudio and Microsoft Excel software were used.

2.1 Data Collection and Pre-processing

As data is acquired from the official U.K. Department of Police website (https://data.police.uk/data/), the dataset used for this report is accurate, true and genuine. The data set contains a total of 11 attributes out of which 5 attributes were considered for the study, they are crime type, lsoa code, date, latitude, and longitude. In this phase, the history of crimes from the year 2017-20 was considered as the training dataset. In the pre-processing phase, removal of the inconsistent data (such as missing values, redundant information, etc.) and transformation of the data is done that is required for the predicting the crime in the following phases of analysis. Crime data of two British counties namely: South Yorkshire and Derbyshire were selected as a subject of study for this report. The reason for selecting these two regions is because both counties can be considered as student cities as they have many colleges and universities situated in their respective regions. And hence it is important for law enforcement to maintain peace in these regions.

To download data for any specific police department in the UK, go to the above-mentioned link and then select the range for date, then select the department of police force specific to the region of interest and click on "Generate File", then a zip file for your selection will be created and then to download click on "Download now".

The shapefile is a common and influential spatial representation for data that we can use to construct a thematic map. The shapefile basically gives us information about the boundary of the regions which we can then fill to create thematic maps. Shapefiles for these two regions were obtained from official UK government open database which available freely (https://borders.ukdataservice.ac.uk/bds.html).
To download data for any region in the UK, from the drop-down boxes you can select "England", "Geography" and "2011 and later" respectively (then click find). In the "Boundaries" box select option "English Lower Layer Super Output Areas 2011" and then "List Areas". You

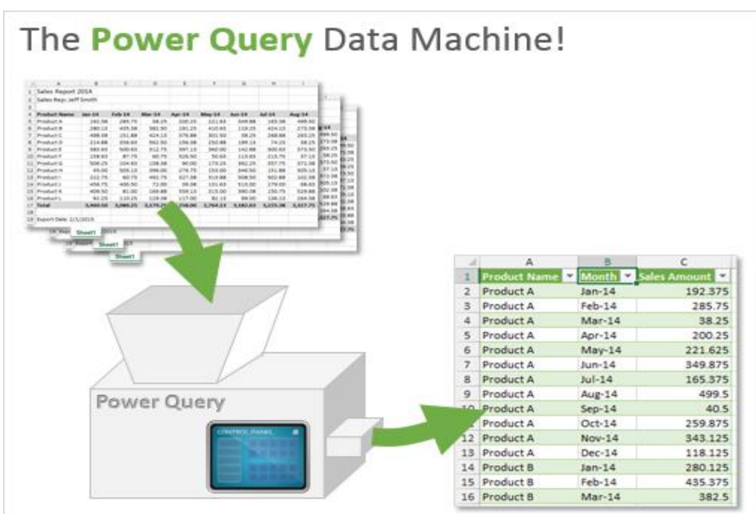
will see a list of UK cities/counties. Scroll down to the required region and then "Extract Boundary Data" to download the file ("BoundaryData.zip")

In the case of this report there was no one specific shapefile for the region of South Yorkshire. Hence shapefiles of the sub-regions: Sheffield, Rotherham, Doncaster and Barnsley were downloaded and read into different variables using function "readOGR()" from "rgdal" package. And then they were merged into one master file. This was done by using the union() function from the raster package in Rstudio.

```
library(raster)
#Obtaining shapefiles for different regions of south Yorkshire
sheffield<-readOGR(dsn="./BoundaryData", layer="england_lsoa_2011")
barnsley<-readOGR(dsn="./Barnsley", layer="england_lsoa_2011")
rotherham<-readOGR(dsn="./Rotherham", layer="england_lsoa_2011")
doncaster<-readOGR(dsn="./Doncaster", layer="england_lsoa_2011")

#Merging shapefiles to create a single shapefile

s_Yorkshire<-raster::union(sheffield,barnsley)
s_Yorkshire<-raster::union(s_Yorkshire,rotherham)
s_Yorkshire<-raster::union(s_Yorkshire,doncaster)
```

This wasn't the case with the shapefile for Derbyshire.

For the crime data of both the counties, as it was tabulated in different CSV files based on month, merging of the data was necessary to create a master file. Hence Power BI service provided by the Microsoft was used to merge all the data from different files into one single file.



source: Google Images

After creating a single master file for the crime data, it was read by using the function "read_xlsx()" from "readxl" package.

```r
#Loading South Yorkshire crime data
crimeData<- read_xlsx("South Yorkshire Crime Data.xlsx", sheet=2)

#Changing Names of the columns
crimeData<-rename(crimeData,c("LSOA code"="LSOA_code",
                              "Crime type"="crime_type"))


crimeData<- crimeData[!is.na(crimeData$Longitude),]

View(s_Yorkshire@data)

#Joing crime data and shapefile data by LSOA_code
s_Yorkshire@data<-left_join(s_Yorkshire@data, crimeData,
                              by=c('LSOA_code'='LSOA_code'))

#Removing unwanted columns
s_Yorkshire@data = subset(s_Yorkshire@data, select = c(crime_type,LSOA_code,Longitude,Latitude,Month))
```

To make the column names understandable for the R environment "rename()" function from dplyr package was used. To clean the data frame from NA value which were observed during inspection of the data was done using the built-in function "is.na()". Then the two data frames were joined together using a common variable that is "LSOA_code" with the function "left_join()". And then the unwanted data columns from the final data frame were removed using the "subset()" function.

