

MA304_2321248.Rmd

Sukanya Das

Packages

```
library(ggplot2)  
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(lubridate)
```

```
##  
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':  
##  
##   date, intersect, setdiff, union
```

```
library(corrplot)
```

```
## corrrplot 0.92 loaded
```

```
library(tidyr)  
library(leaflet)  
library(plotly)
```

```
##  
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':  
##  
##   last_plot
```

```
## The following object is masked from 'package:stats':  
##  
##   filter
```

```
## The following object is masked from 'package:graphics':  
##  
##   layout
```

Introduction

The datasets from Colchester 2023 i.e., crime and temperature, are provided for analysis and data exploration. The crime dataset contains street-level crime incidents which includes:

- Category: Category of the crime
- Persistent ID: Unique identifier for each crime
- Date: Date of the crime (YYYY-MM)
- Latitude: Latitude coordinates
- Longitude: Longitude coordinates

- Street ID: Unique identifier for the street
- Street Name: Name of the location where the crime happened
- Context: Extra information about the crime (if applicable)
- ID: ID of the crime (related to the API, not a police identifier)
- Location Type: Type of location (Force or BTP)
- Location Subtype: Type of location at which crime was recorded (BTP locations)
- Outcome Status: Category and date of the latest recorded outcome for the crime

Similarly, the temperature dataset contains daily climate data collected from a weather station close to Colchester, which includes:

- station_ID: WMO station identifier
- Date: Date of observations (YYYY-MM-DD)
- TemperatureCAvg (°C): Average air temperature
- TemperatureCMax (°C): Maximum air temperature
- TemperatureCMin (°C): Minimum air temperature
- TdAvgC (°C): Average dew point temperature
- HrAvg (%): Average relative humidity
- WindkmhDir : Direction of the wind
- WindkmhInt (km/h): Speed of the wind
- WindkmhGust (km/h): Wind gust
- PresslevHp (hPa): Sea level pressure
- Precmm (mm): Precipitation totals
- TotCLOct (octants): Total cloudiness
- lowCLOct (octants): Cloudiness by low-level clouds
- SunD1h (hours): Duration of sunshine
- VisKm (km): Visibility

- PreselevHp (hPa): Atmospheric pressure measured at the altitude of the station
- SnowDepcm (cm): Depth of snow cover

For these given datasets, data visualisation will be performed by outlining the key aspects of the datasets and providing a detailed explanation of the visualisation procedure in the subsequent steps.

Read Datasets

```
# Load the datasets i.e., Crime and Temperature  
crime_dataset <- read.csv("crime23.csv")  
temp_dataset <- read.csv("temp2023.csv")
```

The crime and temperature datasets are imported in order to proceed with the analysis.

