Analyzing the Relationship Between Crime and Weather Patterns in Colchester: A Data Visualization Approach

Reg No: 2401510

2025-04-26

1. Introduction

Here, we are investigating how weather affects street-level crime in Colchester during 2024. The main goal is to find out if things like temperature, rain, and wind are related to how often crimes happen and what types of crimes occur. We use two datasets: one with crime reports and one with daily weather data.

First, we clean both datasets by fixing missing values, removing unnecessary columns, and making sure the formats match. Then, we group the crime data by month so it can be compared with the weather data, which we also average by month. This helps us combine the two datasets in a meaningful way.

Next, we use different kinds of charts to explore the data. Bar charts and tables show how many crimes happened and what types they were. Pie charts and dot plots show more about the categories of crimes. Histograms and density plots show how weather data is spread out. We also use more advanced visuals like smoothed time series, scatter plots, and correlation charts to find trends over time and how different factors are connected. These visual tools help us discover patterns and can be useful for planning public safety policies in the future.

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(lubridate)
library(ggplot2)
library(leaflet)
```

```
## Warning: package 'leaflet' was built under R version 4.4.3
```

```r
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
```

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.4.3
```

```
## corrplot 0.95 loaded
```

```
library(DT)
```

```
## Warning: package 'DT' was built under R version 4.4.3
```

```
library(naniar)
```

```
## Warning: package 'naniar' was built under R version 4.4.3
```

2. Data Loading

We're using two main datasets in this project: crime24.csv, which has street-level crime data for Colchester in 2024, and temp24.csv, which has daily weather information. To keep things consistent, we renamed the "Date" column in the weather data to lowercase "date" so it matches with the crime data. We then used the head() function to look at the first few rows of each dataset. This gives us a quick idea of what the data looks like and helps us check that everything loaded correctly.

```
# Loading datasets
crime <- read_csv("crime24.csv")
```

```
## New names:
## Rows: 6304 Columns: 13
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (7): category, persistent_id, date, street_name, location_type, location... dbl
## (5): ...1, lat, long, street_id, id lgl (1): context
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```r
temp <- read_csv("temp24.csv")
```

```
## Rows: 366 Columns: 18
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr   (1): WindkmhDir
## dbl  (15): station_ID, TemperatureCAvg, TemperatureCMax, TemperatureCMin, Td...
## lgl   (1): PreselevHp
## date  (1): Date
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Renaming 'Date' to 'date' in temp
temp <- temp %>% rename(date = Date)

# This is how we will observe the structure
head(crime)
```

```
## # A tibble: 6 x 13
##    ...1 category   persistent_id date    lat  long street_id street_name context
##   <dbl> <chr>      <chr>         <chr> <dbl> <dbl>     <dbl> <chr>       <lgl>
## 1     1 anti-soci~ <NA>          2024~  51.9 0.901   2153130 On or near~ NA
## 2     2 anti-soci~ <NA>          2024~  51.9 0.899   2153105 On or near~ NA
## 3     3 anti-soci~ <NA>          2024~  51.9 0.902   2153147 On or near~ NA
## 4     4 anti-soci~ <NA>          2024~  51.9 0.888   2152856 On or near~ NA
## 5     5 anti-soci~ <NA>          2024~  51.9 0.890   2152871 On or near~ NA
## 6     6 anti-soci~ <NA>          2024~  51.9 0.899   2153107 On or near~ NA
## # i 4 more variables: id <dbl>, location_type <chr>, location_subtype <chr>,
## #   outcome_status <chr>
```

```r
head(temp)
```

```
## # A tibble: 6 x 18
##   station_ID date       TemperatureCAvg TemperatureCMax TemperatureCMin TdAvgC
##        <dbl> <date>               <dbl>           <dbl>           <dbl>  <dbl>
## 1       3590 2024-12-31             6.5             7.7               5    4.4
## 2       3590 2024-12-30             5.6             6.9             3.4    4.9
## 3       3590 2024-12-29             3.3             4.9             2.2    3.2
## 4       3590 2024-12-28             4               5.8             2.3    3.7
## 5       3590 2024-12-27             5.3             6.7             4.3    5.1
## 6       3590 2024-12-26             6.7            10               5.6    6.4
## # i 12 more variables: HrAvg <dbl>, WindkmhDir <chr>, WindkmhInt <dbl>,
## #   WindkmhGust <dbl>, PresslevHp <dbl>, Precmm <dbl>, TotClOct <dbl>,
## #   lowClOct <dbl>, SunD1h <dbl>, VisKm <dbl>, SnowDepcm <dbl>,
## #   PreselevHp <lgl>
```

3. Data Summary and Missing Values

Summary statistics and missing data checks were conducted to better understand the structure and quality of the datasets. The summary() function was used to explore variable types, ranges, and typical values, which is critical before moving on to data cleaning and visualization. To identify missing data, colSums(is.na()) was applied to both datasets.

```r
# Let's look at the summary for each dataset
summary(crime)
```

```
##       ...1        category         persistent_id          date          
##  Min.   :   1   Length:6304        Length:6304        Length:6304       
##  1st Qu.:1577   Class :character   Class :character   Class :character  
##  Median :3152   Mode  :character   Mode  :character   Mode  :character  
##  Mean   :3152                                                           
##  3rd Qu.:4728                                                           
##  Max.   :6304                                                           
##       lat             long          street_id        street_name       
##  Min.   :51.88   Min.   :0.8788   Min.   :2152686   Length:6304       
##  1st Qu.:51.89   1st Qu.:0.8966   1st Qu.:2153025   Class :character  
##  Median :51.89   Median :0.9013   Median :2153155   Mode  :character  
##  Mean   :51.89   Mean   :0.9029   Mean   :2153873                     
##  3rd Qu.:51.89   3rd Qu.:0.9088   3rd Qu.:2153366                     
##  Max.   :51.90   Max.   :0.9246   Max.   :2343256                     
##  context              id            location_type      location_subtype  
##  Mode:logical   Min.   :115954844   Length:6304        Length:6304       
##  NA's:6304      1st Qu.:118009952   Class :character   Class :character  
##                 Median :120228058   Mode  :character   Mode  :character  
##                 Mean   :120403000                                        
##                 3rd Qu.:122339060                                        
##                 Max.   :125550731                                        
##  outcome_status    
##  Length:6304       
##  Class :character  
##  Mode  :character  
##                    
##                    
## 
```

```r
summary(temp)
```

```
##    station_ID        date             TemperatureCAvg TemperatureCMax
##  Min.   :3590   Min.   :2024-01-01   Min.   :-2.60   Min.   : 1.10  
##  1st Qu.:3590   1st Qu.:2024-04-01   1st Qu.: 7.00   1st Qu.:10.72  
##  Median :3590   Median :2024-07-01   Median :10.95   Median :14.75  
##  Mean   :3590   Mean   :2024-07-01   Mean   :10.98   Mean   :15.08  
##  3rd Qu.:3590   3rd Qu.:2024-09-30   3rd Qu.:14.50   3rd Qu.:19.60  
##  Max.   :3590   Max.   :2024-12-31   Max.   :23.10   Max.   :29.80  
##  TemperatureCMin      TdAvgC           HrAvg        WindkmhDir       
##  Min.   :-6.100   Min.   :-6.000   Min.   :59.60   Length:366        
##  1st Qu.: 3.325   1st Qu.: 4.725   1st Qu.:75.90   Class :character  
##  Median : 6.800   Median : 8.200   Median :82.75   Mode  :character  
##  Mean   : 6.486   Mean   : 7.752   Mean   :81.74                     
##  3rd Qu.: 9.500   3rd Qu.:11.000   3rd Qu.:88.80                     
##  Max.   :16.700   Max.   :16.900   Max.   :98.60                     
## 
```

```
##    WindkmhInt     WindkmhGust       PresslevHp         Precmm      
##  Min.   : 3.90   Min.   : 11.10   Min.   : 978.9   Min.   : 0.000  
##  1st Qu.:12.22   1st Qu.: 31.50   1st Qu.:1007.5   1st Qu.: 0.000  
##  Median :15.80   Median : 38.90   Median :1013.8   Median : 0.200  
##  Mean   :16.52   Mean   : 40.81   Mean   :1013.7   Mean   : 1.864  
##  3rd Qu.:19.80   3rd Qu.: 48.20   3rd Qu.:1021.0   3rd Qu.: 1.600  
##  Max.   :42.50   Max.   :105.60   Max.   :1037.3   Max.   :38.000  
##                                                    NA's   :24      
##     TotClOct        lowClOct         SunD1h           VisKm      
##  Min.   :0.000   Min.   :1.000   Min.   : 0.000   Min.   : 0.10  
##  1st Qu.:3.800   1st Qu.:5.800   1st Qu.: 0.325   1st Qu.:20.73  
##  Median :5.600   Median :6.900   Median : 3.500   Median :30.95  
##  Mean   :5.304   Mean   :6.609   Mean   : 4.203   Mean   :31.42  
##  3rd Qu.:7.200   3rd Qu.:7.600   3rd Qu.: 7.100   3rd Qu.:41.20  
##  Max.   :8.000   Max.   :8.000   Max.   :15.600   Max.   :71.20  
##                  NA's   :5                                       
##    SnowDepcm    PreselevHp    
##  Min.   :1.00   Mode:logical  
##  1st Qu.:1.25   NA's:366      
##  Median :1.50                 
##  Mean   :1.50                 
##  3rd Qu.:1.75                 
##  Max.   :2.00                 
##  NA's   :364                  
```

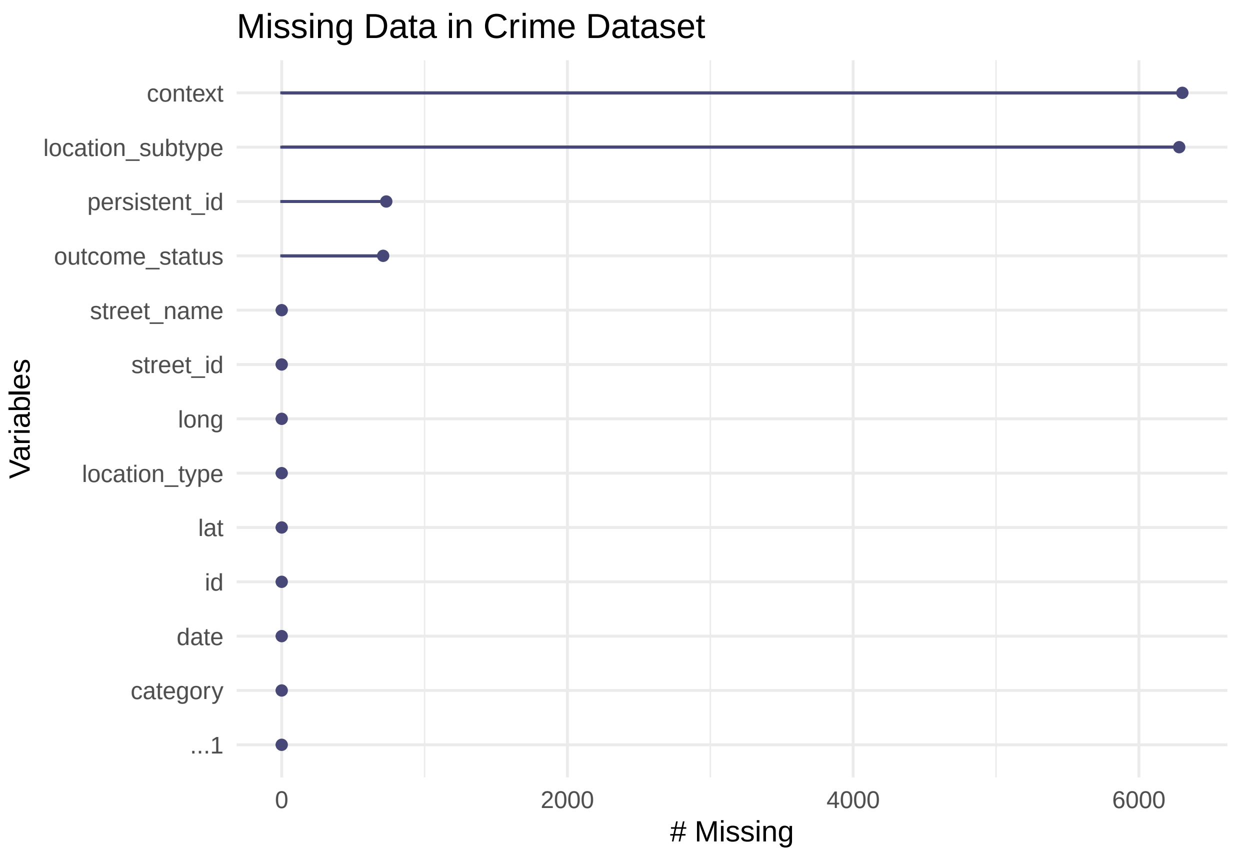
```
# Checking for NA values
colSums(is.na(crime))
```

```
##             ...1         category    persistent_id             date 
##                0                0              732                0 
##              lat             long        street_id      street_name 
##                0                0                0                0 
##          context               id    location_type location_subtype 
##             6304                0                0             6282 
##   outcome_status 
##              710 
```