The crime data and shapefile for the region of Derbyshire were handled and pre-processed in the same as that for the region of South Yorkshire.

The picture below shows a part of the final representation of the data frame after performing data cleaning and pre-processing functions and acquiring the necessary data for this study.

	crime_type	LSOA_code	Longitude	Latitude	Month
1	Bicycle theft	E01033271	-1.479108	53.37119	2017-11-01
2	Bicycle theft	E01033271	-1.476195	53.37221	2017-11-01
3	Criminal damage and arson	E01033271	-1.478018	53.37062	2017-11-01
4	Criminal damage and arson	E01033271	-1.481360	53.37383	2017-11-01
5	Criminal damage and arson	E01033271	-1.478622	53.37405	2017-11-01
6	Criminal damage and arson	E01033271	-1.476195	53.37221	2017-11-01
7	Drugs	E01033271	-1.476195	53.37221	2017-11-01
8	Other theft	E01033271	-1.478219	53.37259	2017-11-01
9	Other theft	E01033271	-1.476438	53.37324	2017-11-01
10	Other theft	E01033271	-1.478219	53.37259	2017-11-01

Attributes	
Name	Description
crime_type	Type of crimes occurred
LSOA_code	Geographic hierarchy codes
Longitude & Latitude	Geographic Co-ordinates
Month	date of the crime

Attributes Table

2.2 Data Visualization

Data visualization uses mathematical graphics, charts, info graphics and other methods to convey information accurately and effectively. The effective visualization helps us to analyze and reason about data and evidence. The aim of visualization is to gain insight into different aspects relevant to some process that we are interested in, such as a science simulation or some real-world process, by means of interactive graphics (Telea A. 2014). Two methods exist to visualize crime statistics on a map, they are: Hotspot Mapping and Heat map. Both hotspot analysis and heat maps are used to evaluate characteristics of spatial crime and present geographically mapped crime data with colorful representation on a map.

As described by (Wang, Ding, Lo, Stepinski, Salazar, & Morabito, 2013), Hotspot mapping is a method used to chart statistical data and display the density of crime from low incidents, i.e., "cold-spots" to high occurrences, i.e., "hotspots". Whereas Heat map presents the density of crime in these areas that actively experience illegal crimes, as opposed to hotspot mapping.

2.2.1 Visualizing Crime Trends

Line Graph or otherwise known as Trendline Graph are used over a constant interval or time span to show quantitative values. In order to display patterns and evaluate how the data has evolved over time, a line graph is most commonly used (Data Visualization Catalogue, n.d.). In this report, crime data of the two counties have been used to study the changes in occurrences of crimes over different months of the year to analyse the trend of number of crimes which in-turn would help the law enforcement to take necessary actions before-hand.

Certain violent or more damage incurring types of crimes were shortlisted for the purpose of this study. They are: Burglary, Criminal damage and Arson, Other theft, Vehicle crime and Violence and Sexual offences.

Firstly to acquire appropriate data for this graph a “dplyr” function called “group_by()” was used to separate necessary data columns from the main data frame and then function “summarise()” was used to get the count of crimes occurring each month. These two functions were applied together using the piping method. Then, to subset the required types of crime “subset()” function from “raster” package was used. As the “Month” column in the dataset was not of “date” format “as.Date()” function was used to convert it. Finally a visualization function from “ggplot2” package called “ggplot()” was used to plot the line graph where the Y-axis was scaled to 3 months break.

Same procedure was followed for visualizing the crime trends of both the counties as shown in image below.

```r
#***********************CODE FOR TREND LINES*********************************************************


#Creating a new variable to plot trend lines
s_Y<-group_by(s_Yorkshire@data,Month,crime_type) %>% dplyr::summarise(count=n())

#Subsetting the top 5 crimes happeing in South Yorkshire over the years
s_Y<-subset(s_Y, s_Y$crime_type %in% c("Burglary","Criminal damage and arson",
                                       "Other theft","Violence and sexual offences","Vehicle crime"))

#changing the data type of Month column from double to data format
s_Y$Month<-as.Date(s_Y$Month)

#PLotting trend lines for different crimes from 2017-11 to 2020-03
ggplot(data = s_Y,aes(x=Month, y=count, color= crime_type))+
  geom_line(size=1)+
  scale_x_date(labels= date_format("%Y-%m"),breaks = date_breaks("3 months"))+ #scaling x-axis on gaps of 3 months
  scale_color_manual(values = c("violet", "orange","green","darkblue","red"))+ #custom colors for crime types
  labs(title = "Trends of different crimes over the years (South Yorkshire)",
       caption = "source:UK crime data", x="Date", y="Number of Crimes", color="Crime Types")
```

2.2.2 Choropleth Heat Map Visualisation

One of the most commonly used applications of Geospatial data is the Choropleth map. This is a type of thematic map in which we use color to reflect statistics of an attribute feature that we are interested in in proportion to its position, in this case, Number of crimes happening in a particular region. (Abdishakur, May 2020) Choropleth maps are used widely in conveying statistical values in different geographical scale, from global to local. They have some drawbacks, though. One specific downside to using choropleth maps is that regions are not standardized, and thus the results shown do not reflect the correct results. For example, large geographic areas might dominate the visual.