Explore Crime and Temperature Datasets

```
# Get rows and columns of crime dataset  
crime_row <- nrow(crime_dataset)  
crime_col <- ncol(crime_dataset)  
head(crime_dataset)
```

```
##          category persistent_id    date    lat    long street_id
## 1 anti-social-behaviour      2023-01 51.88306 0.909136  2153366
## 2 anti-social-behaviour      2023-01 51.90124 0.901681  2153173
## 3 anti-social-behaviour      2023-01 51.88907 0.897722  2153077
## 4 anti-social-behaviour      2023-01 51.89122 0.901988  2153186
## 5 anti-social-behaviour      2023-01 51.89416 0.895433  2153012
## 6 anti-social-behaviour      2023-01 51.88050 0.909014  2153379
##          street_name context      id location_type
## 1      On or near Military Road    NA 107596596      Force
## 2              On or near          NA 107596646      Force
## 3 On or near Culver Street West    NA 107595950      Force
## 4      On or near Ryegate Road     NA 107595953      Force
## 5      On or near Market Close     NA 107595979      Force
## 6      On or near Lisle Road       NA 107595985      Force
## location_subtype outcome_status
## 1                <NA>
## 2                <NA>
## 3                <NA>
## 4                <NA>
## 5                <NA>
## 6                <NA>
```

```
# Get rows and columns of crime dataset
temp_row <- nrow(temp_dataset)
temp_col <- ncol(temp_dataset)
head(temp_dataset)
```

```
## station_ID      Date TemperatureCAvg TemperatureCMax TemperatureCMin TdAvgC
## 1      3590 2023-12-31           8.7           10.6           4.4      7.2
## 2      3590 2023-12-30           6.6           9.7           4.4      4.2
## 3      3590 2023-12-29           9.9          11.4           6.9      6.0
## 4      3590 2023-12-28           9.9          11.5           4.0      7.5
## 5      3590 2023-12-27           5.8          10.6           3.9      3.7
## 6      3590 2023-12-26           9.8          12.7           6.3      7.6
## HrAvg WindkmhDir WindkmhInt WindkmhGust PresslevHp Precmm TotClOct lowClOct
## 1  89.6          S      25.0          63.0      999.0      6.2      8.0      8.0
## 2  85.5         WSW      22.7          50.0     1006.9      0.4      4.6      6.5
## 3  77.2         SW      32.8          61.2     1003.6      0.8      6.5      6.7
## 4  84.6         SSW      32.2          70.4     1003.2      2.8      6.8      7.1
## 5  86.4         SW      13.2          37.1     1016.4      2.0      4.0      6.9
## 6  86.9         WSW      23.5          46.3     1006.2      4.4      6.5      7.4
## SunD1h Viskm PreselevHp SnowDepcm
## 1   0.0  26.3          NA          NA
## 2   1.1  48.3          NA          NA
## 3   0.1  26.7          NA          NA
## 4   0.0  25.1          NA          NA
## 5   3.2  30.1          NA          NA
## 6   0.0  45.8          NA          NA
```

```
# Get the structure of the datasets
str(crime_dataset)
```

```
## 'data.frame':   6878 obs. of  12 variables:
## $ category      : chr  "anti-social-behaviour" "anti-social-behaviour" "anti-social-behaviour" "anti-social
-behaviour" ...
## $ persistent_id : chr  "" "" "" "" ...
## $ date          : chr  "2023-01" "2023-01" "2023-01" "2023-01" ...
## $ lat          : num  51.9 51.9 51.9 51.9 51.9 ...
## $ long         : num  0.909 0.902 0.898 0.902 0.895 ...
## $ street_id    : int   2153366 2153173 2153077 2153186 2153012 2153379 2153105 2153541 2152937 2153107 ...
## $ street_name  : chr   "On or near Military Road" "On or near " "On or near Culver Street West" "On or near
Ryegate Road" ...
## $ context      : logi   NA NA NA NA NA NA ...
## $ id          : int   107596596 107596646 107595950 107595953 107595979 107595985 107596603 107596291 1075
96305 107596453 ...
## $ location_type : chr   "Force" "Force" "Force" "Force" ...
## $ location_subtype: chr   "" "" "" "" ...
## $ outcome_status : chr   NA NA NA NA ...
```

```
str(temp_dataset)
```

```
## 'data.frame': 365 obs. of 18 variables:
## $ station_ID : int 3590 3590 3590 3590 3590 3590 3590 3590 3590 3590 ...
## $ Date : chr "2023-12-31" "2023-12-30" "2023-12-29" "2023-12-28" ...
## $ TemperatureCAvg: num 8.7 6.6 9.9 9.9 5.8 9.8 12.5 10 9.6 10 ...
## $ TemperatureCMax: num 10.6 9.7 11.4 11.5 10.6 12.7 14.3 12 10.8 12.6 ...
## $ TemperatureCMin: num 4.4 4.4 6.9 4 3.9 6.3 9.5 8.4 8.1 8.1 ...
## $ TdAvgC : num 7.2 4.2 6 7.5 3.7 7.6 10.1 7 6.5 6.2 ...
## $ HrAvg : num 89.6 85.5 77.2 84.6 86.4 86.9 85.3 81.5 81.2 78.2 ...
## $ WindkmhDir : chr "S" "WSW" "SW" "SSW" ...
## $ WindkmhInt : num 25 22.7 32.8 32.2 13.2 23.5 34.1 32.7 34.1 37.5 ...
## $ WindkmhGust : num 63 50 61.2 70.4 37.1 46.3 72.3 61.2 68.6 77.8 ...
## $ PresslevHp : num 999 1007 1004 1003 1016 ...
## $ Precmm : num 6.2 0.4 0.8 2.8 2 4.4 0.8 0.8 0 2 ...
## $ TotClOct : num 8 4.6 6.5 6.8 4 6.5 7.8 5 8 7.5 ...
## $ lowClOct : num 8 6.5 6.7 7.1 6.9 7.4 7.8 6.7 8 7.5 ...
## $ SunD1h : num 0 1.1 0.1 0 3.2 0 0 2.9 0 1.4 ...
## $ VisKm : num 26.3 48.3 26.7 25.1 30.1 45.8 61.8 72.9 69.4 34.3 ...
## $ PreselevHp : logi NA NA NA NA NA NA ...
## $ SnowDepcm : int NA NA NA NA NA NA NA NA NA NA ...
```

```
# Check for missing values
colSums(is.na(crime_dataset))
```

```
##      category    persistent_id      date      lat
##          0              0          0          0
##      long      street_id  street_name  context
##          0              0          0      6878
##      id    location_type location_subtype outcome_status
##          0              0          0      677
```

```
colSums(is.na(temp_dataset))
```



```
##      station_ID      Date TemperatureCAvg TemperatureCMax TemperatureCMin
##           0           0              0              0              0
##      TdAvgC      HrAvg      WindkmhDir      WindkmhInt      WindkmhGust
##           0           0              0              0              0
##      PresslevHp      Precmm      TotCl0ct      lowCl0ct      SunD1h
##           0           27              0              13              82
##      VisKm      PreselevHp      SnowDepcm
##           0           365              364
```

```
# Explore the distributions of variables
summary(crime_dataset)
```

```
##      category      persistent_id      date      lat
## Length:6878      Length:6878      Length:6878      Min.   :51.88
## Class :character      Class :character      Class :character      1st Qu.:51.89
## Mode  :character      Mode  :character      Mode  :character      Median :51.89
##                                     Mean   :51.89
##                                     3rd Qu.:51.89
##                                     Max.   :51.90
##      long      street_id      street_name      context
## Min.   :0.8793      Min.   :2152702      Length:6878      Mode:logical
## 1st Qu.:0.8964      1st Qu.:2153025      Class :character      NA's:6878
## Median :0.9014      Median :2153158      Mode  :character
## Mean   :0.9030      Mean   :2153877
## 3rd Qu.:0.9088      3rd Qu.:2153365
## Max.   :0.9246      Max.   :2343256
##      id      location_type      location_subtype      outcome_status
## Min.   :107582824      Length:6878      Length:6878      Length:6878
## 1st Qu.:109309182      Class :character      Class :character      Class :character
## Median :111497486      Mode  :character      Mode  :character      Mode  :character
## Mean   :111301793
## 3rd Qu.:113746477
## Max.   :115699577
```

```
summary(temp_dataset)
```

```

##      station_ID      Date      TemperatureCAvg TemperatureCMax
## Min.   :3590      Length:365      Min.    :-2.60      Min.     : 1.70
## 1st Qu.:3590      Class :character 1st Qu.: 7.20      1st Qu.:10.60
## Median :3590      Mode  :character Median :10.40      Median :14.20
## Mean   :3590                      Mean   :10.92      Mean   :15.13
## 3rd Qu.:3590                      3rd Qu.:15.80      3rd Qu.:20.00
## Max.   :3590                      Max.    :23.10      Max.    :30.40
##
## TemperatureCMin      TdAvgC      HrAvg      WindkmhDir
## Min.    :-6.200      Min.    :-4.400      Min.    :43.10      Length:365
## 1st Qu.: 3.200      1st Qu.: 4.400      1st Qu.:75.60      Class :character
## Median : 6.300      Median : 7.600      Median :81.70      Mode  :character
## Mean   : 6.365      Mean   : 7.578      Mean   :81.25
## 3rd Qu.:10.600      3rd Qu.:11.200      3rd Qu.:87.90
## Max.   :16.300      Max.    :17.500      Max.    :97.90
##
##      WindkmhInt      WindkmhGust      PresslevHp      Precmm
## Min.    : 6.20      Min.    :13.00      Min.    : 967.4      Min.    : 0.000
## 1st Qu.:12.00      1st Qu.:31.50      1st Qu.:1006.3      1st Qu.: 0.000
## Median :16.10      Median :38.90      Median :1014.3      Median : 0.000
## Mean   :16.81      Mean   :40.87      Mean   :1013.6      Mean   : 1.866
## 3rd Qu.:20.20      3rd Qu.:48.20      3rd Qu.:1021.7      3rd Qu.: 1.150
## Max.   :37.50      Max.    :98.20      Max.    :1045.1      Max.    :33.600
##                                     NA's    :27
##      TotCl0ct      lowCl0ct      SunD1h      VisKm
## Min.    :0.000      Min.    :1.800      Min.    : 0.000      Min.    : 3.60
## 1st Qu.:3.600      1st Qu.:5.800      1st Qu.: 1.150      1st Qu.:22.70
## Median :5.100      Median :6.700      Median : 4.700      Median :31.50
## Mean   :4.988      Mean   :6.443      Mean   : 5.127      Mean   :32.11
## 3rd Qu.:7.000      3rd Qu.:7.400      3rd Qu.: 8.050      3rd Qu.:41.50
## Max.   :8.000      Max.    :8.000      Max.    :15.400      Max.    :72.90
##                                     NA's    :13      NA's    :82
## PreselevHp      SnowDepcm
## Mode:logical      Min.    :1
## NA's:365          1st Qu.:1
##                                     Median :1
##                                     Mean   :1
##                                     3rd Qu.:1

```

```
##           Max.      :1  
##           NA's      :364
```

The given datasets are explored and analysed using summary to get an idea of the minimum, maximum, mean, and quartile values. These are calculated for each numerical column in both datasets. In addition to that, missing values are observed for some columns in the datasets.

Crime dataset: The dataset contains 6878 rows and 12 columns, which include a range of crime reports. It also provides detailed information on specific crime incidents, such as the location, time, and outcome of those incidents. The limited range of longitude (0.8793 and 0.9246) and latitude (51.88 to a maximum of 51.90) coordinates indicates that the information provided covers a relatively small geographical area. Furthermore, there are few columns, i.e., context and outcome_status, which contain a considerable number of missing values.

