Spatial Visualisation and Clustering Algorithm to Depict Crime in South Australia

Dhruvisha Gosai (JC810547)
MA5800 - Foundations of Data science, James Cook University

Abstract

As with all other fields, data science has the capacity to lead in improvements in technologies and policies in law enforcement. Areas of opportunity in this space would result in less damage, less assault, and potentially saving lives. The purpose of this report is to analyse the hotspots of reported crimes in South Australia and to identify trends in seasonality and type of crime for different regions. As evidenced by the consequent results, there has been a significant increase in crime recorded in South Australia leading into December 2019, compared to prior months that year as well as all other prior Decembers of the previous decade.

Geospatial visualisation used for crime mapping, provides a fresh insight into understanding how crime affects a state. This applied alongside a k-means clustering results in two unique clusters which divide the neighbourhoods into two groups where one of the clusters depicts that neighbourhood that have high volume of theft, fraud, homicide and other criminal activities tend to have low volume of sexual assaults and other related offences.

Applying the visualisation and data science algorithms approaches adopted in this investigation, lends comprehension into changes in the type and frequency of crimes for South Australia – and can be readily applied to all other states. The groundwork of this investigation may assist in future attempts at predicting crime attributes using sophisticated distance-decay models.

Introduction

Understanding the influencing factors that depict crime rates assists in policy makers and law enforcement making appropriate changes to curb rates of offences and reducing the overall damage and response times to these incidents.

Crime attractors are neighbourhoods and districts which create opportunities to criminals offenders - examples might include bar districts; prostitution areas; drug markets; large shopping malls, particularly those near major public transit exchanges; large, insecure parking lots in business or commercial areas.

The objectives of this report are to identify trends in location of crimes committed – whether these are characterised by the type of offence (robbery, sexual assault etc), or whether the time of year is a superior predicator in expecting a rise in offence counts. Knowing that a particular postcode has consistently seen crime rates higher than surrounding areas can assist in influencing decisions such as patrol frequency, and police workforce hire demand. Trends that are identified have the capacity to form the basis of real-world actionable insights.

With recent development in data tools and an abundance of accessibility of data online, geospatial analysis was possible which presents a unique perspective when examining the patterns, occurrences and/or processes that make use of the geographic information which links elements and phenomena surface to their locations – visually and informationally accessible (Medina & Solymosi, 2019).

Data

For this project, crime statistics data for financial years of 2010 to 2020 were collected from “Data SA” - published by South Australian Police. These were observational datasets detailing the crimes against a person or property reported to the police during this period (Police, 2020). A total of 10 csv files were aggregated in R to create a single dataset with 11 variables and 845,739 observations with 4 derived variables, used for further analysis. A list of used variables given below –

Field Name	Derived	Description	Format
reported_Date		Date of reported crime	Date
incident_suburb		Suburb of crime incident	Character
incident_postcode		Postcode of crime incident	Numeric
offence_desc_level1		Most summarised segmentation for committed offence	Character
offence_desc_level2		Summarised segmentation with additional details for committed offence	Character
offence_desc_level3		Detailed segmentation for committed offences	Character
offence_count		Number of criminal offences	Integer
Year	Derived	Year derived from reported date	Numeric
Month	Derived	Month derived from reported date	Character
Weekday	Derived	Weekday of the reported date	Character
month_year	Derived	Reported date rolled up to month and year level	Character

In addition to crime data, a Suburb/Locality Boundaries, and their legal identifiers dataset was also collected for the purpose of conducting spatial analysis, in the format of a shape file (.shp) (Department of Industry, 2014). This data was derived from cadastre data made available by data.gov.au.
Criminal data was cleaned to remove the rows with missing values for the reported date and incident postcodes – resulting in 566 dropped observations. Time frame for the data was updated to range between 01-Jan-2011 and 31-Dec-2020, removing the year 2010.