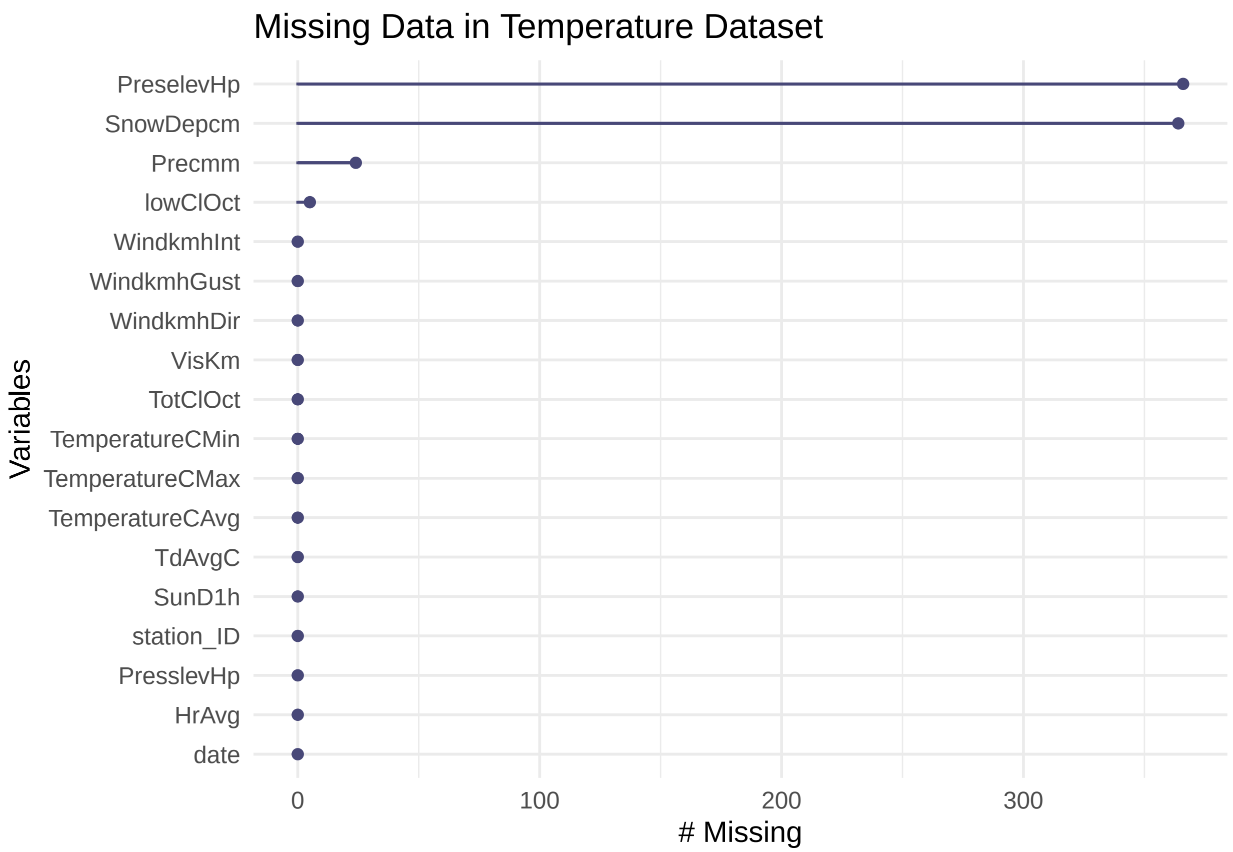
```
colSums(is.na(temp))
```

```
##      station_ID            date TemperatureCAvg TemperatureCMax TemperatureCMin 
##               0               0               0               0               0 
##          TdAvgC           HrAvg      WindkmhDir      WindkmhInt     WindkmhGust 
##               0               0               0               0               0 
##      PresslevHp          Precmm        TotClOct        lowClOct          SunD1h 
##               0              24               0               5               0 
##           VisKm       SnowDepcm      PreselevHp 
##               0             364             366 
```

```
# Missing Data visualisation
gg_miss_var(crime) + ggtitle("Missing Data in Crime Dataset")
```



```
gg_miss_var(temp) + ggtitle("Missing Data in Temperature Dataset")
```



In the crime data, outcome_status had 710 missing values, while context and location_subtype were entirely missing. In the temperature data, Precmm had 24 missing entries, lowClOct had 5, and both SnowDepcm and PreselevHp were largely or completely missing.

To visualize missingness, the gg_miss_var() function from the naniar package was used. This created clear plots showing the proportion of missing data across variables, confirming earlier findings and helping to decide which fields should be dropped, imputed, or retained for analysis.

4. Data Cleaning and Preprocessing

The data was cleaned by removing columns that were either unnecessary or lacked sufficient information. In the crime dataset, variables like context and location_subtype were excluded because they contained either fully missing or nearly empty entries. Likewise, the temperature dataset had columns such as PreselevHp, which was entirely missing, and SnowDepcm, which had too little usable data — both were dropped to minimize noise and focus on meaningful variables.

```
# Droping context and location_subtype Columns
crime_clean <- crime %>%
  select(-context, -location_subtype, -persistent_id )

temp_clean <- temp %>%
  select(-PreselevHp, -SnowDepcm)

# Replacing NA outcome_status with "Unknown" in crime file
crime_clean <- crime_clean %>%
```

```r
  mutate(outcome_status = ifelse(is.na(outcome_status), "Unknown", outcome_status))

# Filling NA in Precmm with 0, and mean-impute lowClOct in temp file
temp_clean <- temp_clean %>%
  mutate(
    Precmm = ifelse(is.na(Precmm), 0, Precmm),
    lowClOct = ifelse(is.na(lowClOct), mean(lowClOct, na.rm = TRUE), lowClOct)
  )
# Making sure there are no missing values

sapply(crime_clean, function(x) sum(is.na(x)))
```

```
##           ...1       category           date            lat           long 
##              0              0              0              0              0 
##      street_id    street_name             id  location_type outcome_status 
##              0              0              0              0              0 
```

```r
sapply(temp_clean, function(x) sum(is.na(x)))
```

```
##      station_ID            date TemperatureCAvg TemperatureCMax TemperatureCMin 
##               0               0               0               0               0 
##         TdAvgC           HrAvg      WindkmhDir      WindkmhInt     WindkmhGust 
##               0               0               0               0               0 
##      PresslevHp          Precmm        TotClOct        lowClOct          SunD1h 
##               0               0               0               0               0 
##           VisKm 
##               0 
```

```r
colnames(temp_clean)
```

```
##  [1] "station_ID"      "date"            "TemperatureCAvg" "TemperatureCMax"
##  [5] "TemperatureCMin" "TdAvgC"          "HrAvg"           "WindkmhDir"     
##  [9] "WindkmhInt"      "WindkmhGust"     "PresslevHp"      "Precmm"         
## [13] "TotClOct"        "lowClOct"        "SunD1h"          "VisKm"          
```

Remaining missing values were then handled: in the crime data, outcome_status NAs were replaced with “Unknown” to retain those records. In the temperature data, missing rainfall (Precmm) was assumed to be zero, while missing values in lowClOct were imputed using the mean.

A final check confirmed that the cleaned datasets were complete and ready for further processing. crime_clean had no remaining missing values, and temp_clean had all relevant numeric fields filled. This comprehensive cleaning ensures the datasets are suitable for accurate aggregation, merging, and visualization in the next steps.

Additionally, the persistent_id column, which could be used for longitudinal tracking of crime records, was dropped due to high sparsity. Since our analysis focuses on aggregation and visualization for a single time period, this identifier was not essential.

5. Time Formatting, Aggregation and Merging

Crime and weather data operated on different time scales—monthly for crime and daily for weather—so the weather data was aggregated to a monthly level. This adjustment allowed for a more meaningful comparison between the two datasets, reducing short-term fluctuations and highlighting broader seasonal trends.

We focused on three weather variables: average temperature, total rainfall, and average wind speed. These influence human behavior and may impact crime. For instance, warm weather can lead to more public activity (and more opportunities for crime), while rain may keep people indoors. Wind affects comfort and visibility, possibly influencing both criminal behavior and police response.

```r
# Treating it as a string, and just assign to 'month'
crime_clean <- crime %>%
  select(-context, -location_subtype) %>%
  mutate(
    outcome_status = ifelse(is.na(outcome_status), "Unknown", outcome_status),
    month = date  # date is already in "YYYY-MM" format
  )
# Grouping by month
crime_monthly <- crime_clean %>%
  group_by(month) %>%
  summarise(crime_count = n())

# Ensuring character type
crime_monthly$month <- as.character(crime_monthly$month)

# Creating temp_monthly by summarising weather data per month
temp_monthly <- temp_clean %>%
  mutate(month = format(date, "%Y-%m")) %>%
  group_by(month) %>%
  summarise(
    avg_temp = mean(TemperatureCAvg, na.rm = TRUE),
    total_rain = sum(Precmm, na.rm = TRUE),
    avg_wind = mean(WindkmhInt, na.rm = TRUE)  #   Corrected here
  )


# Merging with crime_monthly and temp_monthly
merged_monthly <- left_join(crime_monthly, temp_monthly, by = "month")

# View it
head(merged_monthly)
```

```
## # A tibble: 6 x 5
##   month   crime_count avg_temp total_rain avg_wind
##   <chr>         <int>    <dbl>      <dbl>    <dbl>
## 1 2024-01         529     4.25       53.8     19.9
## 2 2024-02         546     7.68       92.6     19.2
## 3 2024-03         502     8.14       58.4     16.6
## 4 2024-04         471     9.08       54.4     20.4
## 5 2024-05         568    13.4        80.6     13.6
## 6 2024-06         490    14.3        25.2     14.7
```

6 Exploratory Data Analysis

We began the analysis by summarizing the frequency of each crime type using a one-way table and then cross-tabulated those with their respective outcome statuses in a two-way table. This approach revealed not only the most common types of crime but also shed light on how often they were resolved—or left unresolved—by the authorities.