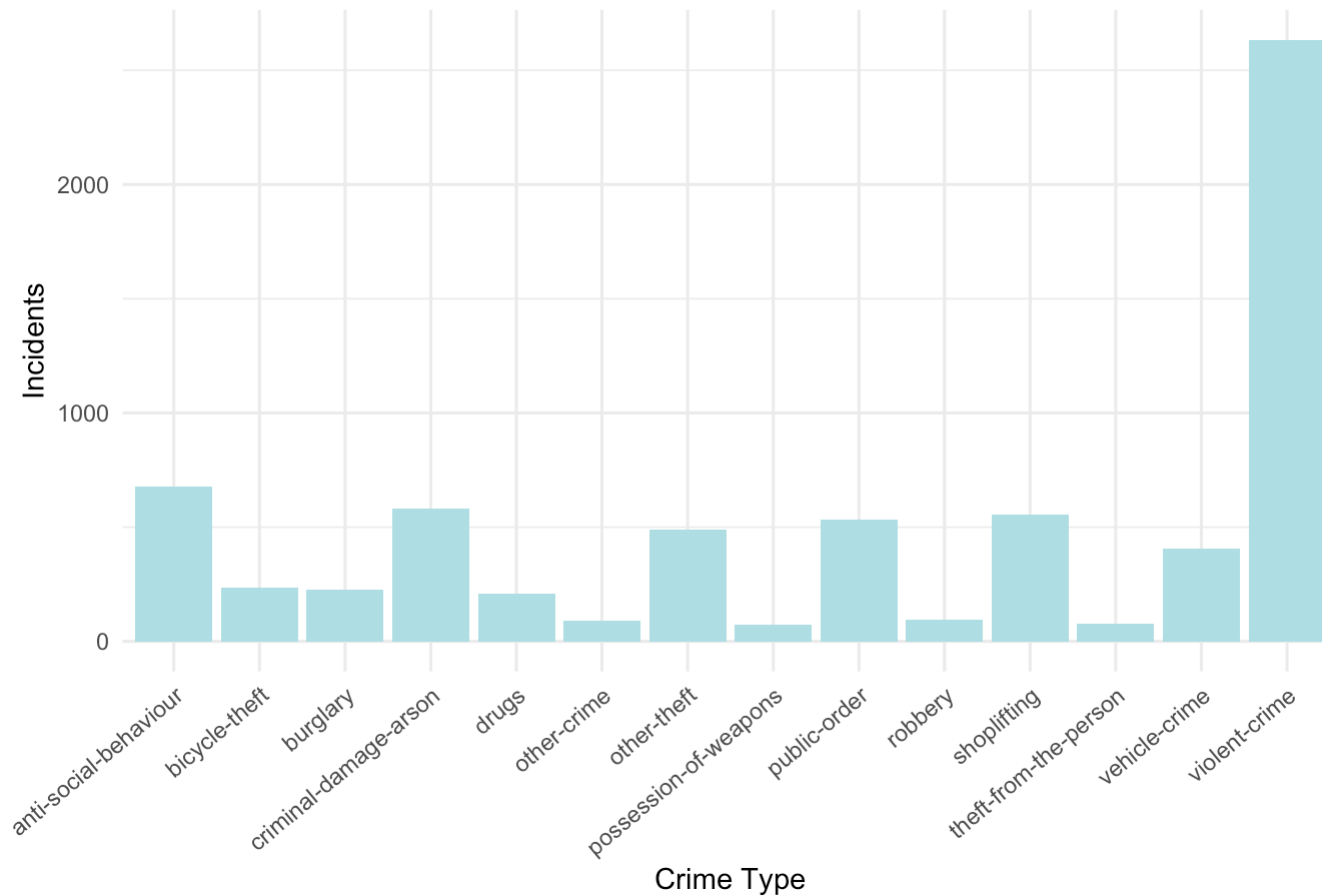
Temperature dataset: The dataset contains 365 rows and 18 columns showing the data of daily weather records from a particular weather station, i.e., ID: 3590, over a period of a year. The summary provides some crucial information, including the range of average temperature, which is from -2.60°C to 23.10°C, with an average of 10.92°C. Also, the dataset has some missing values in the Precmm, LowCIOct, SunD1h, PreselevHp, and SnowDepcm columns.

Both datasets are explored by visually representing the data through appropriate plots and tables in subsequent steps.

Bar Plot

```
# Create bar plot based on crime category  
crime_bar_gg <- ggplot(crime_dataset, aes(x = category))  
  
crime_barplot <- crime_bar_gg + geom_bar(fill = "powderblue") + labs(title = "Bar Plot of Crime Incidents based o  
n Category in Colchester (2023)", x = "Crime Type", y = "Incidents") + theme_minimal() + theme(axis.text.x = eleme  
nt_text(angle = 40, hjust = 1))  
  
# Display bar plot  
crime_barplot
```

Bar Plot of Crime Incidents based on Category in Colchester (2023)



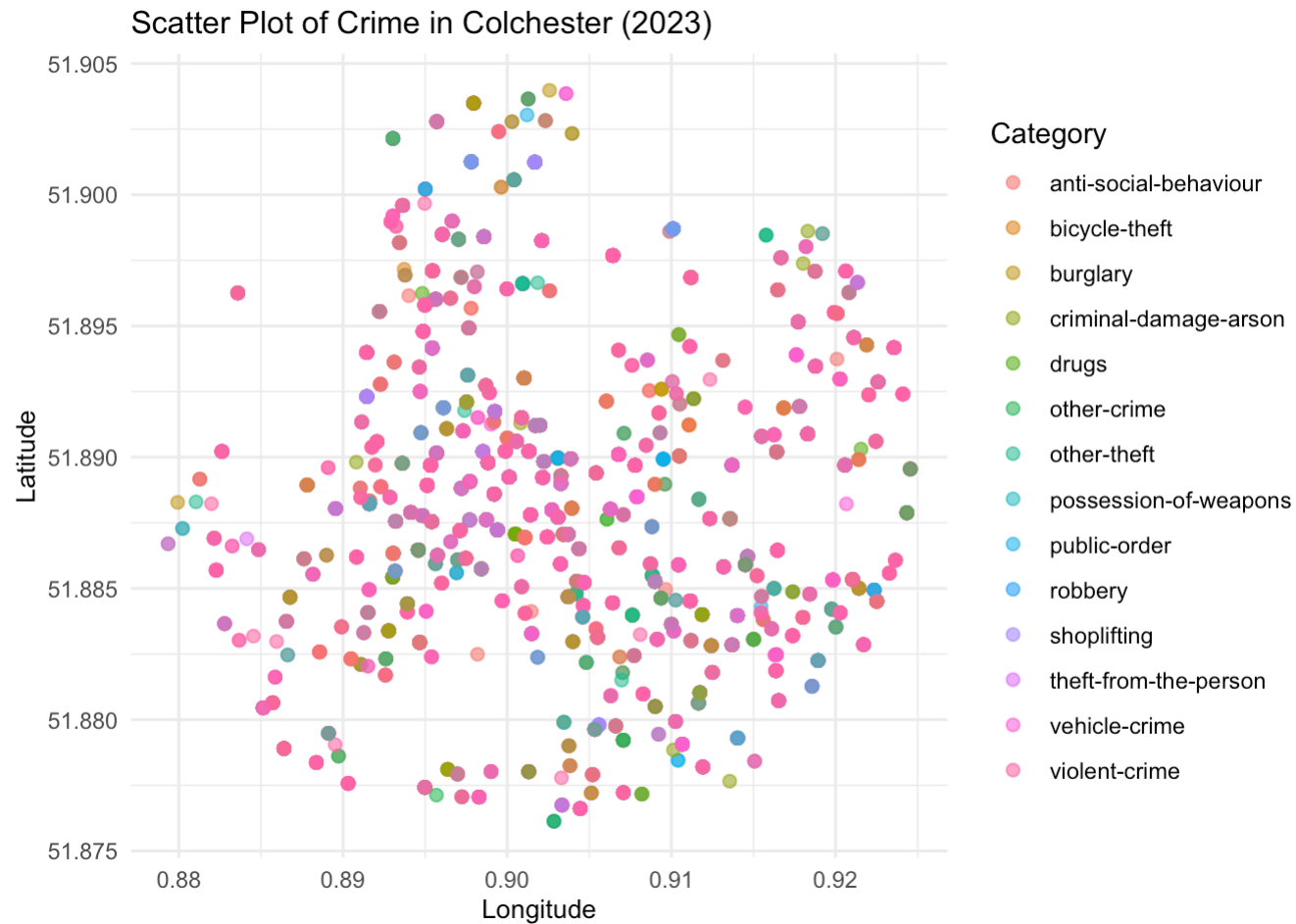
Primarily, the bar plot is observed to represent the instances of crimes based on category in Colchester's 2023. As shown by the plot, violent crime occurs more frequently as compared to any other crime, being the most common crime in Colchester. After violent crime, anti-social behaviour is the second most common crime category. Other crimes such as bicycle theft, burglary, drug and other theft, possession of weapons, public order, criminal damage and arson, shoplifting, and vehicle crime have a significant number of incidents. The crimes with the lowest occurrence rates include other crimes, theft from the person, other theft, and robbery.