Methods

All analysis and data science procedures were performed in R studio version 1.3.959 with a list of libraries referenced in Appendix 1 (RStudio, 2020). Data was imported into R using the list.files() function with a specified file type of .csv. This approach allows multiple files to be imported from a single location with common format. With crime results from 01-Jan-2011 to 31-Dec-2019 of interest, dplyr package is used to filter for date range and to remove any missing dates or postcodes on the records. New variables were created using the year(), month() and, wday() function on *reported date* to illustrate the trends seen over a period of time.
To conduct spatial analysis, a shape file is required consisting of geometric location data and attribute information of geographic features represented by points, lines or areas (ArcMap, 2016). For this report, this information was obtained from data.gov.au and loaded into R using shapefile() function from raster package. It is important to note that raster package needs to be detached once used, to avoid issues with dplyr package as both the packages have select() function. Once imported, data would not be in the form of a data frame or a matrix, it instead requires the function fortify() to convert the shape object into a computable data frame. Applying fortify function allows the geometric objects to be readable in the form of geographic coordinates - latitude and longitude.
Data exploration and visualisation is key to identifying trends of interest and to use as a basis in creating an algorithm crucial to predict crime. There are 4 different types of visualisations implemented during this research to understand the effects of each variable over a period of time.
Figures 1 to 4 were created using the ggplot2 package and function. Because of the high number of observations, data was first summarised based on x and y variables of interest from crime data and then plotted using functions like geom_line() for time series (Fig. 1), geom_point() for data points on time series and crime spots on geo-spatial map (Fig. 4), geom_bar() for crime indicators' bar graph (Fig. 2), geom_tile() for crime frequency by month and year (Fig. 3) and geom_polygon() to create the geo-spatial map using the coordinates from shape file.
From data exploration, it was apparent that the year 2019 saw a rise in the number of crimes. As this was the most recent full year, data was further subset to only use 2019 observations moving forward.
To depict the type of crimes in a particular neighbourhood, K-Means clustering is used. It is one of the more commonly used clustering techniques and operates by splitting the data points into small but meaningful clusters (Piech, 2013). Using the crime data to measure the number of thefts, property damages, trespassing, assaults, and other criminal markers will be grouped together for similar

neighbourhoods. To conduct the cluster analysis, summarisation was done for counting the number of offences and then grouping by *Suburbs* and *Offence Description 2*, followed by transposing *Offence Description 2*. Subsequently, *NA* values were identified using is.na() and then converted to 0 before being rescaled for comparison. It was apparent from the clustering plot (Fig. 5) that there were only two clusters to fit in a model.

Results

The results from this investigation were met with a small degree of success but overall, some characteristics of significance emerged.

1. Time series of offences: Plotting the time series of crime data for over a period of 9 years from 2011 to 2019. With key observations as stated below –
    - Decrease of -8.6% (10,084 offences) over the period of 2011-12,
    - Increase of 5.0% (5,343 offences) between 2018-19.
    - There was a notable increase in the number of offences from approximately June-2018, which has continued to steadily rise through 2019 until today.

It can be observed that there was a dramatic drop in reported offences for the 2018 period, with a clean cut-off point for Jan2018 and Dec2018 before the volumes resumed to their expected trend.

This suggests that there is a potential data anomaly coming through in the data source which was unable to be verified from the SA government website.

For the period of 2012-2014 a gradual but noticeable drop was observed in reported crime, seeing an almost 30% reduction in reported offences in June2014 compared to June of 2012. This was been attributed to a new police commissioner during this period adopting changes in policies and procedures.