6.1 Frequency Tables

```r
# One-way frequency table of crime types
crime_table <- crime_clean %>%
  count(category, sort = TRUE)
datatable(crime_table, options = list(pageLength = 15), caption = "Frequency of Crime Types")
```

```r
# Two-way table: Crime category by outcome status
crime_2way <- table(crime_clean$category, crime_clean$outcome_status)
library(knitr)
kable(crime_2way, caption = "Crime Type by Outcome Status")
```

Table 1: Crime Type by Outcome Status

	Action to be taken by another organisation	Awaiting court outcome	Court result unavailable	Formal action is not in the public interest	Further action is not in the public interest	Further investigation is not in the public interest	Investigation complete; no suspect identified	Local resolution	Offender given a caution	Status update unavailable	Suspect charged as part of another case	Unable to prosecute suspect	Under investigation	Unknown
anti-social-behaviour	0	0	0	0	0	0	0	0	0	0	0	0	0	710
bicycle-theft	3	3	2	0	0	0	113	0	0	2	0	16	10	0
burglary	1	10	6	0	0	0	108	0	0	7	0	26	13	0
criminal-damage-arson	6	28	34	0	1	0	246	9	11	10	0	110	24	0
drugs	2	13	19	7	4	0	18	118	16	9	0	16	43	0
other-crime	4	7	12	2	1	1	14	1	1	5	0	43	9	0
other-theft	1	6	5	2	0	0	283	1	0	12	1	74	27	0
possession-of-weapons	2	8	9	0	1	0	6	4	4	8	0	16	7	0
public-order	8	20	20	11	0	0	165	12	3	27	0	160	32	0
robbery	0	7	4	1	0	0	41	0	0	4	0	25	3	0
shoplifting	5	82	65	2	0	0	313	25	1	7	0	92	37	0
theft-from-the-person	2	2	0	0	0	0	66	0	0	2	0	14	5	0
vehicle-crime	1	7	6	0	0	0	208	0	0	3	0	33	12	0
violent-crime	84	96	107	12	7	1	446	31	20	141	0	1195	280	0

```
crime_2way
```

```
##
##                         Action to be taken by another organisation
##   anti-social-behaviour                                          0
##   bicycle-theft                                                  3
##   burglary                                                       1
##   criminal-damage-arson                                          6
##   drugs                                                          2
##   other-crime                                                    4
##   other-theft                                                    1
##   possession-of-weapons                                          2
##   public-order                                                   8
##   robbery                                                        0
##   shoplifting                                                    5
##   theft-from-the-person                                          2
##   vehicle-crime                                                  1
##   violent-crime                                                 84
##
##                         Awaiting court outcome Court result unavailable
##   anti-social-behaviour                      0                        0
##   bicycle-theft                              3                        2
##   burglary                                  10                        6
##   criminal-damage-arson                     28                       34
##   drugs                                     13                       19
##   other-crime                                7                       12
##   other-theft                                6                        5
##   possession-of-weapons                      8                        9
##   public-order                              20                       20
##   robbery                                    7                        4
##   shoplifting                               82                       65
##   theft-from-the-person                      2                        0
##   vehicle-crime                              7                        6
##   violent-crime                             96                      107
##
##                         Formal action is not in the public interest
##   anti-social-behaviour                                           0
##   bicycle-theft                                                   0
##   burglary                                                        0
##   criminal-damage-arson                                           0
##   drugs                                                           7
##   other-crime                                                     2
##   other-theft                                                     2
##   possession-of-weapons                                           0
##   public-order                                                   11
##   robbery                                                         1
##   shoplifting                                                     2
##   theft-from-the-person                                           0
##   vehicle-crime                                                   0
##   violent-crime                                                  12
##
##                         Further action is not in the public interest
##   anti-social-behaviour                                            0
##   bicycle-theft                                                0
##   burglary                                                     0
##   criminal-damage-arson                                        1
##   drugs                                                        4
##   other-crime                                                  1
##   other-theft                                                  0
##   possession-of-weapons                                        1
##   public-order                                                 0
##   robbery                                                      0
##   shoplifting                                                  0
##   theft-from-the-person                                        0
##   vehicle-crime                                                0
##   violent-crime                                                7
##                        
##                         Further investigation is not in the public interest
##   anti-social-behaviour                                                   0
##   bicycle-theft                                                           0
##   burglary                                                                0
##   criminal-damage-arson                                                   0
##   drugs                                                                   0
##   other-crime                                                             1
##   other-theft                                                             0
##   possession-of-weapons                                                   0
##   public-order                                                            0
##   robbery                                                                 0
##   shoplifting                                                             0
##   theft-from-the-person                                                   0
##   vehicle-crime                                                           0
##   violent-crime                                                           1
##                        
##                         Investigation complete; no suspect identified
##   anti-social-behaviour                                             0
##   bicycle-theft                                                   113
##   burglary                                                        108
##   criminal-damage-arson                                           246
##   drugs                                                            18
##   other-crime                                                      14
##   other-theft                                                     283
##   possession-of-weapons                                             6
##   public-order                                                    165
##   robbery                                                          41
##   shoplifting                                                     313
##   theft-from-the-person                                            66
##   vehicle-crime                                                   208
##   violent-crime                                                   446
##                        
##                         Local resolution Offender given a caution
##   anti-social-behaviour                0                        0
##   bicycle-theft                        0                        0
##   burglary                             0                        0
##   criminal-damage-arson                9                       11
##   drugs                              118                       16
##   other-crime                          1                        1
##   other-theft                          1                        0
##   possession-of-weapons                4                        4
##   public-order                        12                        3
##   robbery                              0                        0
##   shoplifting                         25                        1
##   theft-from-the-person                0                        0
##   vehicle-crime                        0                        0
##   violent-crime                       31                       20
##
##                          Status update unavailable
##   anti-social-behaviour                         0
##   bicycle-theft                                 2
##   burglary                                      7
##   criminal-damage-arson                        10
##   drugs                                         9
##   other-crime                                   5
##   other-theft                                  12
##   possession-of-weapons                         8
##   public-order                                 27
##   robbery                                       4
##   shoplifting                                   7
##   theft-from-the-person                         2
##   vehicle-crime                                 3
##   violent-crime                               141
##
##                          Suspect charged as part of another case
##   anti-social-behaviour                                       0
##   bicycle-theft                                               0
##   burglary                                                    0
##   criminal-damage-arson                                       0
##   drugs                                                       0
##   other-crime                                                 0
##   other-theft                                                 1
##   possession-of-weapons                                       0
##   public-order                                                0
##   robbery                                                     0
##   shoplifting                                                 0
##   theft-from-the-person                                       0
##   vehicle-crime                                               0
##   violent-crime                                               0
##
##                          Unable to prosecute suspect Under investigation Unknown
##   anti-social-behaviour                           0                   0     710
##   bicycle-theft                                  16                  10       0
##   burglary                                       26                  13       0
##   criminal-damage-arson                         110                  24       0
##   drugs                                          16                  43       0
##   other-crime                                    43                   9       0
##   other-theft                                    74                  27       0
##   possession-of-weapons                          16                   7       0
##   public-order                                  160                  32       0
##   robbery                                        25                   3       0
##   shoplifting                                    92                  37       0
##   theft-from-the-person                          14                   5       0
##   vehicle-crime                                  33                  12       0
##    violent-crime                                  1195                  280        0
```

Violent crime emerges as the most common category, with shoplifting, other theft, and criminal damage/arson following. Notably, 1,195 violent crimes were marked 'unable to prosecute suspect' and 446 as 'no suspect identified', highlighting investigative challenges.

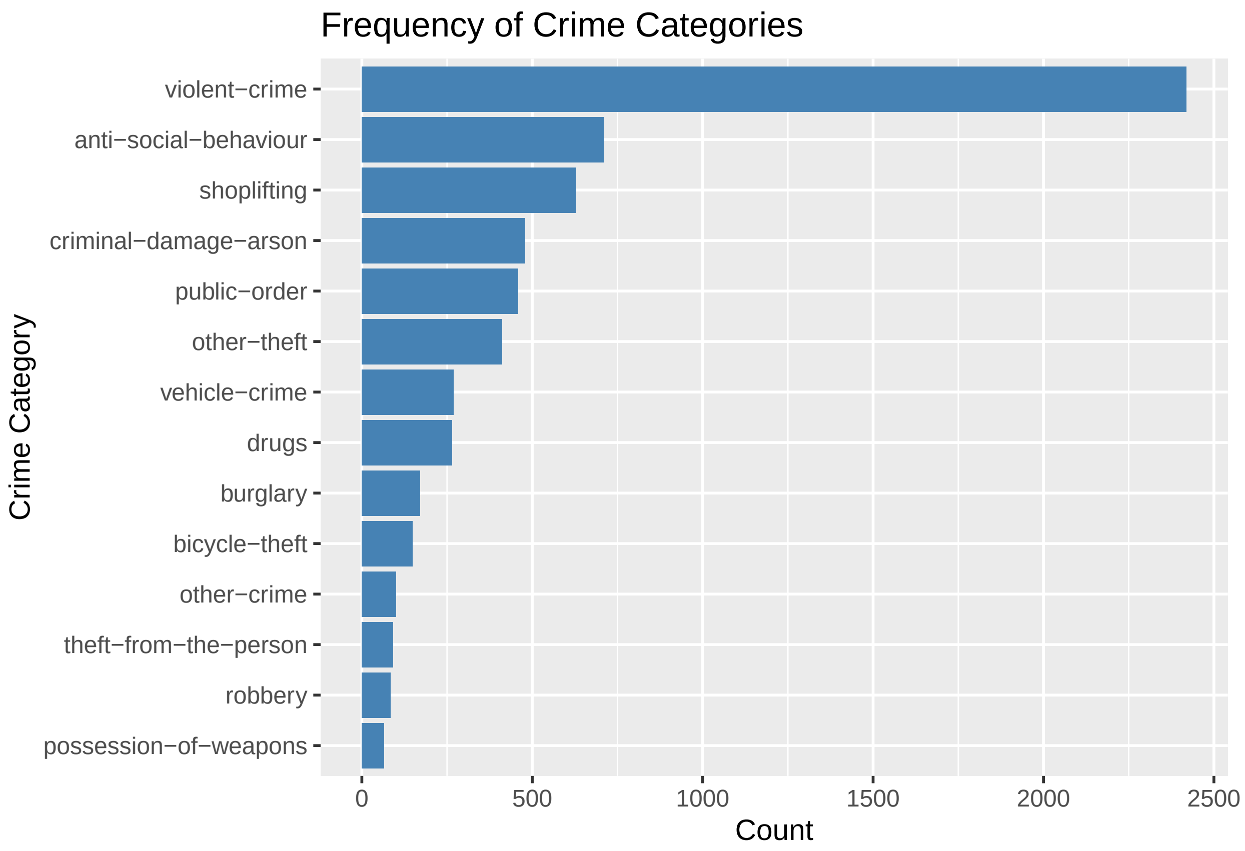
Drug offences often result in clear outcomes—cautions, formal actions, or local resolutions—suggesting easier prosecution. Weapon possession and public order crimes also show higher resolution rates.

In contrast, all 710 anti-social behaviour cases are marked "Unknown," indicating widespread underreporting or classification gaps, and raising concerns about how these incidents are tracked and addressed.