Scatter Plot

```
# Create Scatter plot based on crime incident locations
crime_scatter_gg <- ggplot(crime_dataset, aes(x = long, y = lat, color = category))

crime_scatterplot <- crime_scatter_gg + geom_point(alpha = 0.6, size = 2) +
  labs(title = "Scatter Plot of Crime in Colchester (2023)",
    x = "Longitude", y = "Latitude", color = "Category") +
  theme_minimal() + theme(plot.title = element_text(size = 12),
    axis.title = element_text(size = 10))

# Display Scatter Plot
crime_scatterplot
```



Further, these crime incidents can be observed using a scatter plot based on geographical coordinates, which can provide a more precise and accurate visual representation. The scatter plot shows the latitude on the y-axis and longitude on the x-axis, with each point representing a crime incident. Regions with higher crime rates can be observed by the points that are concentrated in specific areas. Moreover, there is a significant distribution of crime over a wide range of latitudes and longitudes. This distribution suggests that the occurrence of crime is not confined to a certain region but rather extends across the entire area of Colchester.

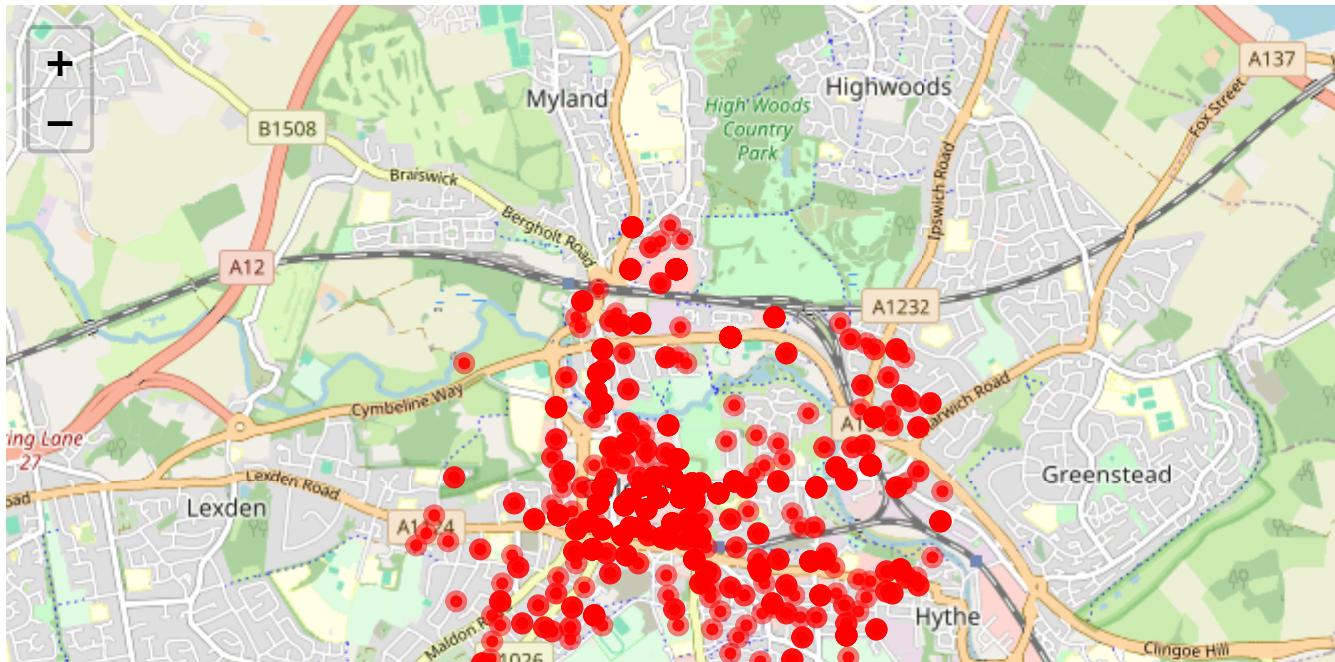
Leaflet Plot

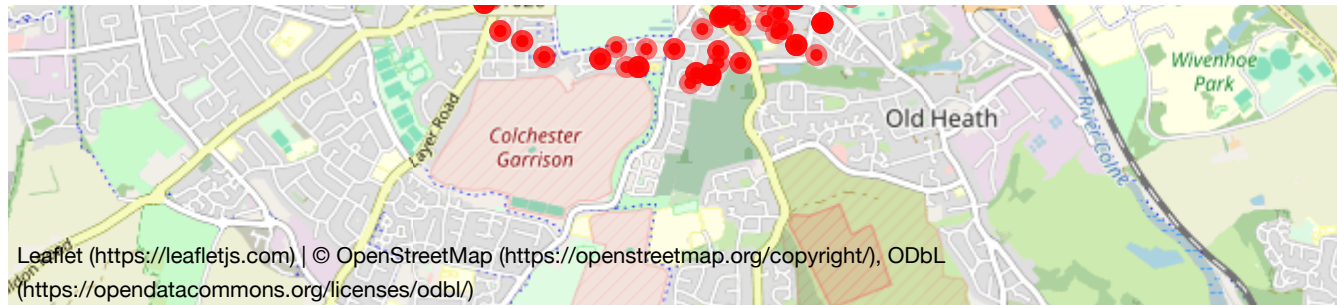
```
# Get 1000 random rows as a sample data
random_crime_data <- crime_dataset[sample(nrow(crime_dataset), 1000), ]
```

```
# Create Leaflet
leaflet_plot <- leaflet() %>%
  addTiles() %>%
  addCircleMarkers(data = random_crime_data,
    radius = 3,
    color = "red",
    fillOpacity = 0.9,
    popup = ~paste("Crime Type: ", category))
```

```
## Assuming "long" and "lat" are longitude and latitude, respectively
```

```
# Display leaflet/map
leaflet_plot
```





Afterwards, the crime types can be viewed in a specific location using a leaflet plot. This will give an idea of where most of the crimes occurred in 2023. In this scenario, the leaflet plot shows 1000 randomly chosen rows to illustrate a sample of crime data. The points in the plot suggest a particular crime case or a group of crime cases in certain areas of Colchester. The points are mainly concentrated in the central area of the plot, which indicates that this area has more reported crimes. Some areas, such as Highwoods, Colchester Garrison, and Greenstead, show fewer incidents. In general, the plot with sample data indicates that crime is more common in the central parts, i.e., Colchester Town, Hythe, etc.