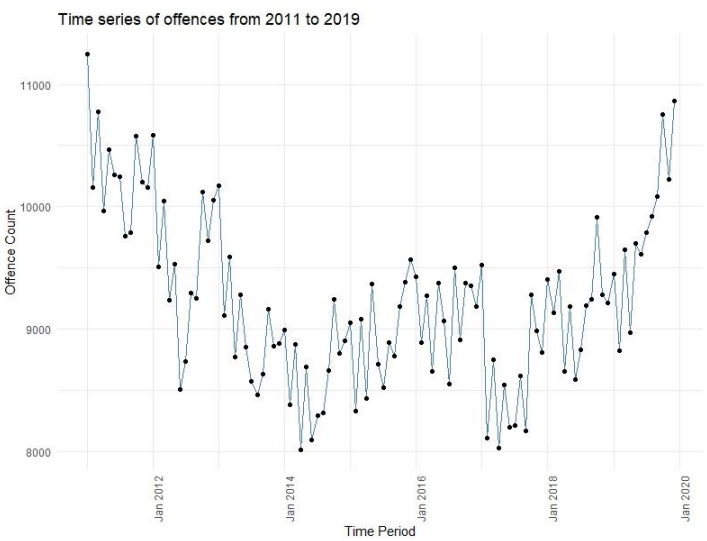

Figure 1: Time series of all offence types in last 9 years

As 2019 data is the most recent full year it was selected for further analysis, particularly as there has been a steady increase in reported crime leading into 2020. This is especially apparent for December 2019 which has seen a stark increase in crime compared to prior months of 2019 and indeed all other Decembers preceding that one – as visible in the heat map below.

2. Major crime indicators: A bar graph with *Offence Description 2* is an apt indication of what sort of crimes are most prevalent for the rise of the volume of offences in 2019. Seemingly, Theft comprised of 46% of the total offences committed in 2019.

3. Crime frequency by month and year tile: With theft being the more prominent type of crime, making up almost half of all the crimes committed in 2019, it was of interest