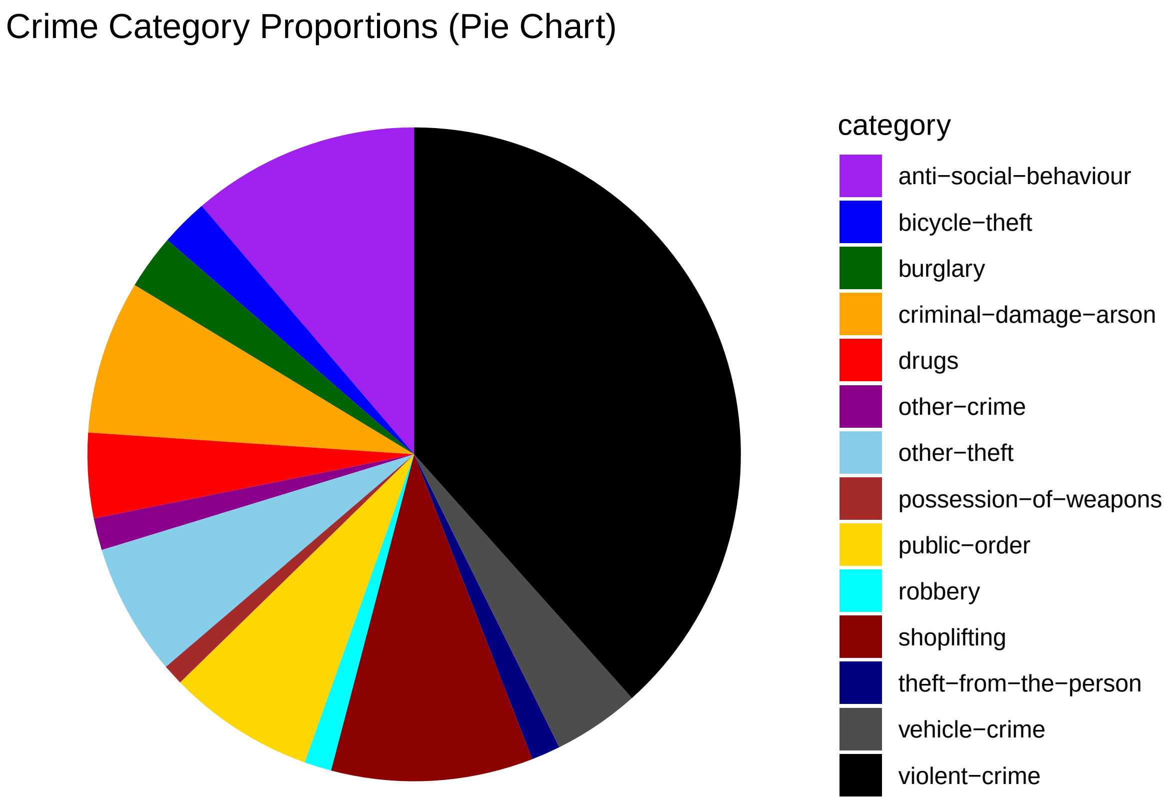
6.2 Bar Plot / Pie Chart / Dot Plot

We visualized the frequency of each category using three distinct yet complementary methods: a bar plot, a pie chart, and a dot plot to better understand the distribution of crime categories in Colchester during 2024. These plots all used the same data but offered different lenses through which to interpret it.

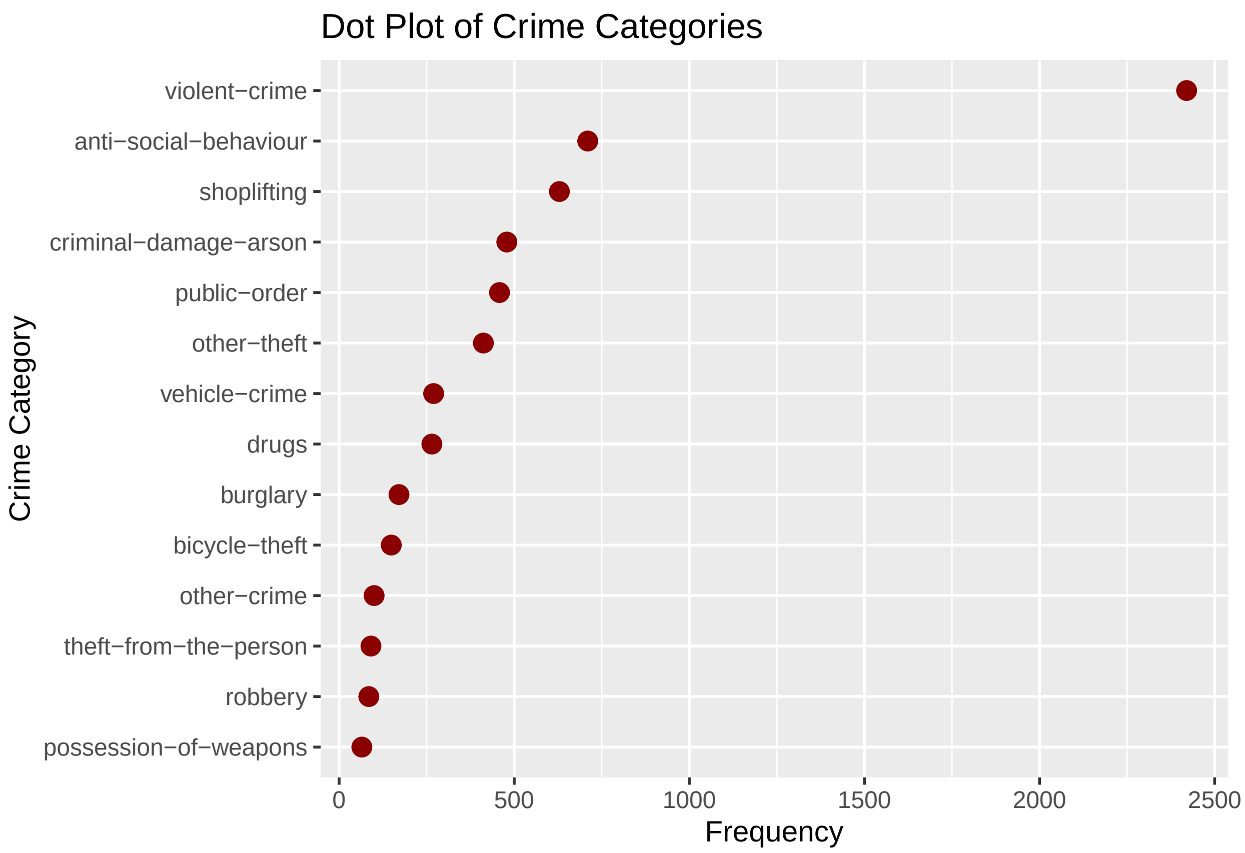
```r
# Bar plot of crime categories
crime_clean %>%
  count(category, sort = TRUE) %>%
  ggplot(aes(x = fct_reorder(category, n), y = n)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  coord_flip() +
  labs(title = "Frequency of Crime Categories",
       x = "Crime Category", y = "Count")
```



```r
# Pie Chart
crime_pie <- crime_clean %>%
  count(category)


my_colors <- c(
  "anti-social-behaviour" = "purple",
  "bicycle-theft" = "blue",
  "burglary" = "darkgreen",
  "criminal-damage-arson" = "orange",
  "drugs" = "red",
  "other-crime" = "darkmagenta",
  "other-theft" = "skyblue",
  "possession-of-weapons" = "brown",
  "public-order" = "gold",
  "robbery" = "cyan",
  "shoplifting" = "darkred",
  "theft-from-the-person" = "navy",
  "vehicle-crime" = "grey30",
  "violent-crime" = "black"
)

ggplot(crime_pie, aes(x = "", y = n, fill = category)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y") +
  scale_fill_manual(values = my_colors) +
  theme_void() +
  labs(title = "Crime Category Proportions (Pie Chart)")
```



```
# Dot Plot - Alternative to Bar
crime_dot <- crime_clean %>%
  count(category, sort = TRUE)

ggplot(crime_dot, aes(x = reorder(category, n), y = n)) +
  geom_point(color = "darkred", size = 3) +
  coord_flip() +
  labs(title = "Dot Plot of Crime Categories",
       x = "Crime Category", y = "Frequency")
```



The bar plot highlights violent crime as the most frequent, far surpassing other types like anti-social behaviour, shoplifting, and criminal damage/arson. These four dominate the overall crime pattern, with the bar plot enabling clear comparisons.

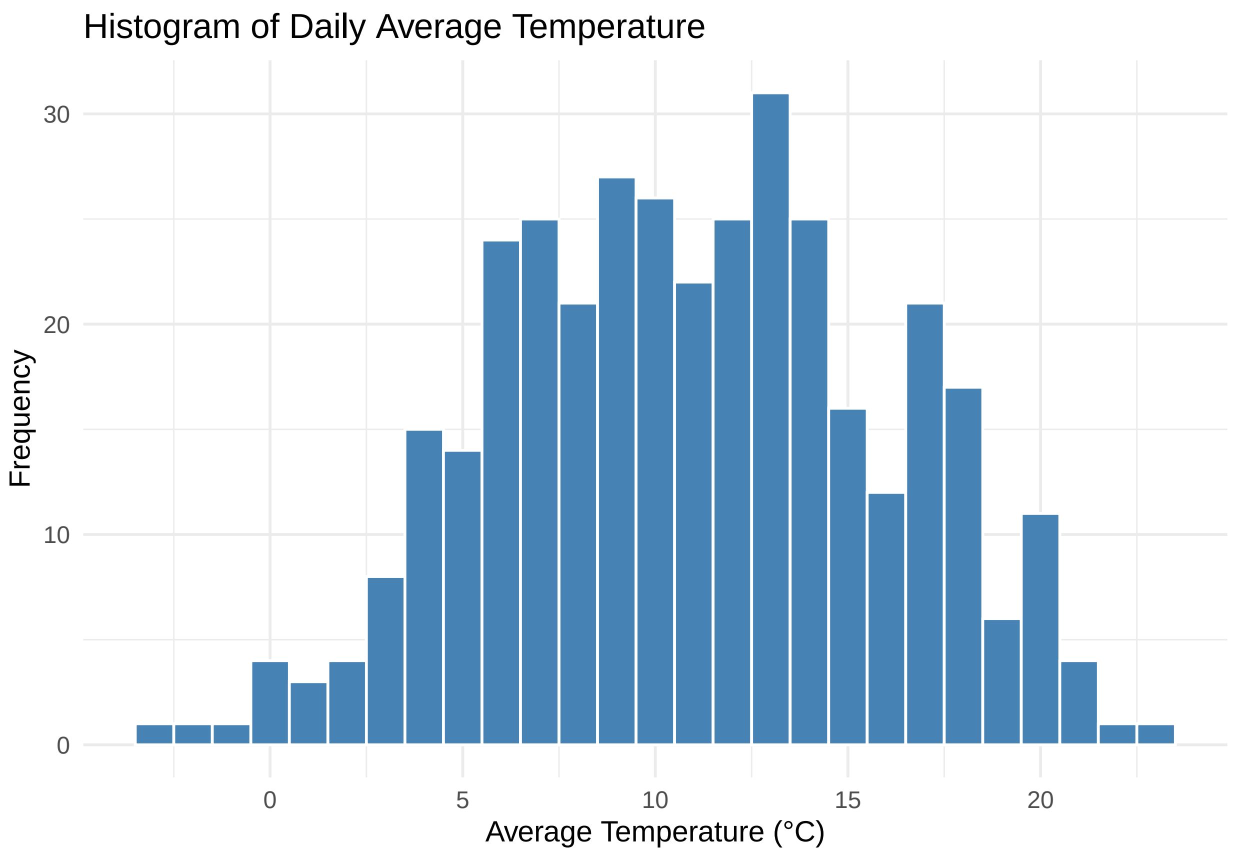
The pie chart, though less precise, emphasizes proportions—most notably the large black segment for violent crime. Custom colors enhance clarity and contrast.

The dot plot offers a clean, position-based view of frequencies. It supports earlier insights, showing violent crime as most common, while crimes like weapon possession, robbery, and personal theft rank lowest.