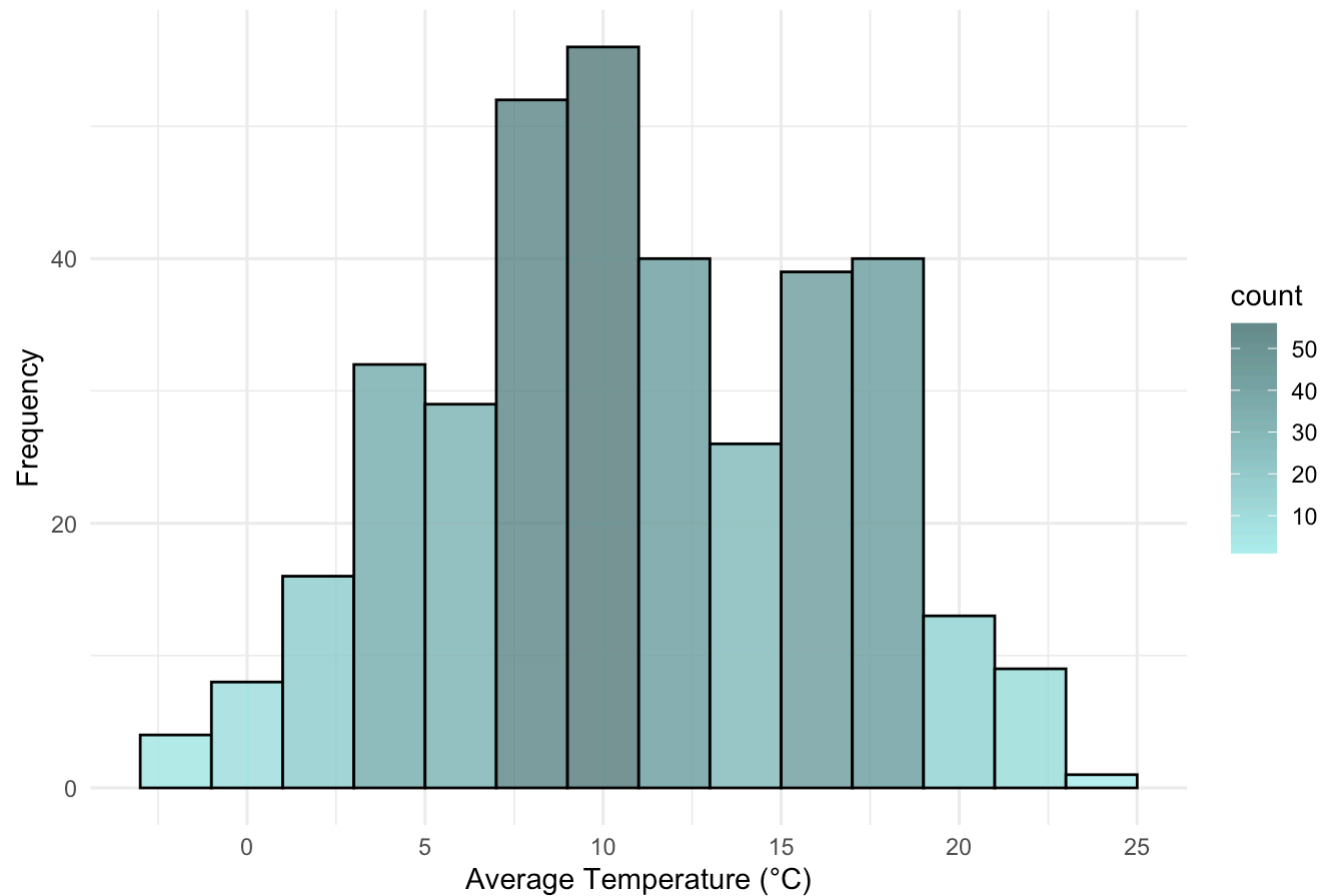
Histogram Plot

```
# Create histogram for average temperature
temp_hist_gg <- ggplot(temp_dataset, aes(x = TemperatureCAvg, fill = ..count..))

temp_histogram <- temp_hist_gg + geom_histogram(binwidth = 2, color = "black", alpha = 0.9) +
  labs(title = "Histogram of Temperature Averages Distribution in Colchester (2023)", x = "Average Temperature (°C)", y = "Frequency") + theme_minimal() + scale_fill_gradient(low = "paleturquoise", high = "paleturquoise4")

# Display histogram
temp_histogram
```


Histogram of Temperature Averages Distribution in Colchester (2023)



After observing the crime data for the year 2023, we will next explore the temperature dataset in order to find out the climatic condition for the entire year 2023. For this purpose, the distribution of average temperatures in Colchester is observed using histograms. The minor skewness towards the right indicates that the temperature mostly lies in the centre, i.e., around the middle temperature range. Furthermore, there are more instances of higher temperatures as compared to lower temperatures. The highest bar shows a frequent average temperature range of around 10°C to 15°C. Also, there is a wide range of temperatures, ranging from below 5°C to above 20°C, suggesting varying weather conditions throughout the year.

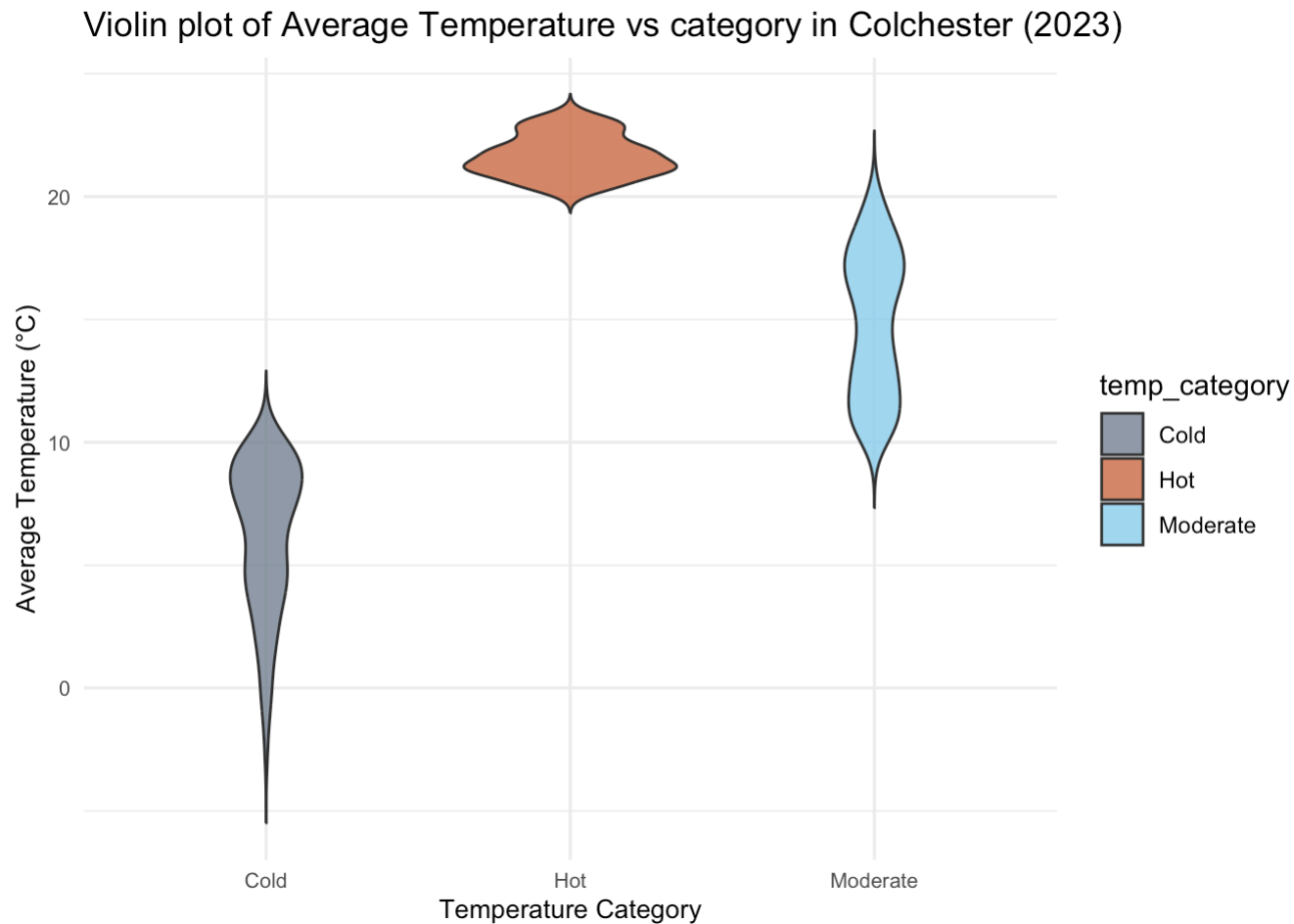
Violin Plot

```
# Set average temperatures to cold, moderate, and hot
temp_dataset <- temp_dataset %>%
  mutate(temp_category = case_when( TemperatureCAvg <= 10 ~ "Cold", TemperatureCAvg <= 20 ~ "Moderate", TRUE ~ "Hot"))

# Create violin plot
violin_gg <- ggplot(temp_dataset, aes(x = temp_category, y = TemperatureCAvg, fill = temp_category))

violin_plot <- violin_gg + geom_violin(trim = FALSE, width = 0.7, alpha = 0.8) +
  scale_fill_manual(values = c("Cold" = "slategray", "Moderate" = "skyblue", "Hot" = "sienna3")) + theme_minimal() +
  labs(title = "Violin plot of Average Temperature vs category in Colchester (2023)", x = "Temperature Category", y = "Average Temperature (°C)") +
  theme(axis.title = element_text(size = 10), axis.text = element_text(size = 8))

# Display the violin plot
violin_plot
```