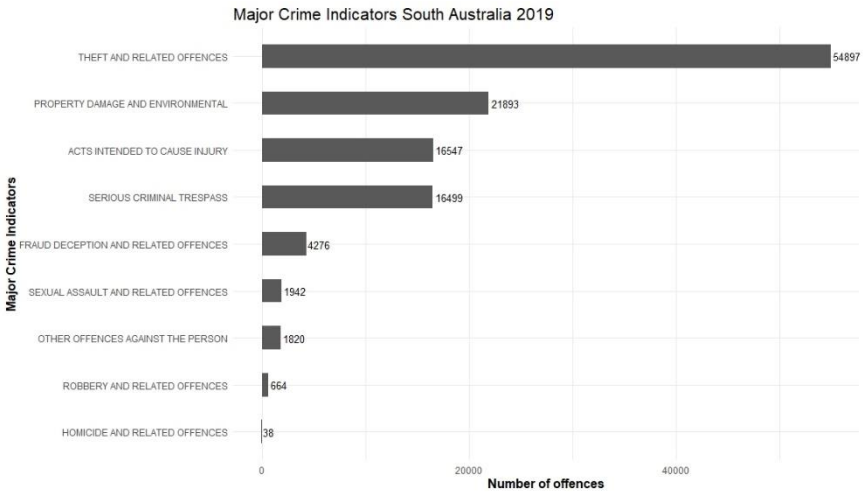

Figure 2: Crime indicators for 2019

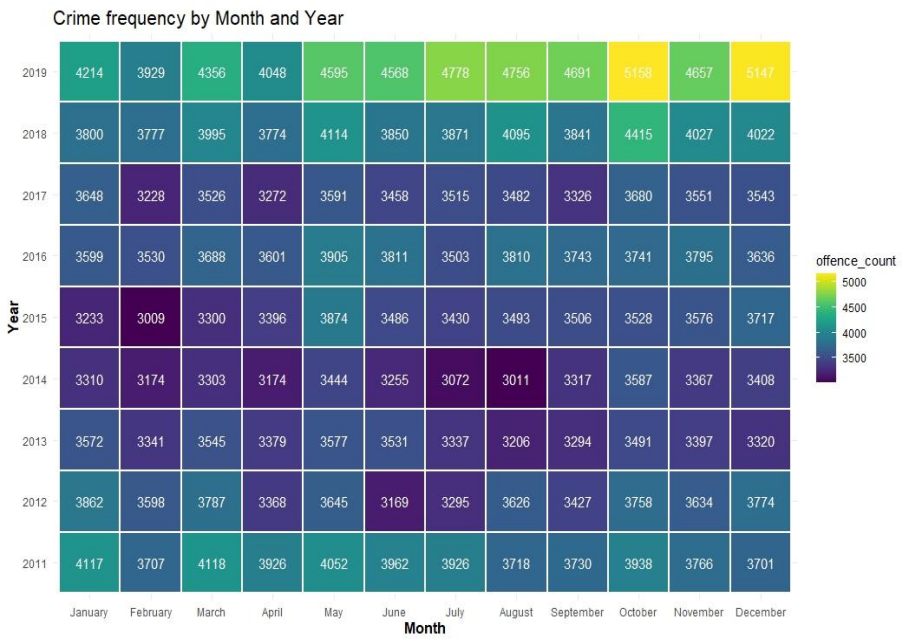

Figure 3: Theft offences reported in last 9 years, split by each month

to see understand the spread of theft by months over a period of 9 years from 2011 to 2019. As noted from the findings, theft in total looks significantly higher towards the second half of 2019 with offences reaching an all-time high of 5,000+ reported incidents in the months of October and December.

4. Geospatial map: This was generated to understand the major crime spots of SA. With the high-density population centre of metropolitan Adelaide, it is correlated to the high density of the crimes recorded.

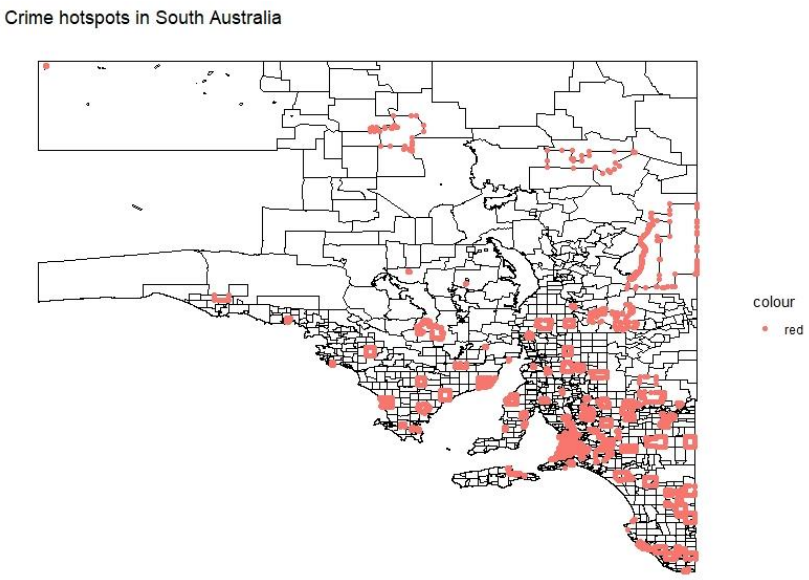

Figure 4: Crime hotspots in South Australia in 2019

5. K-Means Clustering: This was used to identify the relationship between the neighbourhoods or suburbs and the type of crimes committed. From the results of the clustering model (appendix 7), it was evident that there were 2 maximum possible clusters with one cluster comprising of 63 neighbourhoods or suburbs and second cluster comprising of 1,539 neighbourhoods. First cluster seemed to have low offences recorded for fraud, theft, property damage, criminal trespassing, and homicide. But a higher count recorded for sexual assaults and vice versa for cluster 2 where sexual assaults were lower and the other crimes were higher. Plotting the results of clustering, it appears that the choice of number of clusters is not too good as there is more noise and likely to have more accurate results with higher number of clusters. However despite not being statistically significant, these two clusters explain 76.75% of the point variability, indicating that the two segmentations are mildly appropriate.