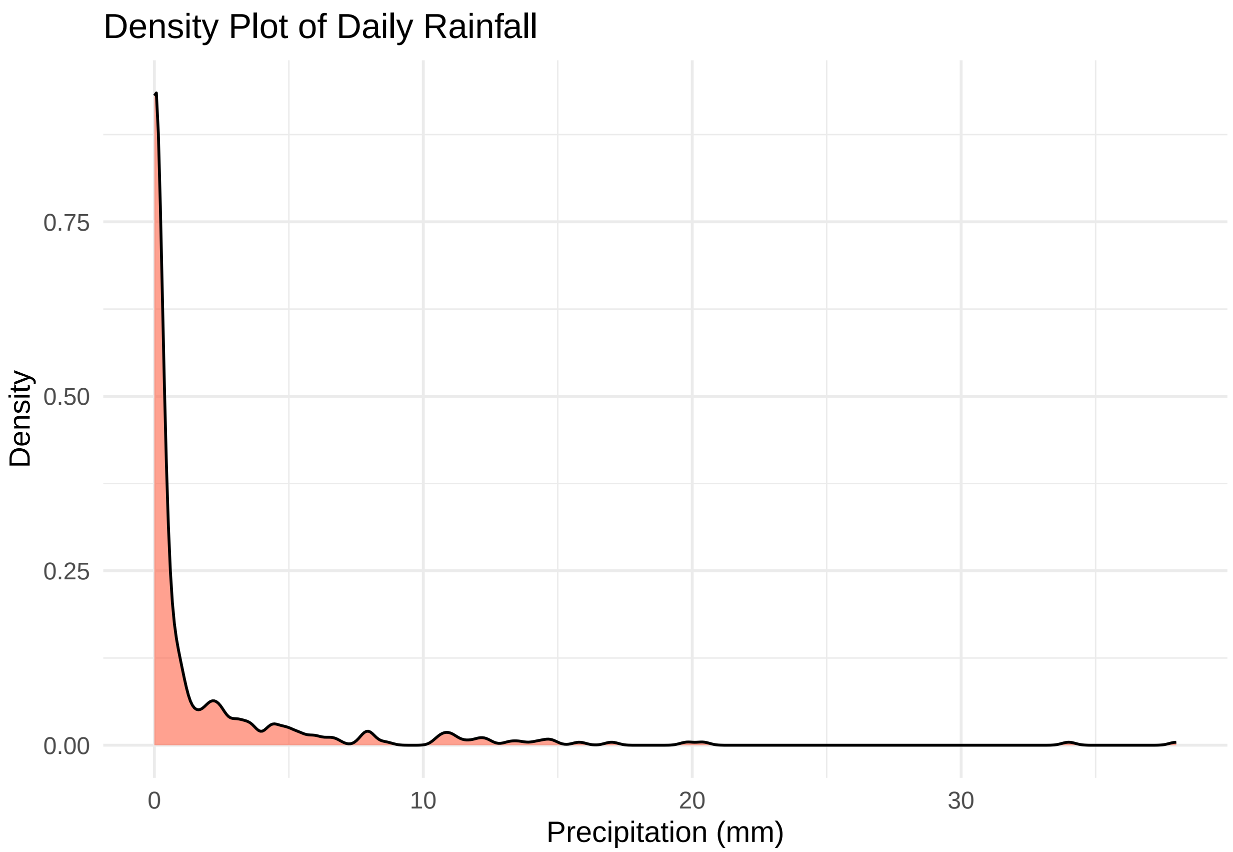
6.3. Histogram and Density Plots

Let us look into the distribution of key weather variables—average daily temperature and precipitation—using histograms and density plots. These visualizations provide insight into the central tendencies, variability, and skewness of weather conditions in Colchester during 2024.

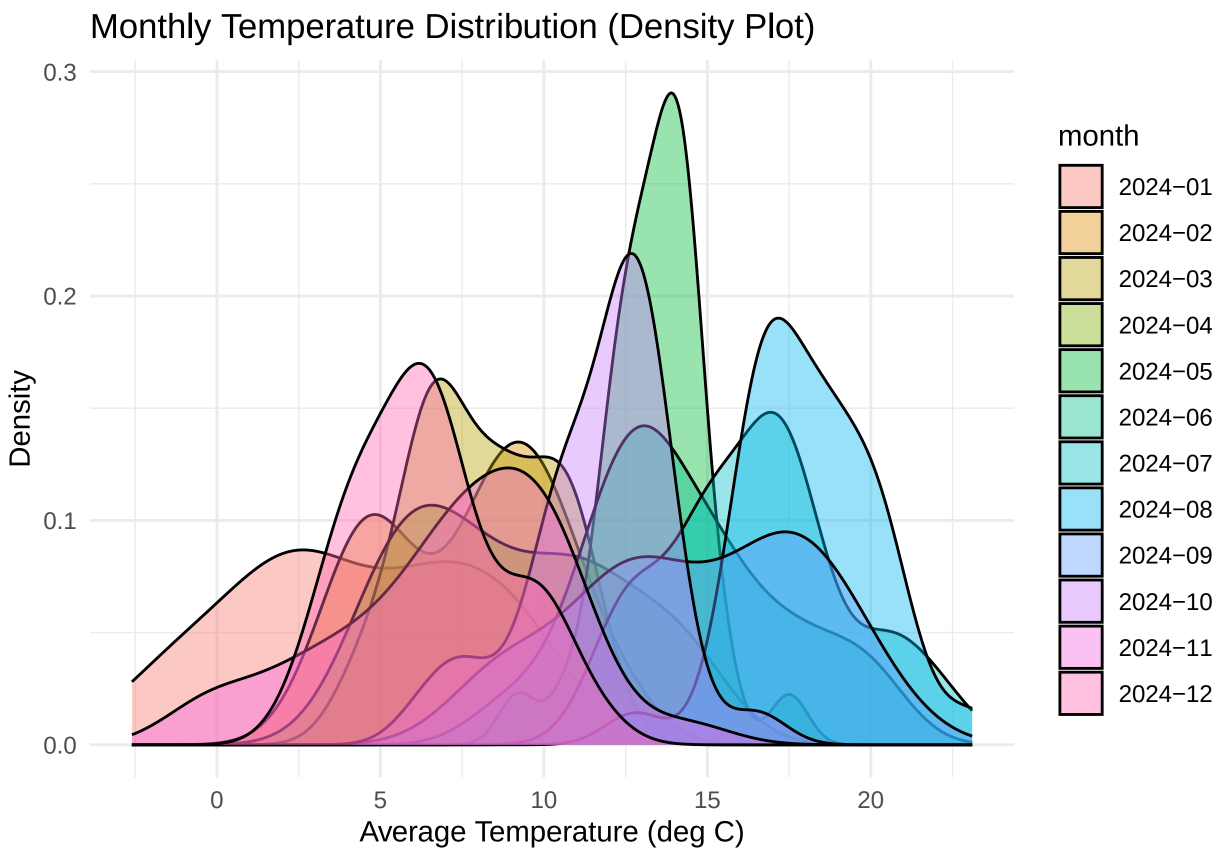
```r
# Histogram of Daily Average Temperature

ggplot(temp_clean, aes(x = TemperatureCAvg)) +
  geom_histogram(binwidth = 1, fill = "steelblue", color = "white") +
  labs(title = "Histogram of Daily Average Temperature",
       x = "Average Temperature (\u00B0C)", y = "Frequency") +
  theme_minimal()
```



```
# Density Plot of Daily Rainfall (Precmm)

ggplot(temp_clean, aes(x = Precmm)) +
  geom_density(fill = "tomato", alpha = 0.6) +
  labs(title = "Density Plot of Daily Rainfall",
       x = "Precipitation (mm)", y = "Density") +
  theme_minimal()
```



```
# Compare Temperature Distribution by Month

# Extract month from date
temp_clean <- temp_clean %>%
  mutate(month = format(date, "%Y-%m"))

# Monthly temperature density plot
ggplot(temp_clean, aes(x = TemperatureCAvg, fill = month)) +
  geom_density(alpha = 0.4) +
  labs(title = "Monthly Temperature Distribution (Density Plot)",
       x = "Average Temperature (deg C)",  # <-- Safe replacement
       y = "Density") +
  theme_minimal()
```



The daily temperature histogram shows a bell-shaped curve centered around 10–15°C, suggesting mild weather is typical in Colchester. Few days fall below 5°C or above 20°C, reflecting a temperate climate.

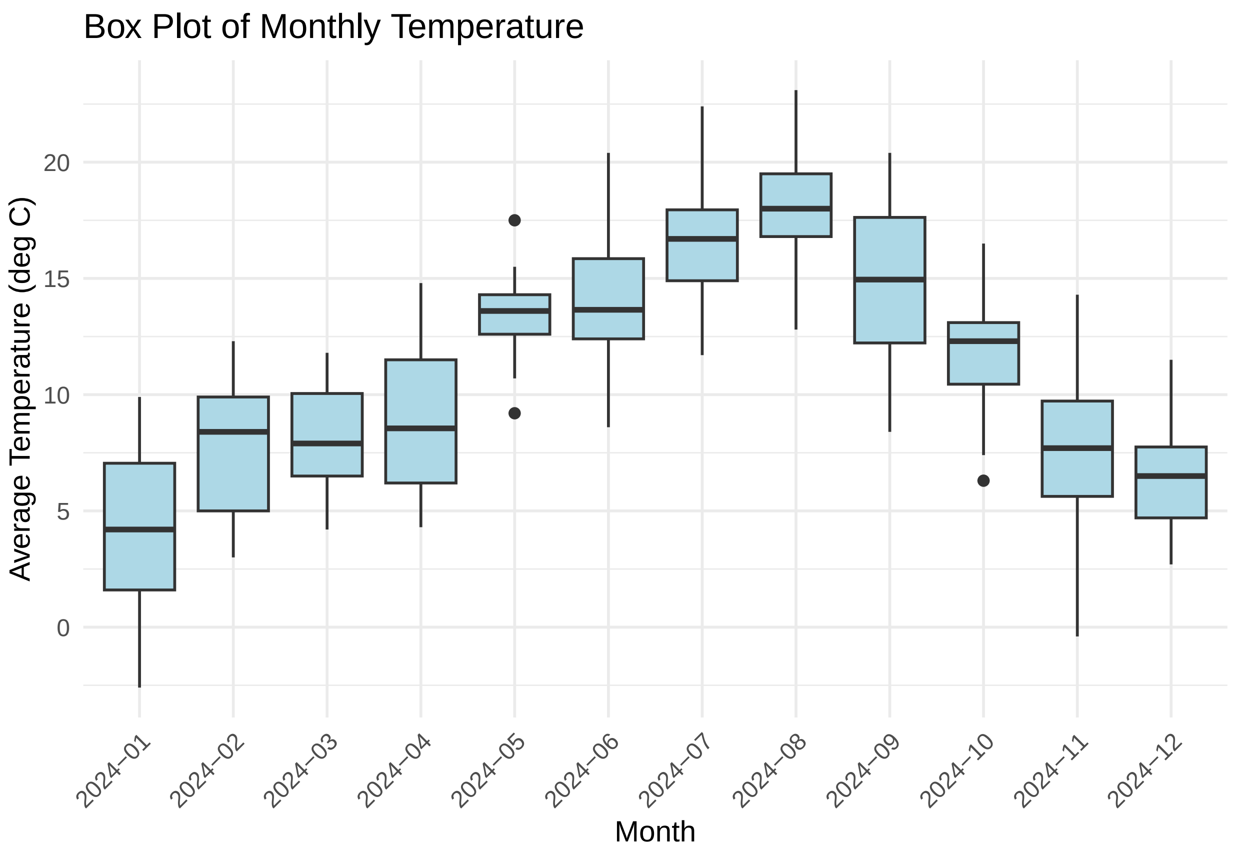
Rainfall density is sharply right-skewed, with most days seeing little to no rain and a few showing heavy rainfall. This pattern highlights the dominance of dry days and occasional extremes.

The monthly temperature density plot reveals seasonal shifts—warmer days cluster in summer, colder ones in early and late months. These trends offer useful context for exploring seasonal links to crime rates.