Further, the above finding can be analysed using a violin plot to get a precise result. Here, the violin plot is divided into three groups of average temperatures, i.e., cold, moderate, and hot, which makes it easier to view the climatic/weather conditions more accurately. This distribution reflects the below observations:

- Cold temperatures show a narrow pattern, indicating low variation in the average temperature. From the plot, it can be observed that the median temperature seems to be between 5°C and 6°C.
- Moderate temperatures show a wider pattern, suggesting higher variation. In general, the range of the moderate type is almost from 7°C to 13°C.
- Hot temperatures indicate a minimal variance in the temperature. The shape of the pattern implies a relatively uniform spread of average temperatures within the hot category.

In general, cold temperatures fluctuate the least, moderate temperatures vary the most, and hot temperatures also reflect considerable variation.

```
# Assign additional column of date to both the datasets
crime_dataset$Date_new <- as.character(crime_dataset$date)
temp_dataset$Date_new <- substring(temp_dataset$Date, 1, 7)

# Combine both the datasets based on new formatted date
crime_temp_combined <- merge(temp_dataset, crime_dataset, by.x = "Date_new", by.y = "Date_new")
```

After that, by combining the crime and temperature datasets, we can determine the number of crimes that occur in specific climate conditions. From this analysis, we will verify if there's a relationship between temperature and climatic conditions.

Table

```
# Get temperature ranges
cold <- subset(crime_temp_combined, TemperatureCAvg <= 10)
moderate <- subset(crime_temp_combined, TemperatureCAvg > 10 & TemperatureCAvg <= 20)
hot <- subset(crime_temp_combined, TemperatureCAvg > 20)

# Get the counts within each temperature range
counts_cold <- table(cold$category)
counts_moderate <- table(moderate$category)
counts_hot <- table(hot$category)

# Bind the data and form a table
crime_temp_table <- rbind(counts_cold, counts_moderate, counts_hot)

# Add row names and column names
row.names(crime_temp_table) <- c("Cold", "Moderate", "Hot")
colnames(crime_temp_table) <- c("Anti social behaviour","Bicycle theft", "Burglary","Criminal damage","Drugs","Other crime","Other theft","Possession of weapons", "Public order","Robbery","Shoplifting","Theft from the person", "Vehicle crime","Violent crime")

# Form table, reduce margin, separate rows and columns, and adjust padding
table <- knitr::kable(crime_temp_table, "html", table.attr = "style='border-collapse: collapse; border: 1px solid black; margin: 0 auto;'",
table <- gsub("<table", "<table style='border: 1px solid black; margin: 0 auto;'", table)
table <- gsub("<tr", "<tr style='border: 1px solid black;'", table)
table <- gsub("<td", "<td style='border: 1px solid black; padding: 14px;'", table)

table
```

	Anti social behaviour	Bicycle theft	Burglary	Criminal damage	Drugs	Other crime	Other theft	Possession of weapons	Public order	Robbery	Shoplifting	Theft from the person	Vehicle crime	Violent crime
Cold	7884	3123	3109	9014	3045	1275	6951	1053	8155	1249	8153	1131	7179	37452

	Anti social behaviour	Bicycle theft	Burglary	Criminal damage	Drugs	Other crime	Other theft	Possession of weapons	Public order	Robbery	Shoplifting	Theft from the person	Vehicle crime	Violent crime
Moderate	11860	3741	3475	8152	3044	1410	7532	1137	7549	1470	8231	1097	4837	39974
Hot	862	285	248	529	234	114	477	71	485	137	526	83	375	2767

A table is created by combining both datasets to show the count of each crime within a specific temperature range. Below are the observations:

- The rate of anti-social behaviour is particularly high in moderate temperatures, followed by cold, with a minimal frequency in hot temperatures.
- Bicycle theft and burglary seem to occur more frequently in moderate temperatures and then in cold temperatures, with a substantial drop in hot temperatures.
- Criminal damage and arson are more frequent in moderate and cold temperatures as compared to hot climatic conditions.
- The frequency of drug-related crimes is most common in moderate temperatures, cold temperatures, and much lower in hot temperatures.
- Other types of crimes and thefts show a similar pattern, which occurs mostly in moderate and cold temperatures.
- Weapons possession is the most frequent case in moderate temperatures and the lowest in hot temperatures.
- Public order offences are most frequent in cold temperatures, followed by moderate temperatures, and are least frequent in hot temperatures.
- Shoplifting, robbery, and theft from the person are more common in moderate temperatures as compared to cold and hot.
- Vehicle crime is more common in cold temperatures, and then moderate, with a drop in hot climatic conditions.
- Violent crime is the major concern, with significantly higher cases in moderate temperatures compared to cold and hot temperatures.

In general, the crime occurs most in moderate or slightly warmer conditions.

Correlation Analysis

```
# Correlation analysis
correlation_matrix <- cor(crime_temp_combined[, c("TemperatureCAvg", "lat", "long")])

# Populate data to form correlation matrix
temp <- correlation_matrix[1,1]
temp_lat <- correlation_matrix[1,2]
temp_long <- correlation_matrix[1,3]
lat_temp <- correlation_matrix[2,1]
lat <- correlation_matrix[2,2]
lat_long <- correlation_matrix[2,3]
long_temp <- correlation_matrix[3,1]
long_lat <- correlation_matrix[3,2]
long <- correlation_matrix[3,3]
```

Correlation Matrix

Parameters	Average Temperature	Latitude	Longitude
Average Temperature	1	0.03	-0.01
Latitude	0.03	1	-0.09
Longitude	-0.01	-0.09	1

Next, the correlation is verified between the numeric variables, such as temperature and geographic coordinates. This analysis can help in determining the patterns in how temperature changes with latitude and longitude in Colchester in 2023. This information acts as a key to predict the weather conditions and their patterns. As shown in the table, the correlation for average temperature is 1 which represents a precise correlation with itself. There's a very weak positive linear relationship between average temperature and latitude, which is around 0.0272. Additionally, there's a very weak negative linear relationship between average temperature and longitude, with a correlation coefficient of approximately -0.0064. The correlation coefficient of approximately -0.0902 suggests a very weak negative relationship between latitude and longitude. Thus, the correlation coefficients suggest a weak linear relationship between average temperature and the geographical coordinates. This information does not give enough evidence for the relationship between temperature, latitude, and longitude. Therefore, more plots and tables are used for further analysis of these datasets in later steps to ensure a precise and accurate result.