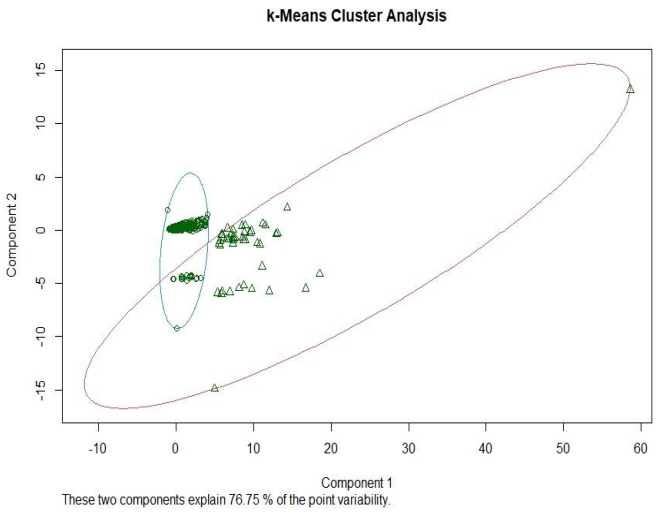

Figure 5: K-Means cluster plotting

Conclusion

As observed from these results, data mining in crimes for local, state, and national governments has the potential to yield returns in future data science and modelling endeavours. Attempts to anticipate crime waves and opportunities in improving response rates in local police stations serve as achievable goals possible from understanding the relevant drivers. This analysis intends to highlight these influencing factors so that further research in this field may benefit. Using geospatial information in real-time alongside of distance-decay model could assist law enforcement to prevent the delay in solving a crime and fast-track the allocation and dispatch of police force at the required location.

Further analysis would be beneficial in observing the continued crime rates to see if there are deviations from current trends leading into 2021, and to what extent COVID19 has impacted these – if at all.

It is acknowledged that as evidenced by the sudden drop crime volume in 2018 data that there may be some data integrity issues within the sourced data, leading to some doubts over the validity of volumes raised in prior and subsequent years. Of additional note is that as laws change over time where new ones are implemented and old ones are removed, this may lead to artificial disparity in rises and troughs over time.

References

ArcMap. (2016). *Environmental Systems Research Institute, Inc.* Retrieved from desktop.arcgis.com: https://desktop.arcgis.com/en/arcmap/10.3/manage-data/shapefiles/what-is-a-shapefile.htm

Department of Industry, S. E. (2014, 09 09). *SA Suburb/Locality Boundaries - Geoscape Administrative Boundaries*. Retrieved from data.gov.au: https://data.gov.au/dataset/ds-dga-bcfcfc9a-7c8d-479a-9bdf-b95ca66ad29a/details?q=

Medina , J., & Solymosi, R. (2019, 04 02). *Crime Mapping in R*. Retrieved from github: https://maczokni.github.io/crimemapping_textbook_bookdown/

Piech, C. (2013). *Stanford CS221*. Retrieved from stanford.edu: https://stanford.edu/~cpiech/cs221/handouts/kmeans.html

Police, S. A. (2020, 09 15). *Crime Statistics - Datasets*. Retrieved from data.sa.gov.au: https://data.sa.gov.au/data/dataset/crime-statistics

RStudio. (2020). *RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA*. Retrieved from rstudio.com: http://www.rstudio.com/

Appendix

1. R studio libraries used:

```
library(plyr)
library(readr)
library(MASS)
library(ggplot2)
library(tidyr)
library(data.table)
```

```
library(crosstalk)
library(plotly)
library(shiny)
library(leaflet)
library(zoo)
library(tidyverse)
library(tidyr)
library(hrbrthemes)
library(ggthemes)
library(lubridate)
library(reshape2)
library(ggmap)
library(maps)
library(viridis)
library(cluster)
library(highcharter)
library(UsingR)
library(cowplot)
library(dplyr)
library(raster)*
```

*Note: Raster is used to import .shp file and need to be detached once used or else it will throw error when running select() in pipe statement.