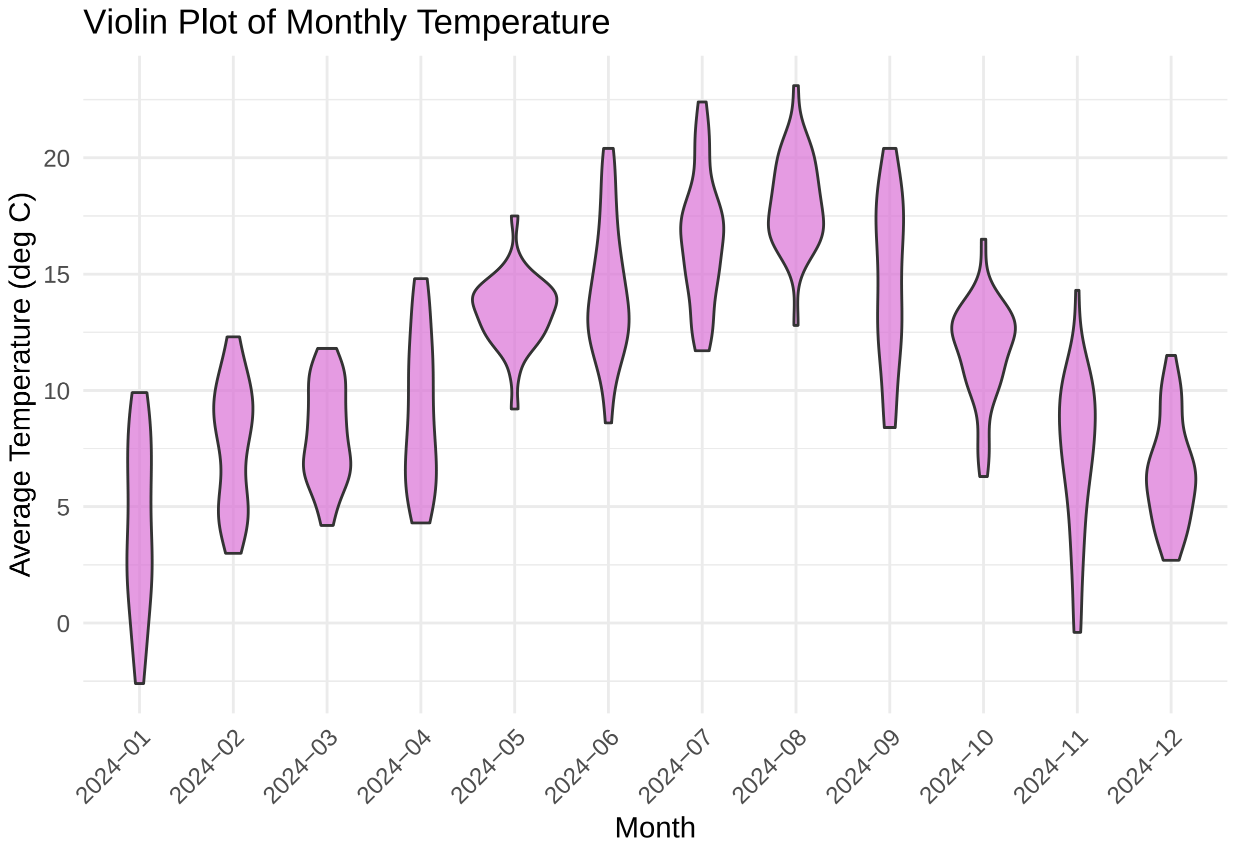
6.4. Box Plot / Violin Plot / Sina Plot

Average daily temperatures across the year were visualized using box plots, violin plots, and sina plots. Each of these visualization methods offers a unique lens on the data, allowing us to explore overall trends, variability, and outliers with greater clarity and depth.

```r
# Box Plot of Temperature by Month

ggplot(temp_clean, aes(x = month, y = TemperatureCAvg)) +
  geom_boxplot(fill = "lightblue") +
  labs(title = "Box Plot of Monthly Temperature",
       x = "Month", y = "Average Temperature (deg C)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



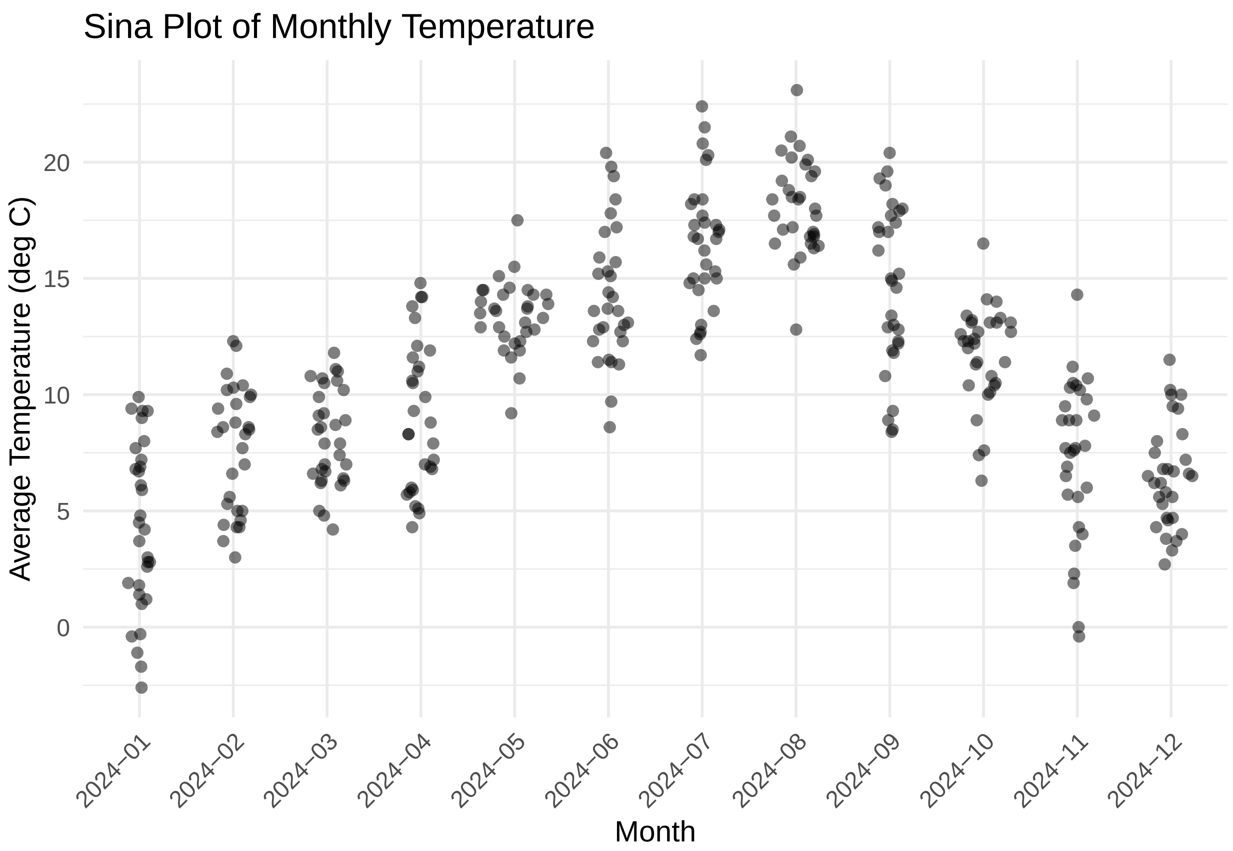
```r
# Violin Plot of Temperature by Month
ggplot(temp_clean, aes(x = month, y = TemperatureCAvg)) +
  geom_violin(fill = "orchid", alpha = 0.7) +
  labs(title = "Violin Plot of Monthly Temperature",
       x = "Month", y = "Average Temperature (deg C)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```
# Sina Plot (if ggforce is installed)
library(ggforce)
```

```
## Warning: package 'ggforce' was built under R version 4.4.3
```

```
ggplot(temp_clean, aes(x = month, y = TemperatureCAvg)) +
  geom_sina(fill = "steelblue", alpha = 0.5) +
  labs(title = "Sina Plot of Monthly Temperature",
       x = "Month", y = "Average Temperature (deg C)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

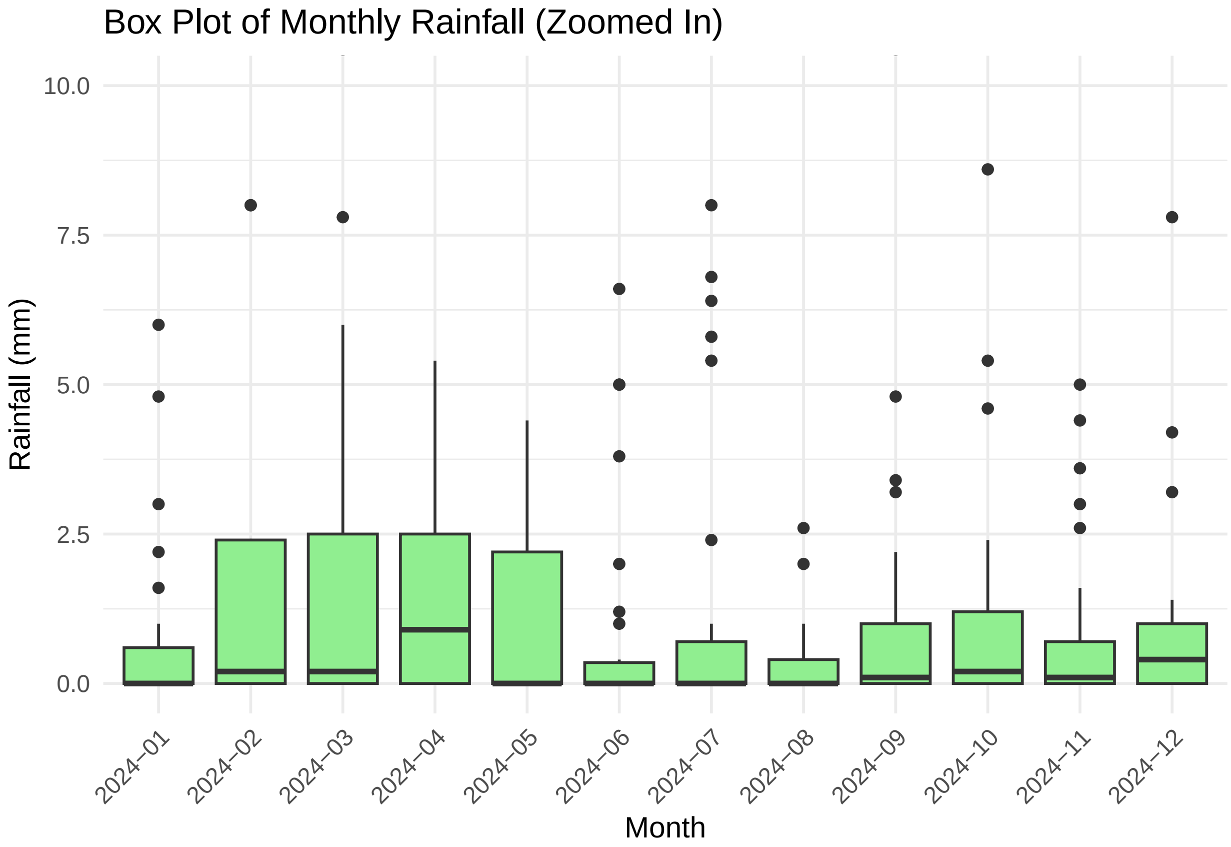


The box plot shows seasonal temperature trends—cold in winter, warm from May to September—with higher variability in April and October. Outliers, especially in spring and autumn, suggest occasional temperature extremes.