1: For Heat Map based on Number of Crimes:

For this graphical representation, firstly a new variable with the main data is created for better understanding and to reduce confusion. Then we use the "group_by()" and "summarise()" together to find the total number of crimes happening in a region based on "LSOA_code". Then the data from both variables were merged together using "join()" function from dplyr package. To overcome the problem of duplicate LSOA codes, "filter()" function embedded with "duplicated()" function was used which resulted in unique values for each region with their respective data. Finally, to plot this resultant data, visualization function "tm_shape()" from the "tmap" package was used. But before that the plotting screen was transformed into an interactive screen, which is done by the function "tmap_mode("view")".

Same procedure was followed for visualizing the Choropleth Heat Map for both the counties as shown in the image below.

```r
#***********************************CODE FOR HEAT MAP BASED ON NO OF CRIMES************************

#Creating extra variable for Heat Map
s_Y_heatMap<-s_Yorkshire

#counting the number of LSOA_code which inturn gives the number of crimes in that region
LSOA_code_count<-group_by(s_Y_heatMap@data,LSOA_code,) %>% dplyr::summarise(count=n())

#joining the two data
s_Y_heatMap@data<-join(s_Y_heatMap@data, LSOA_code_count,by=c('LSOA_code'='LSOA_code'))

#filtering all the duplicate LSOA_codes from the data_frame
s_Y_heatMap@data <- s_Y_heatMap@data %>% filter(duplicated(s_Y_heatMap@data["LSOA_code"])==FALSE)

#changing the mode of tmap to interactive
tmap_mode("view")

#plotting the crime count data
tm_shape(s_Y_heatMap)+tm_fill("count",palette=c("blue","white","red"),
                        breaks=c(0,100,200,500,4500,5500,6500,8000,9000),
                        title ="No of Crimes") +
  tm_layout(legend.outside = TRUE)+ tm_borders() +
  tm_layout(main.title = "Heat Map for South Yorkshire based (No of Crimes)", title.size = 1)

```

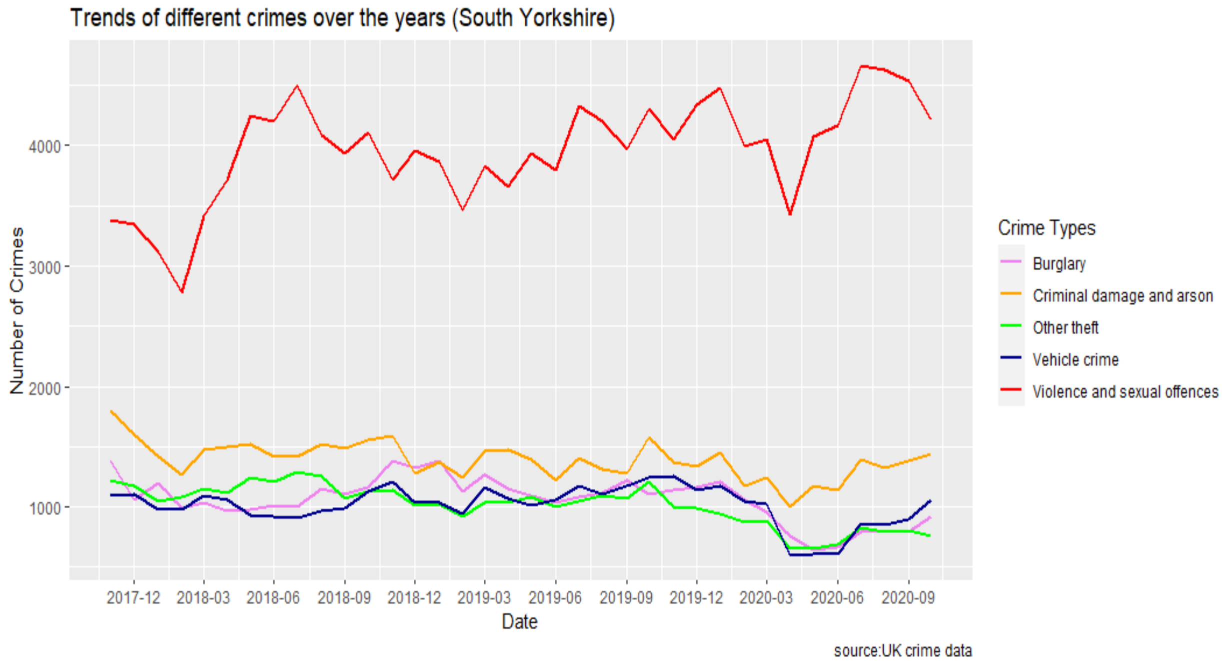
2. For Heat Map based on Major Crime Type:

For this visualisation, the aim was to find which crime occurred the most in a region over the period of three years that were considered for this study. Hence, to find the count, “group_by()” function was used to extract necessary columns and piped them to “summarise()” function. Which finally gave a count of each crime type in a region. To make use of the “count” data column later in the code and to avoid clash with the built-in R function “count()”, the column was renamed for better access. Then all the duplicate “LSOA_code” in the main file were removed. For the main step, to actually find which crime occurred the most in a particular area “setDT()” function embedded with “order()” and “head()” from “data.table” package were used. The “setDT()” function converts the “list” and “data.frame” to “data.table” by reference (rdrr.io, n.d.). The “order()” functions lists the values in the table in descending order of the “Count” and “head()” selects one unique “LSOA_code” from starting from top of the table which gives the crime type happening the most in a region. Then we join the two data files and then plot using the “tm_shape()” function.