2. Import datasets:

```
crime = "C:/Users/dhru/Desktop/JCU - Master of Data science/MA5800 - Foundations of Data
science/Assignment 5 - Capstone/Import Files/Crimes/"

import_crime <- list.files(path=crime, pattern="*.csv", full.names=TRUE)
import_crime

crime_csv <- ldply(import_crime, read_csv)

library(raster)
# Import shape file
shp <- shapefile("C:/Users/dhru/Desktop/JCU - Master of Data science/MA5800 - Foundations of Data
science/Assignment 5 - Capstone/Import Files/SA_LOCALITY_POLYGON_SHP-GDA2020.shp")
head(shp)
detach("package:raster", unload = TRUE)
```

2. Clean data:

```
#Give headings to the column for CRIME DATA
names(crime_csv) <- c("reported_date", "incident_suburb", "incident_postcode", "offence_desc_level1"
                      ,"offence_desc_level2", "offence_desc_level3","offence_count")

#Give columns a data type - factor, numeric, date for CRIME DATA
crime_csv$reported_date       <- as.Date(crime_csv$reported_date,'%d/%m/%Y')
crime_csv$incident_suburb     <- as.character(crime_csv$incident_suburb)
crime_csv$incident_postcode   <- as.character(crime_csv$incident_postcode)
crime_csv$offence_desc_level1 <- as.character(crime_csv$offence_desc_level1)
crime_csv$offence_desc_level2 <- as.character(crime_csv$offence_desc_level2)
crime_csv$offence_desc_level3 <- as.character(crime_csv$offence_desc_level3)
crime_csv$offence_count       <- as.numeric(crime_csv$offence_count)

#Create a new variable for year and month of crime
crime_csv$year                <- year(crime_csv$reported_date)
crime_csv$month               <- month(crime_csv$reported_date, label=TRUE, abbr=FALSE)
crime_csv$weekday             <- wday(crime_csv$reported_date, label=TRUE, abbr=FALSE)
crime_csv$weekday_num         <- wday(crime_csv$reported_date)
crime_csv$month_year          <- as.yearmon(crime_csv$reported_date, "%Y %m")

#removing missing records and only keeping data from 01-01-2011 to 31-12-2019 for reliable number of
occurances
crime_csv <- crime_csv %>%
    filter(!is.na(year)
           ,!is.na(incident_postcode)
           ,"2011-01-01" <= reported_date
           ,reported_date < "2020-01-01")
```