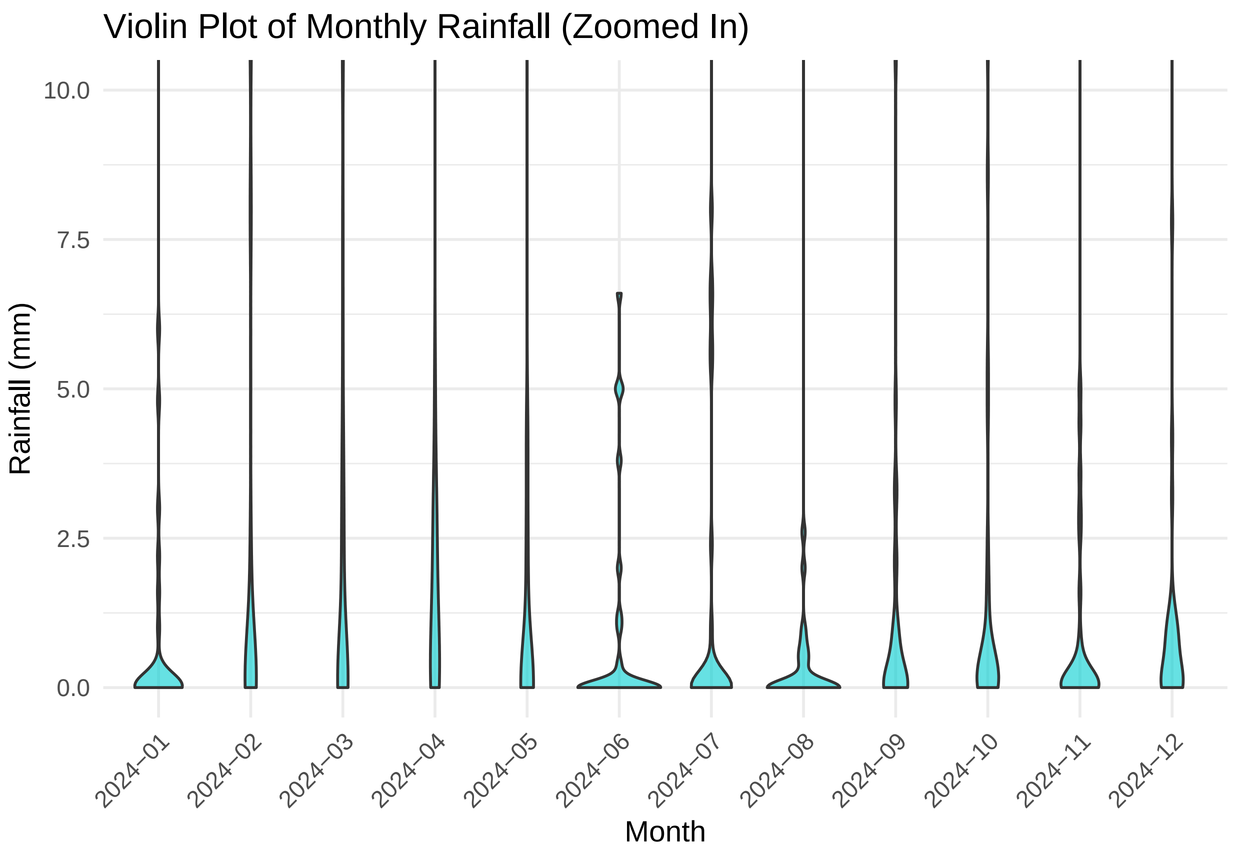
The violin plot expands on this by displaying full distributions. Summer months like July and August show tight clusters around 18–20°C, while winter months have wider spreads, highlighting more variable cold conditions.

The sina plot adds detail by showing each temperature data point. It reveals clustering, gaps, and density patterns across months, providing a granular view that reinforces seasonal trends.

```r
#Box Plot of Rainfall by Month
ggplot(temp_clean, aes(x = month, y = Precmm)) +
  geom_boxplot(fill = "lightgreen") +
  coord_cartesian(ylim = c(0, 10)) +
  labs(title = "Box Plot of Monthly Rainfall (Zoomed In)",
       x = "Month", y = "Rainfall (mm)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```r
#Violin Plot of Rainfall by Month
ggplot(temp_clean, aes(x = month, y = Precmm)) +
  geom_violin(fill = "darkturquoise", alpha = 0.6) +
  coord_cartesian(ylim = c(0, 10)) +
  labs(title = "Violin Plot of Monthly Rainfall (Zoomed In)",
       x = "Month", y = "Rainfall (mm)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



We used box and violin plots to explore monthly rainfall, focusing on the 0–10 mm range with coord_cartesian() to better show common values.

Box plots reveal low median rainfall—often below 2 mm—across months, with greater variability and more outliers from January to May, indicating occasional heavy rain.

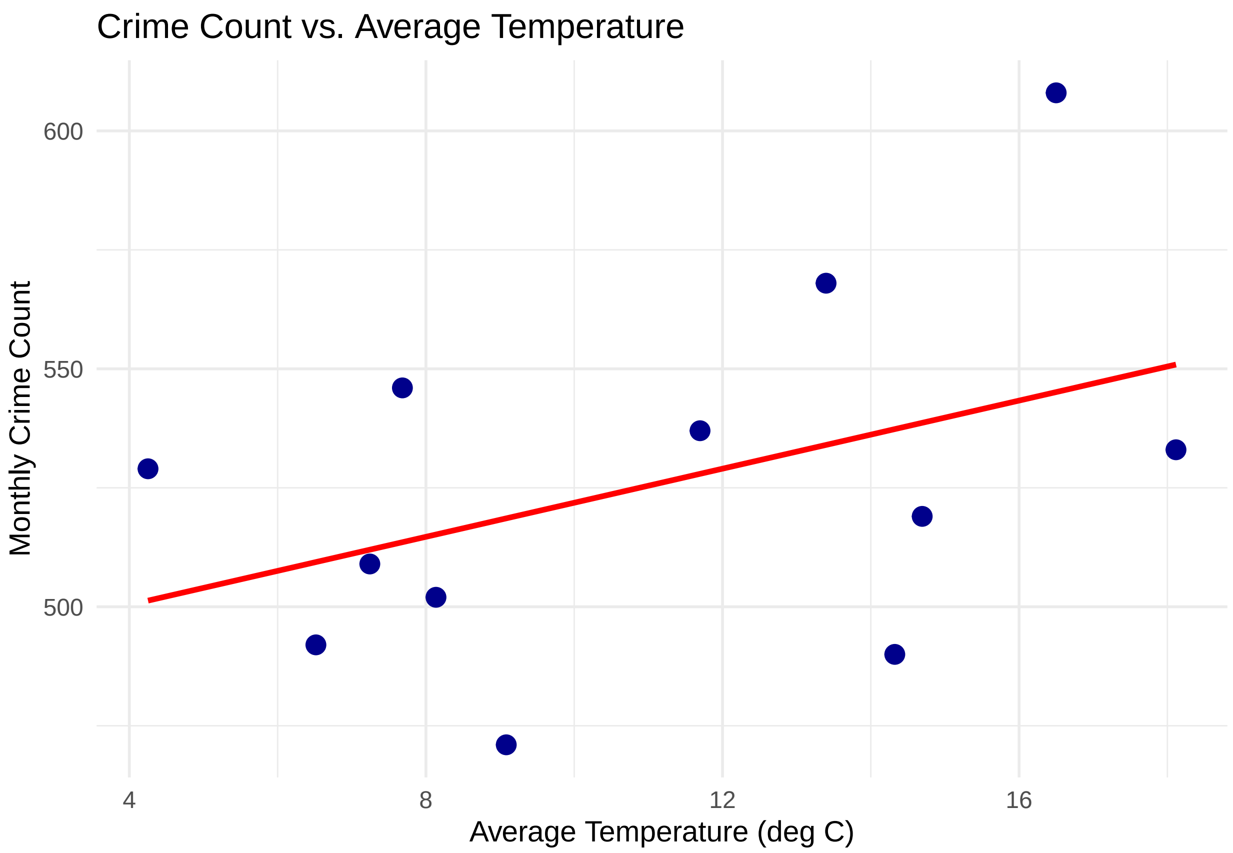
The violin plots show rainfall clustered near zero, but wider shapes in early spring reflect a broader range of outcomes. Narrower violins in summer highlight consistently dry conditions, reinforcing seasonal rainfall patterns.

6.5. Scatter Plot / Pair Plot

We explored the potential relationship between weather conditions and crime levels in Colchester by creating scatter plots and a pair plot using monthly-aggregated data. These visualizations made it easier to spot patterns or correlations that might be missed when relying solely on summary statistics.

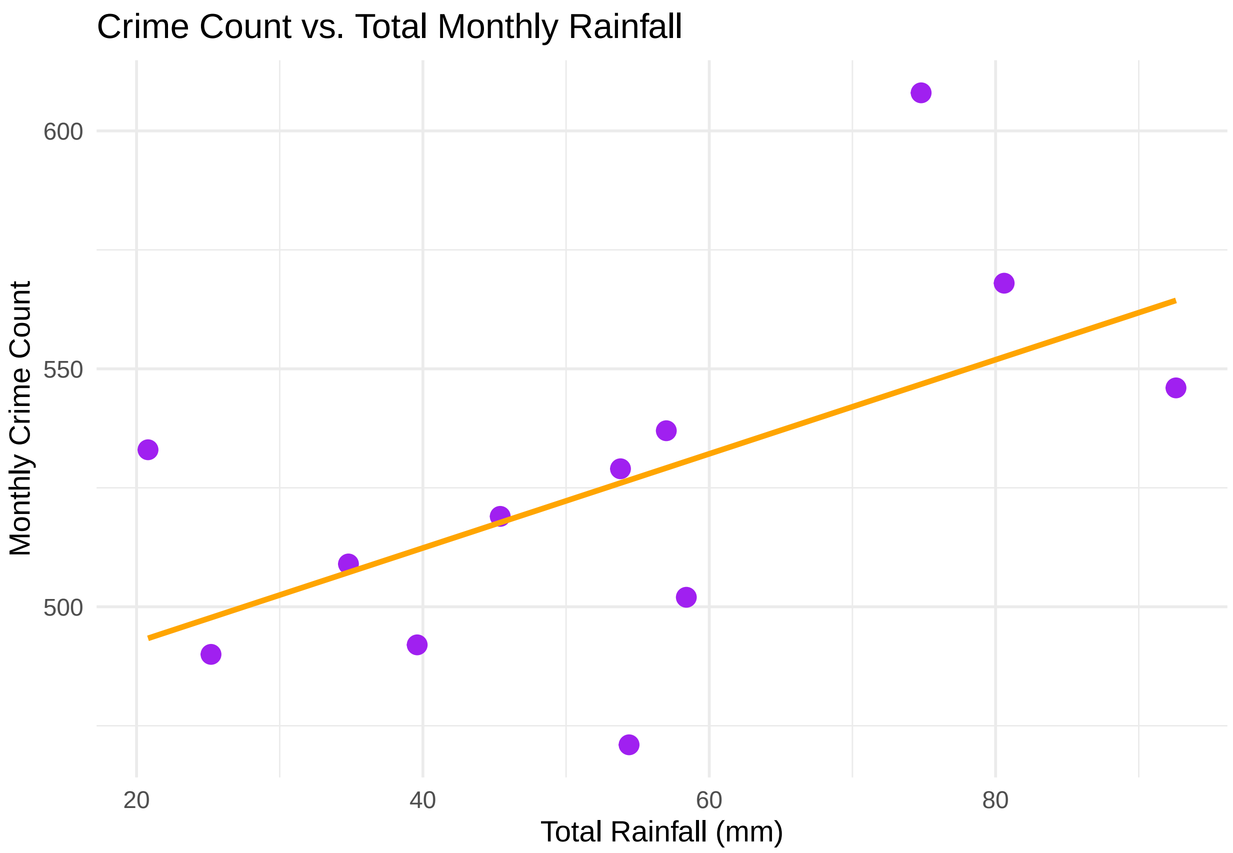
```r
# Scatter Plot: Crime vs. Temperature
ggplot(merged_monthly, aes(x = avg_temp, y = crime_count)) +
  geom_point(color = "darkblue", size = 3) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Crime Count vs. Average Temperature",
       x = "Average Temperature (deg C)", y = "Monthly Crime Count") +
  theme_minimal()
```

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