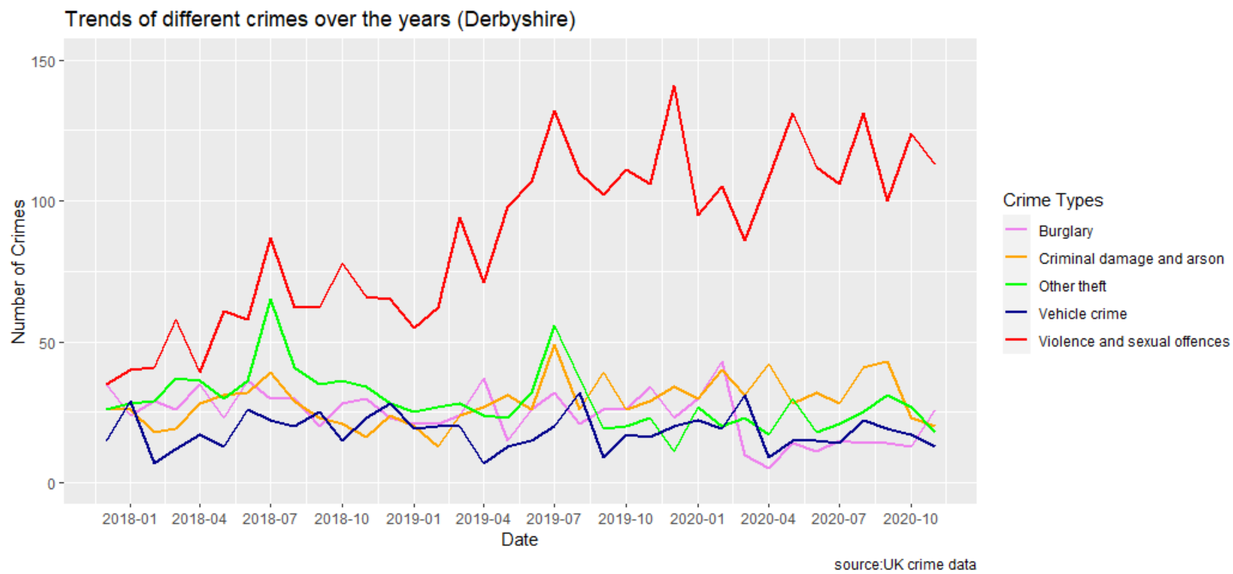
Same procedure was followed for visualizing the Choropleth Heat Map for both the counties as shown in the image below.

```r
#*******************************CODE FOR HEAT MAP BASED ON MOST HAPPENING CRIME IN A REGION*************

#Creating extra variable for Heat Map
s_Y_maxCrime<-s_Yorkshire

#to count the no of each type of crime happening in a region
crime_count<-group_by(s_Y_maxCrime@data,LSOA_code,crime_type) %>% dplyr::summarise(count=n())

#changing the name to avoid clash in-built function count()
crime_count<-rename(crime_count,c("count"="Count"))

#filtering duplicate LSOA_codes in the Shapefile data.frame
s_Y_maxCrime@data <- s_Y_maxCrime@data %>% filter(duplicated(s_Y_maxCrime@data["LSOA_code"])==FALSE)

#obtaning max count of a crime happening in a region
crime_count<-setDT(crime_count)[order(-Count, LSOA_code), head(.SD, 1), by = LSOA_code]


#joining both data.frames
s_Y_maxCrime@data<-join(s_Y_maxCrime@data, crime_count,by=c('LSOA_code'='LSOA_code'))


#interactive mode
tmap_mode("view")

#plotting data based on crime_type to get heatmap of which region has which crime the most
tm_shape(s_Y_maxCrime)+tm_fill("crime_type",title ="Types of Crimes") +
  tm_layout(legend.outside = TRUE)+ tm_borders() +
  tm_layout(main.title = "Heat Map for South Yorkshire (Crime Type)", title.size = 1)
```

3. Results



The above graph shows the trends of crime in the county of South Yorkshire from 09-2017 to 10-2020. The above graph depicts that there is a very high rate of crime related to "Violence and sexual offences" throughout the year as compared to other crimes. It can also be seen that every year during the months of winter there is possibility of increase in the crime rates of certain crimes like "Arson", "Vehicle crime" and "Violence and sexual offences". But a drop can be seen in the case of "Burglary". This is because most people prefer to stay indoors during these times. As it is also a festive environment during the winters it contributes to the increase in crime rate for certain crimes. Considering the recent pandemic situation, a sharp drop can be noticed at point at which it was begun, that is during the month of march in the year 2020. And as soon as the restrictions on the lockdown were eased around the month of June, a growth in the crime rate is easily visible. Hence, this proved to be a major contributing factor in the occurrences of a crime.



Similar to the trends in South Yorkshire, in Derbyshire, it can be seen that "Violence and sexual offences" is crime with the most occurrences. Also, same trend can be noticed in this graph too, for the time during the outbreak of the pandemic situation. The only difference between the two counties is that the spike in crimes is during the month of June to August in each year as compared to the colder months in South Yorkshire.

This type of trendline graph or line graph can be very helpful for the law enforcement departments to take prior action where ever possible to avoid public harm.