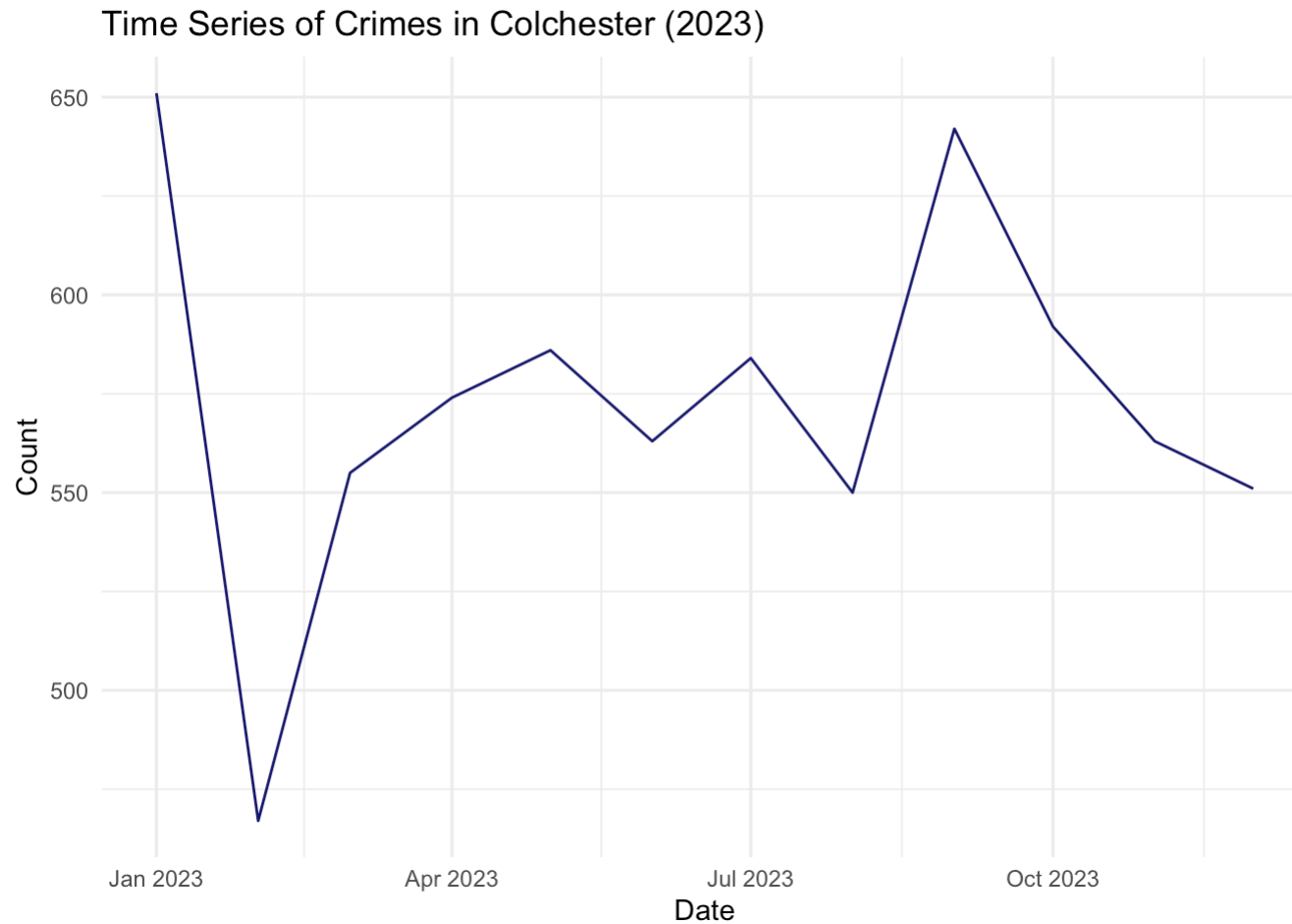
Time Series Plot

```
# Format Date
crime_dataset$date <- as.Date(paste(crime_dataset$date,"01", sep = "-"), format = "%Y-%m-%d")

# Time series plot of crime incidents
crime_ts_gg <- ggplot(crime_dataset, aes(x = date))

crime_ts <- crime_ts_gg + geom_line(stat = "count", color = "midnightblue") +
  theme_minimal() + labs(title = "Time Series of Crimes in Colchester (2023)",
    x = "Date", y = "Count")

# Display Time Series
crime_ts
```

Thereafter, the crime incidents are analysed based on the months of the year 2023. Here, the time series plot depicts the incidence of crime in Colchester for the entire year 2023. Based on the observation, there's a variation in the crime rate over the months. For instance, there's a substantial decrease in crime in February, which is below 500 cases. The peak can be observed around May and also in July, when crime cases surpass 550. Afterwards, there's a drop in August, followed by a significant increase around September exceeding 650 cases, i.e., the highest peak over the months, and a further drop in October. In general, this may suggest that high crime rates mostly occur at moderate temperatures and slightly warmer climatic conditions.