3. Data Visualisation:

```
#--------Time series - Total ----------#
crime_timeSeries_total <- crime_csv %>%
    group_by(month_year) %>%
    summarise(offence_count = sum(offence_count)) %>%
    select(month_year,offence_count)

ggplot(data=crime_timeSeries_total, aes(x=month_year, y=offence_count)) +
    geom_line( color="steelblue") +
    geom_point() +
```

```
    xlab("Time Period") +
    ylab("Offence Count") +
    ggtitle("Time series of offences over a period of 10 years") +
    theme_minimal() +
    theme(axis.text.x=element_text(angle=90, hjust=1))

#--------bar plot 2019----------#
crime_barplot <- crime_csv %>%
    filter(year == "2019") %>%
    group_by(offence_desc_level2) %>%
    summarise(offence_count = sum(offence_count)) %>%
    select(offence_desc_level2,offence_count)

ggplot(data=crime_barplot, aes(x = reorder(offence_desc_level2, offence_count), y = offence_count)) +
    geom_bar(stat = 'identity', width = 0.5) +
    geom_text(aes(label = offence_count), stat = 'identity', data = crime_barplot, hjust = -0.1, size
= 3.5) +
    coord_flip() +
    xlab('Major Crime Indicators') +
    ylab('Number of offences') +
    ggtitle('Major Crime Indicators South Australia 2019') +
    theme_minimal() +
    theme(plot.title = element_text(size = 16),
          axis.title = element_text(size = 12, face = "bold"))

#-------- visualisation tiles month by year ----------#
crime_titles <- crime_csv %>%
    filter(offence_desc_level2 =="THEFT AND RELATED OFFENCES") %>%
    group_by(month,year) %>%
    summarise(offence_count = sum(offence_count)) %>%
    select(month,year,offence_count)

ggplot(crime_titles, aes(month,as.factor(year) , fill = offence_count)) +
    geom_tile(size = 1, color = "white") +
    scale_fill_viridis()  +
    geom_text(aes(label=offence_count), color='white') +
    ggtitle("Crime frequency by Month and Year") +
    xlab('Month') +
    ylab('Year') +
    theme_minimal()+
    theme(plot.title = element_text(size = 16),
          axis.title = element_text(size = 12, face = "bold"))

#-------- Time series - Theft----------#
crime_timeSeries1 <- crime_csv %>%
    filter(year == "2019", offence_desc_level2 == "THEFT AND RELATED OFFENCES") %>%
    group_by(month_year) %>%
    summarise(offence_count = sum(offence_count)) %>%
    select(month_year,offence_count)

timeSeries1<- ggplot(data=crime_timeSeries1, aes(x=month_year, y=offence_count)) +
    geom_line( color="steelblue") +
    geom_point() +
    xlab("Time Period") +
    ylab("Offence Count") +
    ggtitle("Theft") +
    theme_minimal() +
    theme(axis.text.x=element_text(angle=90, hjust=1))

#--------Time series - Property damage ----------#
crime_timeSeries2 <- crime_csv %>%
    filter(year == "2019", offence_desc_level2 == "PROPERTY DAMAGE AND ENVIRONMENTAL") %>%
    group_by(month_year) %>%
    summarise(offence_count = sum(offence_count)) %>%
    select(month_year,offence_count)

timeSeries2 <- ggplot(data=crime_timeSeries2, aes(x=month_year, y=offence_count)) +
    geom_line( color="steelblue") +
    geom_point() +
    xlab("Time Period") +
    ylab("Offence Count") +
    ggtitle("Property Damage") +
    theme_minimal() +
    theme(axis.text.x=element_text(angle=90, hjust=1))

#--------Time series - Acts intended to cause injury ----------#
crime_timeSeries3 <- crime_csv %>%
    filter(year == "2019", offence_desc_level2 == "ACTS INTENDED TO CAUSE INJURY") %>%
    group_by(month_year) %>%
```

```
    summarise(offence_count = sum(offence_count)) %>%
    select(month_year,offence_count)

timeSeries3 <- ggplot(data=crime_timeSeries3, aes(x=month_year, y=offence_count)) +
    geom_line( color="steelblue") +
    geom_point() +
    xlab("Time Period") +
    ylab("Offence Count") +
    ggtitle("Acts intended to cause injury") +
    theme_minimal() +
    theme(axis.text.x=element_text(angle=90, hjust=1))

#--------Time series - serious criminal trespassing ----------#
crime_timeSeries4 <- crime_csv %>%
    filter(year == "2019", offence_desc_level2 == "SERIOUS CRIMINAL TRESPASS") %>%
    group_by(month_year) %>%
    summarise(offence_count = sum(offence_count)) %>%
    select(month_year,offence_count)

timeSeries4 <- ggplot(data=crime_timeSeries4, aes(x=month_year, y=offence_count)) +
    geom_line( color="steelblue") +
    geom_point() +
    xlab("Time Period") +
    ylab("Offence Count") +
    ggtitle("Serious criminal trespassing") +
    theme_minimal() +
    theme(axis.text.x=element_text(angle=90, hjust=1))

p <- plot_grid(timeSeries1, timeSeries2, timeSeries3, timeSeries4, labels = "AUTO")
timeSeries_title <- ggdraw() + draw_label("Time series of incidents in 2019", fontface='bold')
plot_grid(timeSeries_title, p, ncol=1, rel_heights=c(0.1, 1)) # rel_heights values control title
margins
```

4. Data Sub-Setting

```
#filtering to keep only 2019 and Theft for mapping
crime_csv_x1 <- crime_csv %>%
    filter(year =="2019"
           ,offence_desc_level2 == "THEFT AND RELATED OFFENCES") %>%
    group_by(incident_suburb, incident_postcode) %>%
    summarise(offence_count = sum(offence_count)) %>%
    select(incident_suburb,incident_postcode, offence_count)
```