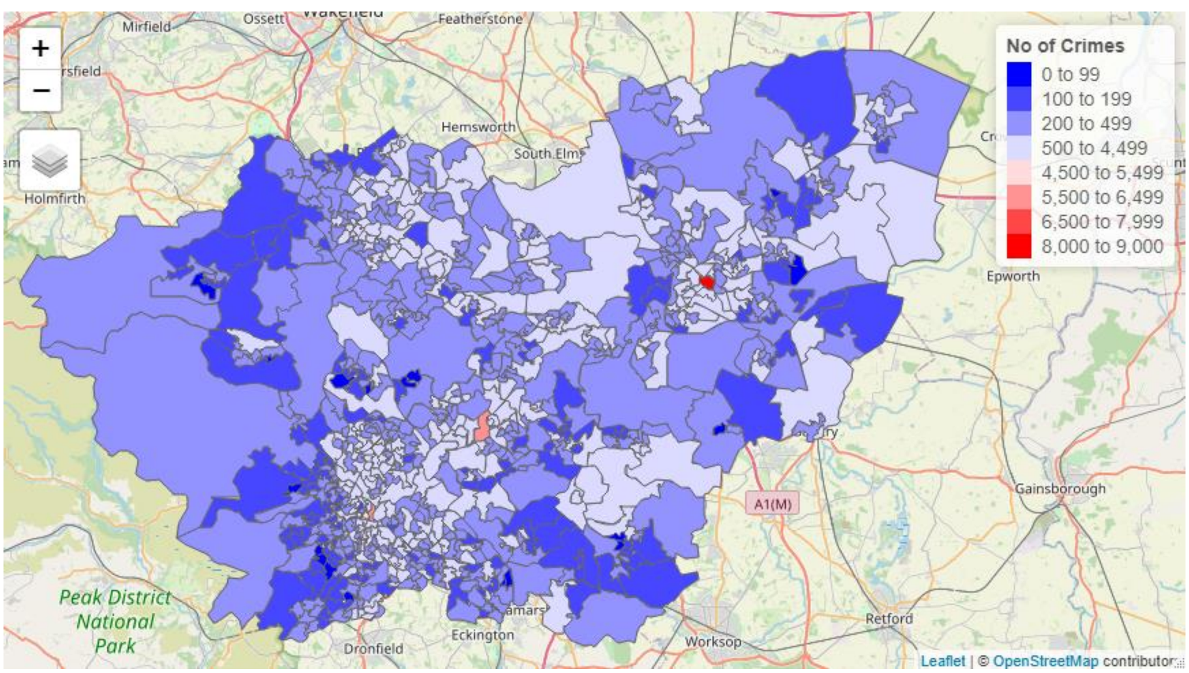


Figure 1:Crime count in South Yorkshire

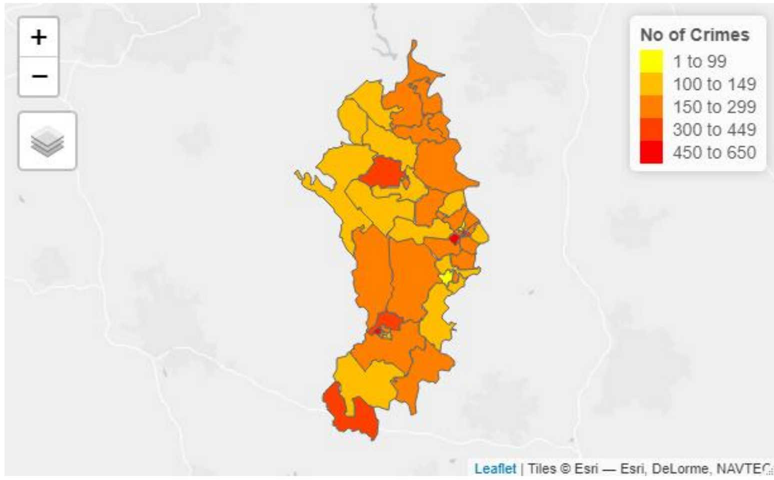


Figure 2: Crime count in Derbyshire

Figure 1 and 2 show the choropleth heat map of the number of crimes in each region of South Yorkshire and Derbyshire respectively. It can be seen that some region in both counties have very high crime count as compared to others. Which would help the police department to divert the focus on these areas more and save on unnecessary expenditure.