Time Series Plot with Smoothing

```
# Get count of the incidents for based on date
crime_summary <- crime_dataset %>%
  group_by(date) %>%
  summarize(count = n())

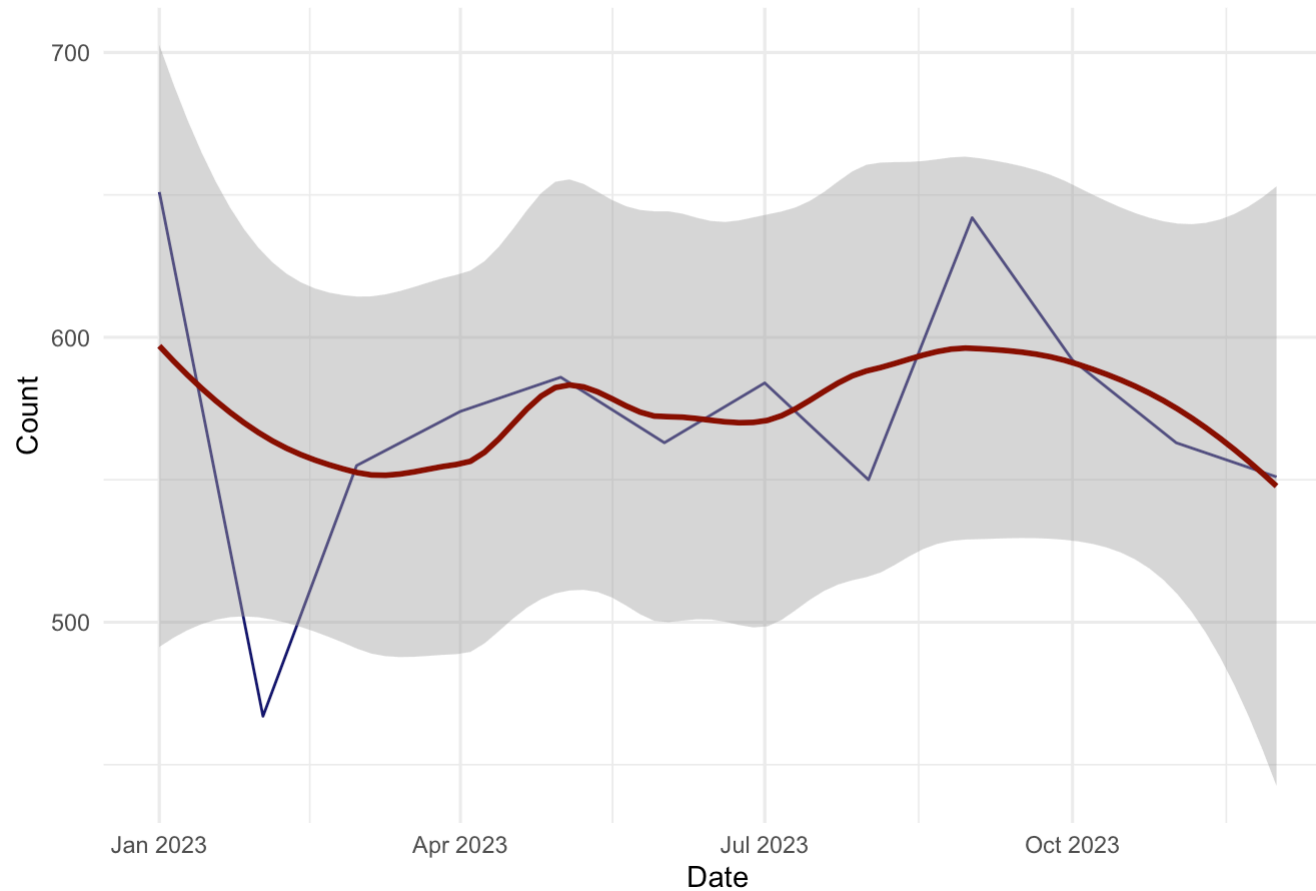
# Time series plot of crime incidents with smoothing
crime_tss_gg <- ggplot(crime_summary, aes(x = date, y = count))

crime_ts_smooth <- crime_tss_gg + geom_line(color = "midnightblue") +
  geom_smooth(method = "loess", color = "red4") + theme_minimal() +
  labs(title = "Time Series of Crimes in Colchester (2023) with smoothing", x = "Date", y = "Count")

# Display time series with smoothing
crime_ts_smooth
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Time Series of Crimes in Colchester (2023) with smoothing



Again, the time series plot is analysed using the smoothing line. As shown in the graph, the red line overlaps with the original data, and the shaded grey area shows some variation related to the smoothed mean. By reducing the variations, the smoothing line helps to determine the data pattern. From the very beginning of the year until the end of April, there's been a slight decrease in the crime rate. The time period between May and July shows a slight increase in crime rates, and there's a marked increase in high crime rates from August until September. Towards the end of October, the smoothed line shows a drop in crime cases. Overall, the plot indicates the majority of the cases occur at mild or slightly warmer temperatures.

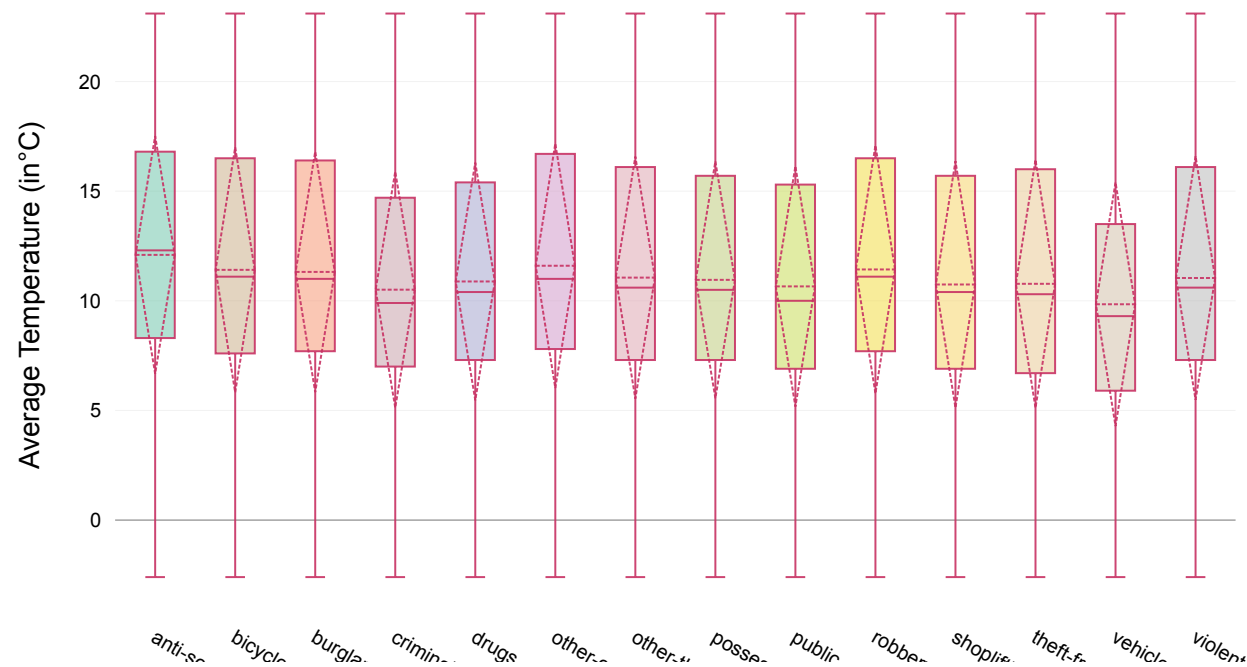
Interactive Plot

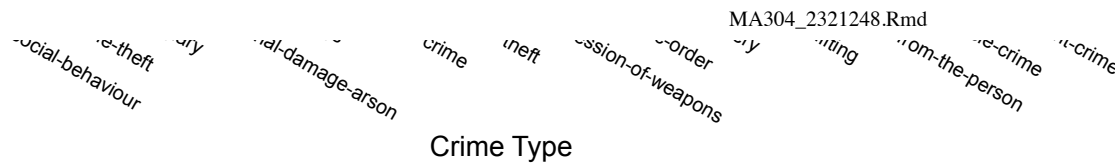
```
# Interactive Box plot
box_plot <- plot_ly( crime_temp_combined, x = ~category, y = ~TemperatureCAvg,
  color = ~category, type = "box", marker = list(color = "rgba(200, 50, 100, 0.7)"),
  line = list(color = "rgba(200, 50, 100, 0.9)", width = 1), boxmean = "sd") %>%

  layout( title = "Interactive plot of Average Temperature and Crime Types in Colchester (2023)", xaxis = list(title = "Crime Type", tickfont = list(family = "Arial", size = 10)), yaxis = list(title = "Average Temperature (in° C)", tickfont = list(family = "Arial", size = 10)),
  showlegend = FALSE, font = list(color = "black", family = "Arial"),
  margin = list(l = 70, r = 20, b = 70, t = 70, pad = 6))

# Display box plot
box_plot
```

Interactive plot of Average Temperature and Crime Types in Colchester (2023)





Finally, an interactive box plot is analysed, showing the min, max, mean, median, and quartiles of the average temperature in °C along with crime types. This plot shows that the median temperatures for each crime category lie mostly between 10°C and 12°C. Also, the maximum temperature for all the crime types is around 23.1°C, which could indicate the highest temperature observed in the dataset. Additionally, the minimum temperature is uniform throughout all crime types, i.e., -2.6°C. As the plot shows, there are temperature variations within different crimes, and seems to be a slight temperature dependence based on the median temperatures. Crimes occur across a wide range of temperatures, with the majority of incidents resulting in moderate to warmer climate conditions. The plot also indicates that crimes are occurring in the most extreme temperatures as well.

Conclusion

In summary, based on various plots and tables, we have observed and analysed how crime is related to temperature, geographic location, and time period in Colchester during 2023. The most common crime types are violent crime and anti-social behaviour, with violent crime being the major concern in Colchester in 2023. Other crimes, such as bicycle theft, burglary, drug use, etc., are less common as compared to violent crime and anti-social behaviour but have a great impact on general safety. Despite being less frequent, these crimes can still have a substantial impact on both society and the individuals affected. Additionally, by analysing the scatter plot and leaflet plot, the crime incidents are widely spread throughout Colchester rather than being confined to a particular area. As per the analysis, central Colchester has a significant rate of crime, with some outskirt areas experiencing fewer cases of crime. After combining the crime and temperature datasets, the table of crime counts within a particular climatic range suggests a relationship between moderate temperatures and a rise in crime for most crime types. In addition, the violin plot shows the most variation during moderate temperatures. Moreover, the most usual temperature range is 10°C to 15°C in Colchester 2023, with a skew towards moderate or slightly warmer temperatures. Also, throughout-the-year variations in the crime rate are shown by the time series plot, with a notable increase in September, followed by May and July. Furthermore, the smoothed version of this plot highlights a pattern of higher crime rates during milder or warmer months. Again, by observing the interactive box plot, median temperatures for each crime type suggest that most crimes take place in moderate climatic conditions. But, the correlation matrix shows a weak correlation between the average temperature and geographical coordinates. This may indicate that the crime might be affected by other factors besides geographic location and temperature. Hence, climatic factors such as moderate or slightly warmer temperatures are one of the factors contributing to the higher crime incidents. Beside this, the economic condition, mental well-being, and policing strategies, might influence crime rates in Colchester for the year 2023.