5. Spatial Analysis

```
# Join shape file to crime data
summary(shp@data)
shp@data <- left_join(shp@data, crime_csv_x1, by = c("NAME" = "incident_suburb"))
head(shp@data, 15)

#passes the spatial data as a data.frame rather than a spatial object
shp@data$id <- rownames(shp@data)
south_aus_shp <- fortify(shp)
south_aus_shp     <- join(south_aus_shp,shp@data, by="id")
class(south_aus_shp)

crime_shp <- south_aus_shp %>%
    filter(offence_count > 100) %>%
    select(lat,long,offence_count) %>%
    group_by(lat,long)

ggplot()+
    geom_polygon(data = south_aus_shp,
                 aes(x = long, y = lat, group = group),
                 fill="white", colour = "black")+
    geom_point(data = crime_shp,
               aes(x = long, y = lat, color = "red"))+
    ggtitle("Crime hotspots in South Australia") +
    xlab("") +
    ylab("") +
    theme_bw()+
    theme(axis.line=element_blank(),
          axis.text.x=element_blank(),
          axis.text.y=element_blank(),
          axis.ticks=element_blank(),
          panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          panel.border = element_blank(),
          panel.background = element_blank())
```

6. K-Means Clustering

```
crime_cluster <- crime_csv %>%
                    filter(year =="2019") %>%
                    group_by(incident_suburb, offence_desc_level2) %>%
                    summarise(offence_count = sum(offence_count))

crime_cluster_x1 <- crime_cluster %>%
                   pivot_wider(names_from = "offence_desc_level2", values_from = "offence_count")

#qualitative data removed from the analysis
crime_cluster_x2 <- crime_cluster_x1[, -c(1,1)]

#remove missing values
crime_cluster_x2[is.na(crime_cluster_x2)] <- 0

#scaling of data
mean <- apply(crime_cluster_x2, 2, mean)
sd <- apply(crime_cluster_x2, 2, sd)
crime_cluster_x3 <- scale(crime_cluster_x2, mean, sd)

#number of clusters
Cluster_count <- (nrow(crime_cluster_x3)-1) * sum(apply(crime_cluster_x3, 2, var))
for (i in 2:20)
    Cluster_count[i] <- sum(kmeans(crime_cluster_x3, centers=i)$withiness)

plot(1:20, Cluster_count, type='b',
     xlab='Number of Clusters',
     ylab='Within groups sum of squares')
title("Number of Clusters for K-Means")

#fitting the model
k_cluster <- kmeans(crime_cluster_x3, 2)
k_cluster

crime_cluster_x4 <- data.frame(crime_cluster_x3, k_cluster$cluster)
clusplot(crime_cluster_x4, k_cluster$cluster, color=TRUE, shade=F, labels=0,
lines=0, main='k-Means Cluster Analysis')
```

7. K-Means Cluster Output

```
> k_cluster
K-means clustering with 2 clusters of sizes 1211, 44

Cluster means:
  ACTS INTENDED TO CAUSE INJURY FRAUD DECEPTION AND RELATED OFFENCES OTHER OFFENCES AGAINST THE PERSON
1                    -0.1172099                              -0.09573948                            -0.1248004
2                     3.2259364                               2.63501165                             3.4348462
  PROPERTY DAMAGE AND ENVIRONMENTAL ROBBERY AND RELATED OFFENCES SERIOUS CRIMINAL TRESPASS THEFT AND RELATED OFFENCES
1                        -0.1376229                   -0.1097044                 -0.1341481                 -0.1117131
2                         3.7877580                    3.0193641                  3.6921223                  3.0746492
  HOMICIDE AND RELATED OFFENCES SEXUAL ASSAULT AND RELATED OFFENCES
1                   -0.05531273                          0.00102562
2                    1.52235707                         -0.02822787

Clustering vector:
   [1] 1 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
  [57] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [113] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [169] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [225] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [281] 1 2 1 2 2 2 1 2 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [337] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [393] 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [449] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [505] 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [561] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [617] 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1
 [673] 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1
 [729] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1
 [785] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [841] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1
 [897] 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [953] 1 1 1 2 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 [ reached getOption("max.print") -- omitted 255 entries ]

Within cluster sum of squares by cluster:
[1] 3505.841 4222.760
 (between_SS / total_SS =  31.5 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"    "size"
[8] "iter"         "ifault"
```