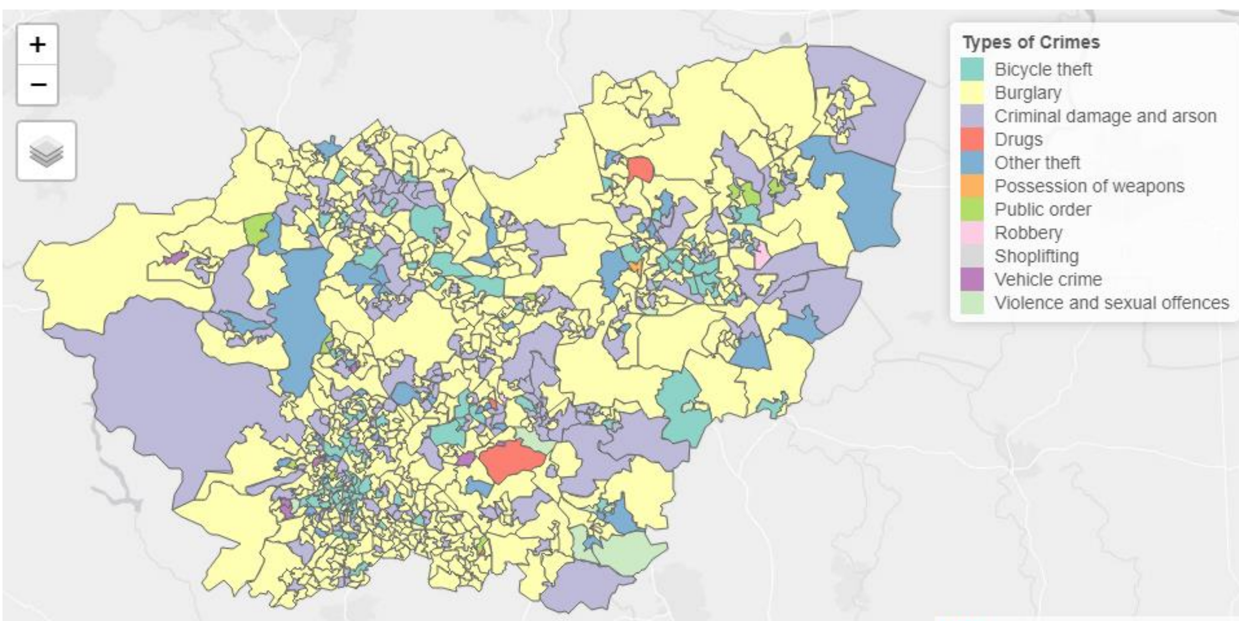


Figure 3: Major crime in a region (South Yorkshire)

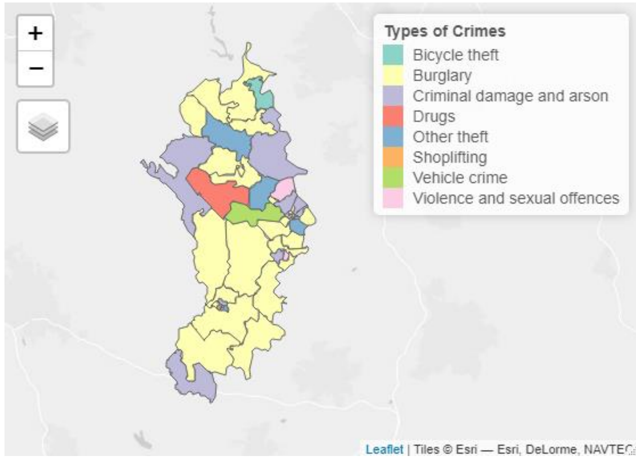


Figure 4: Major crime in a region (Derbyshire)

It can be seen from figures 3 & 4 that both counties have “Burglary” as the most common crime. This type of choropleth Heat Map can provide the police department with the data as to where and what type of resources should be used in order to reduce the crime rate in a region. Even though “Burglary” is the major crime in many region, “Violence and sexual offences” has the most number of cases. This is the drawback of the Choropleth Heat Maps. Hence both, the count of each crime as well as the major crime of a region is necessary to be analysed.

4. Conclusion

As concluded from this study and stated by (Salleh, et al., 2012), factors like demographics, economic instability, health, weather, environmental disasters or global unrest, play an important role in the type of a crime or its frequency. This method of analysing the past data provides a framework for visualizing and evaluating crime networks with Google Maps and different R packages using different approaches. By way of multiple immersive visualizations, it allows crime researchers to analyze these crime networks. It also is of great help to the law enforcement agencies for better use of resources to fight crime and to maintain peace and order.

APPENDIX

5. R Code

5.1 Code for South Yorkshire

```
library(tidyverse)
library(dplyr)
library(readxl)
library(tmap)
library(rgdal)
library(ggplot2)
library(mapdata)
library(raster)

#Obtaining shapefiles for different regions of South Yorkshire
sheffield <- readOGR(dsn="./BoundaryData", layer="england_lsoa_2011")
barnsley <- readOGR(dsn="./Barnsley", layer="england_lsoa_2011")
rotherham <- readOGR(dsn="./Rotherham", layer="england_lsoa_2011")
doncaster <- readOGR(dsn="./Doncaster", layer="england_lsoa_2011")

#Merging shapefiles to create a single shapefile
s_Yorkshire <- raster::union(sheffield,barnsley)
s_Yorkshire <- raster::union(s_Yorkshire,rotherham)
s_Yorkshire <- raster::union(s_Yorkshire,doncaster)

#arranging data from similar category together (Data pre-processing)
a <- s_Yorkshire[1:3]
names(a) <- c("Label","Region","LSOA_code")

b <- s_Yorkshire[4:6]
names(b) <- c("Label","Region","LSOA_code")

c <- s_Yorkshire[7:9]
names(c) <- c("Label","Region","LSOA_code")

#Merging Cleaned data together
s_Yorkshire <- rbind(a,b,c)

#Removing NA values from the acquired data
s_Yorkshire <- s_Yorkshire[!is.na(s_Yorkshire$Label),]


#Loading South Yorkshire crime data
crimeData <- read_xlsx("South Yorkshire Crime Data.xlsx", sheet=2)

#Changing Names of the columns
crimeData <- rename(crimeData,c("LSOA code"="LSOA_code",
                              "Crime type"="crime_type"))

#Removing NA values from the acquired data
crimeData <- crimeData[!is.na(crimeData$Longitude),]

#Joing crime data and shapefile data by LSOA_code
s_Yorkshire@data <- left_join(s_Yorkshire@data, crimeData,
                               by=c('LSOA_code'='LSOA_code'))
```

```r
#Removing unwanted columns
s_Yorkshire@data <- subset(s_Yorkshire@data, select = c(crime_type, LSOA_code,
Longitude, Latitude, Month))


#**********************************CODE FOR TREND LINES*******************************


#Creating a new variable to plot trend lines
s_Y <- group_by(s_Yorkshire@data,Month,crime_type) %>% dplyr::summarise(count=n())


#Subseting the top 5 crimes happeing in South Yorkshire over the years
s_Y <- subset(s_Y, s_Y$crime_type %in% c("Burglary", "Criminal damage and arson",
"Other theft", "Violence and sexual offences", "Vehicle crime"))

#changing the data type of Month column from double to data format
s_Y$Month <- as.Date(s_Y$Month)

#PLotting trend lines for different crimes from 2017-11 to 2020-03

ggplot(data = s_Y,aes(x=Month, y=count, color= crime_type))+ geom_line(size=1)+
  scale_x_date(labels= date_format("%Y-%m"),
               breaks = date_breaks("3 months"))+ #scaling x-axis on gaps of 3 months
  scale_color_manual(values = c("violet", "orange","green",
                                "darkblue","red"))+ #custom colors for crime types
  labs(title = "Trends of different crimes over the years (South Yorkshire)",
       caption = "source:UK crime data", x="Date", y="Number of Crimes", color="Crime
Types")




#**********************CODE FOR HEAT MAP BASED ON NO OF CRIMES**********************

#Creating extra variable for Heat Map
s_Y_heatMap <- s_Yorkshire

#counting the number of LSOA_code which inturn gives the number of crimes in that
region
LSOA_code_count <- group_by(s_Y_heatMap@data,LSOA_code,) %>%
dplyr::summarise(count=n())

#joining the two data
s_Y_heatMap@data <- join(s_Y_heatMap@data, LSOA_code_count, by =
c('LSOA_code'='LSOA_code'))

#filtering all the duplicate LSOA_codes from the data_frame
s_Y_heatMap@data <- s_Y_heatMap@data %>%
filter(duplicated(s_Y_heatMap@data["LSOA_code"])==FALSE)

#changing the mode of tmap to interactive
tmap_mode("view")

#plotting the crime count data
tm_shape(s_Y_heatMap)+
       tm_fill("count",palette=c("blue","white","red"),
                              breaks=c(0,100,200,500,4500,5500,6500,8000,9000),
                              title ="No of Crimes") +
       tm_layout(legend.outside = TRUE)+ tm_borders() +
       tm_layout(main.title = "Heat Map for South Yorkshire based (No of Crimes)",
                               title.size = 1)
```

```
#************CODE FOR HEAT MAP BASED ON MOST HAPPENING CRIME IN A REGION*************

#Creating extra variable for Heat Map
s_Y_maxCrime <- s_Yorkshire

#to count the no of each type of crime happening in a region
crime_count <- group_by(s_Y_maxCrime@data,LSOA_code,crime_type) %>%
dplyr::summarise(count=n())

#changing the name to avoid clash in-built function count()
crime_count <- rename(crime_count,c("count"="Count"))

#filtering duplicate LSOA_codes in the Shapefile data.frame
s_Y_maxCrime@data <- s_Y_maxCrime@data %>%
filter(duplicated(s_Y_maxCrime@data["LSOA_code"])==FALSE)

#obtaning max count of a crime happening in a region
crime_count <- setDT(crime_count)[order(-Count, LSOA_code), head(.SD, 1), by =
LSOA_code]


#joining both data.frames
s_Y_maxCrime@data <- join(s_Y_maxCrime@data,crime_count,by=c('LSOA_code'='LSOA_code'))


#interactive mode
tmap_mode("view")

#plotting data based on crime_type to get heatmap of which region has which crime the
most
tm_shape(s_Y_maxCrime)+tm_fill("crime_type",title ="Types of Crimes") +
  tm_layout(legend.outside = TRUE)+ tm_borders() +
  tm_layout(main.title = "Heat Map for South Yorkshire (Crime Type)", title.size = 1)

####################################################################################
```

5.2 Code For Derbyshire

```
library(tidyverse)
library(dplyr)
library(readxl)
library(rgdal)
library(ggplot2)
library(tmap)
library(raster)


#Reading the shapefile for derbyshire
derbyshire <- readOGR(dsn="./Derbyshire Shapefile", layer="england_lsoa_2011")

#Reading the crime data for derbyshire
derbyData <- read_xlsx("Derbyshire Crime Data.xlsx")
```

```r
#Renaming columns for easier access

derbyData <- rename(derbyData, c("LSOA code"="LSOA_code",
                                 "Crime type"="crime_type"))

#Removing rows containing NA values
derbyData <- derbyData[!is.na(derbyData$Longitude),]



#Joining shapefile and crime data based on LSOA code
derbyshire@data <- left_join(derbyshire@data, derbyData,
                             by=c('code'='LSOA_code'))

#Removing unwanted columns
derbyshire@data <- subset(derbyshire@data, select =
c(crime_type,LSOA_code,Longitude,Latitude,Month))



#*****************************CODE FOR TREND LINES*******************************


#Creating a new variable to plot trend lines
derby_trend <- group_by(derbyshire@data,Month,crime_type) %>%
dplyr::summarise(count=n())

#Subsetting the top 5 crimes happeing in South Yorkshire over the years
derby_trend <- subset(derby_trend, derby_trend$crime_type %in% c("Burglary","Criminal
damage and arson", "Other theft", "Violence and sexual offences", "Vehicle crime"))

#changing the data type of Month column from double to data format
derby_trend$Month <- as.Date(derby_trend$Month)

#PLotting trend lines for different crimes from 2017-11 to 2020-03
ggplot(data = derby_trend,aes(x=Month, y=count, color= crime_type))+
  geom_line(size=1)+ ylim(0,150)+
  scale_x_date(labels= date_format("%Y-%m"),
               breaks = date_breaks("3 months"))+ #scaling x-axis on gaps of 2 months
  scale_color_manual(values = c("violet", "orange","green",
                                "darkblue","red"))+ #custom colors for crime types
  labs(title = "Trends of different crimes over the years (Derbyshire)",
       caption = "source:UK crime data", x="Date", y="Number of Crimes",
       color="Crime Types")


#*****************CODE FOR HEAT MAP BASED ON NO OF CRIMES**********************

#Creating extra variable for Heat Map
derby_heatMap <- derbyshire

#counting the number of LSOA_code which inturn gives the number of crimes in that
region
LSOA_count_derby <- group_by(derby_heatMap@data,code,) %>% dplyr::summarise(count=n())

#joining the two data
derby_heatMap@data <- join(derby_heatMap@data, LSOA_count_derby,by=c('code'='code'))

#filtering all the duplicate LSOA_codes from the data_frame
derby_heatMap@data <- derby_heatMap@data %>%
filter(duplicated(derby_heatMap@data["code"])==FALSE)
```

```r
#changing the mode of tmap to interactive
tmap_mode("view")

#plotting the crime count data
tm_shape(derby_heatMap)+tm_fill("count",palette=c("yellow","red"),
                              breaks=c(1,100,150,300,450,650),
                              title ="No of Crimes") +
  tm_layout(legend.outside = TRUE)+ tm_borders() +
  tm_layout(main.title = "Heat Map for South Yorkshire", title.size = 1)



#*************CODE FOR HEAT MAP BASED ON MAX NO OF A CRIME IN A REGION*************

#Creating extra variable for Heat Map
derby_maxCrime <- derbyshire

#to count the no of each type of crime happening in a region
crime_count_derby <- group_by(derby_maxCrime@data,code,crime_type) %>%
dplyr::summarise(count=n())

#changing the name to avoid clash in-built function count()
crime_count_derby <- rename(crime_count_derby,c("count"="Count"))

#filtering duplicate LSOA_codes in the data.frame
derby_maxCrime@data <- derby_maxCrime@data %>%
filter(duplicated(derby_maxCrime@data["code"])==FALSE)

#obtaning max count of a crime happeing in aa region
crime_count_derby <- setDT(crime_count_derby)[order(-Count,code), head(.SD, 1), by =
code]


#joining both data.frames
derby_maxCrime@data <- join(derby_maxCrime@data,crime_count_derby,by=c('code'='code'))


#interactive mode
tmap_mode("view")

#plotting data based on crime_type to get heatmap of which region has which crime the
most
tm_shape(derby_maxCrime)+tm_fill("crime_type",title ="Types of Crimes") +
  tm_layout(legend.outside = TRUE)+ tm_borders() +
  tm_layout(main.title = "Heat Map for South Yorkshire", title.size = 1)

#********************************************************************************
```

6.References

Abdishakur (May 2020) The 7 Best Thematic Map Types for Geospatial Data
Retrieved January 10, 2021, from
https://towardsdatascience.com/the-7-best-thematic-map-types-for-geospatial-data-34d21b8b475

Bachner, J. (2013). *Predictive policing: preventing crime with data and analytics*. IBM Center for the Business of Government.

Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: a general framework and some examples. *computer*, *37*(4), 50-56.

Line Graph: The Data Visualisation Catalogue Retrieved January 10, 2021, from
https://datavizcatalogue.com/methods/line_graph.html

rdrr.io(n.d.) setDT: Convert lists and data.frames to data.table by reference
Retrieved January 07, 2021, from
https://rdrr.io/rforge/data.table/man/setDT.html

Salleh, S. A., Mansor, N. S., Yusoff, Z., & Nasir, R. A. (2012). The crime ecology: Ambient temperature vs. spatial setting of crime (burglary). Procedia-Social and Behavioral Sciences, 42, 212- 222.

Sarah Marsh (2019) "UK police use of computer programs to predict crime sparks discrimination warning" Retrieved January 10, 2021, from
https://www.theguardian.com/uk-news/2019/feb/03/police-risk-racial-profiling-by-using-data-to-predict-reoffenders-report-warns

Telea, A. C. (2014). *Data visualization: principles and practice*. CRC Press.

Tim Lau (2020) "Predictive Policing Explained" Retrieved January 10, 2021, from
https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained

Wang, D., Ding, W., Lo, H., Stepinski, T., Salazar, J., & Morabito, M. (2013). Crime hotspot mapping using the crime related factors—a spatial data mining approach. *Applied intelligence*, *39*(4), 772-781.