```r
# Scatter Plot: Crime vs. Rainfall
ggplot(merged_monthly, aes(x = total_rain, y = crime_count)) +
  geom_point(color = "purple", size = 3) +
  geom_smooth(method = "lm", se = FALSE, color = "orange") +
  labs(title = "Crime Count vs. Total Monthly Rainfall",
       x = "Total Rainfall (mm)", y = "Monthly Crime Count") +
  theme_minimal()
```

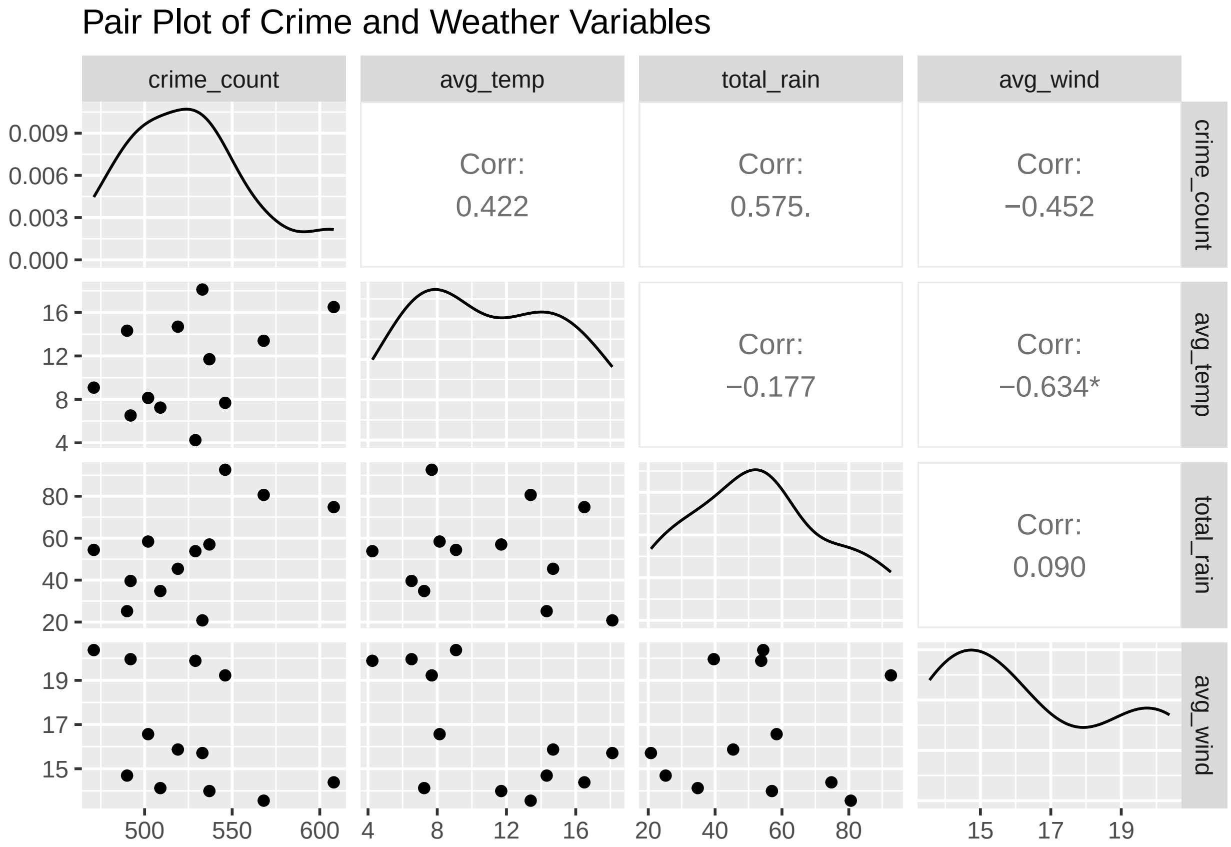
```
## `geom_smooth()` using formula = 'y ~ x'
```



```r
#Pair Plot (Multiple Variables at Once)

library(GGally)
```

```
## Warning: package 'GGally' was built under R version 4.4.3

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```r
# Select relevant columns
ggpairs(merged_monthly[, c("crime_count", "avg_temp", "total_rain", "avg_wind")],
        title = "Pair Plot of Crime and Weather Variables")
```



The scatter plot of crime vs. temperature shows a slight positive trend—crime tends to rise with warmer weather. This aligns with theories linking outdoor activity and social interaction in warmer months to increased crime.

Surprisingly, crime also rises with rainfall, as seen in a second scatter plot. While we'd expect rain to deter crime, this trend may be driven by a few high-rainfall, high-crime months and could reflect underlying factors needing further analysis.

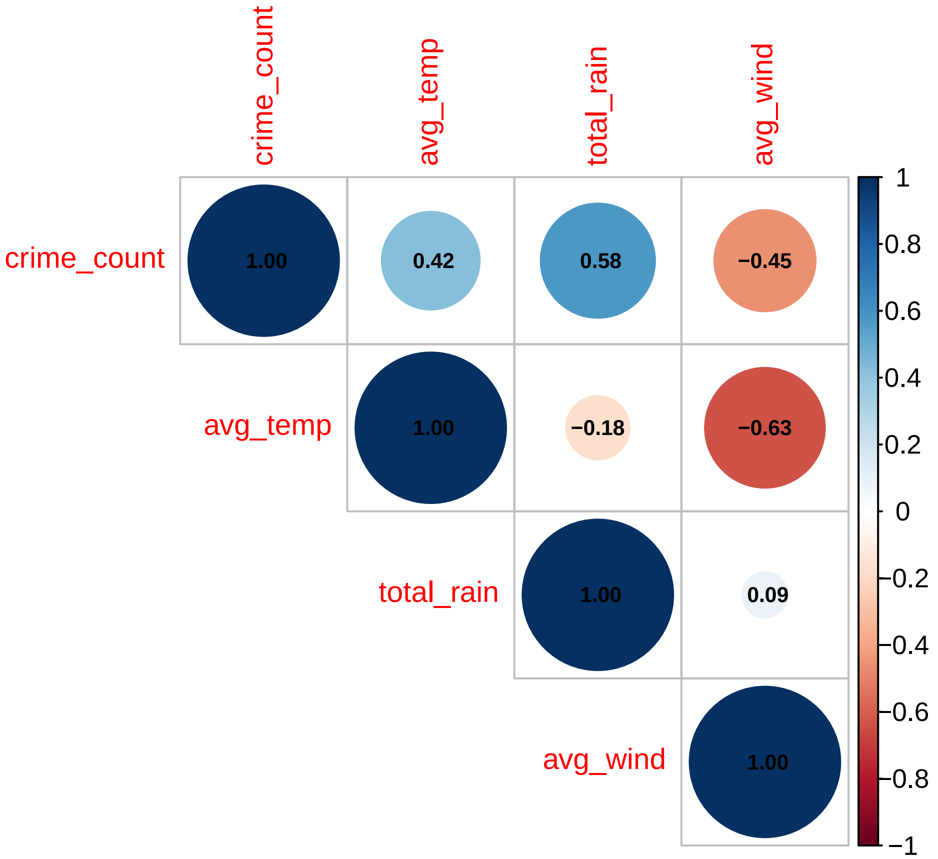
The pair plot provides a comprehensive look at all numerical variables—crime_count, avg_temp, total_rain, and avg_wind. Key takeaways include:

1. A moderate positive correlation between temperature and crime (r = 0.422)
2. A slightly stronger correlation between rainfall and crime (r = 0.575)
3. A negative correlation between crime and wind speed (r = -0.452)
4. The strongest inverse relationship in the matrix is between wind speed and temperature (r = -0.634)

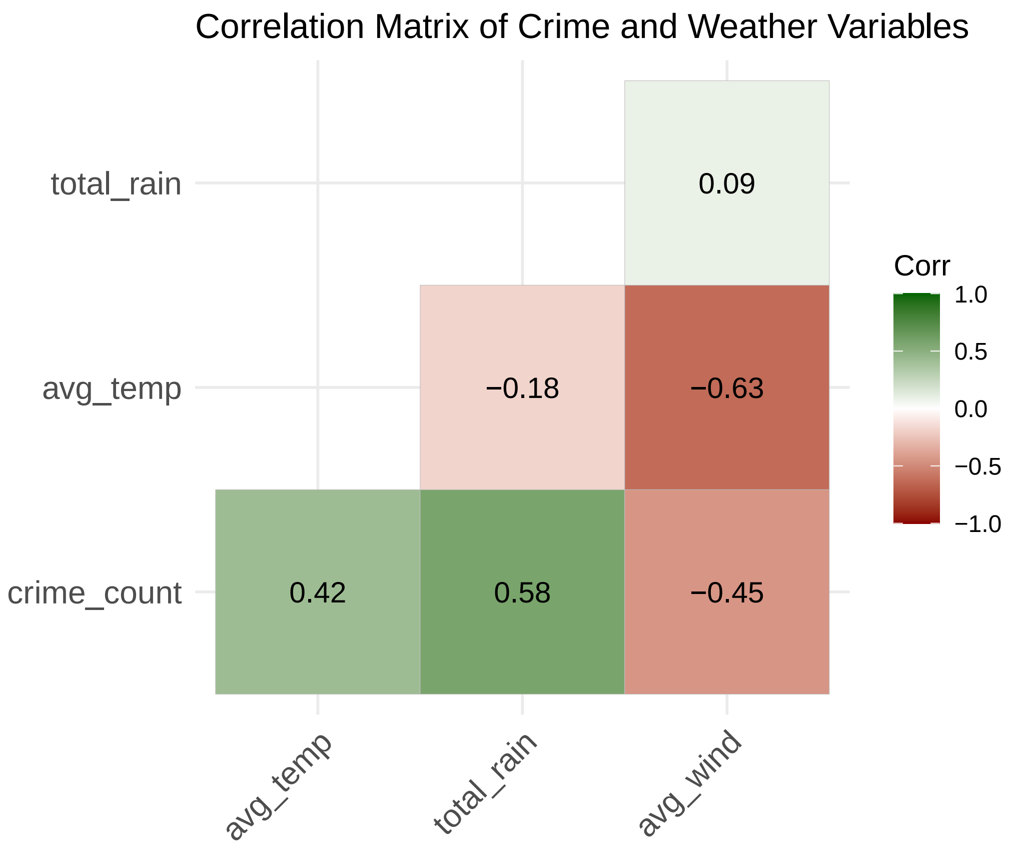
7 Correlation and Trend Analysis

Now, we explore how weather variables—namely temperature, rainfall, and wind speed—correlate with crime levels in Colchester over the year 2024. Using both correlation matrices and time series plots, we aim to identify any consistent patterns or associations that might suggest causality or seasonality.

```r
# Select numeric columns
cor_data <- merged_monthly %>%
  select(where(is.numeric))

# Calculate correlation matrix
cor_matrix <- cor(cor_data, use = "complete.obs")

# Basic circular correlation plot
library(corrplot)

corrplot(cor_matrix, method = "circle", type = "upper",
         tl.cex = 0.9, addCoef.col = "black", number.cex = 0.7)
```



```r
library(ggcorrplot)

ggcorrplot(cor_matrix, lab = TRUE, type = "lower",
           colors = c("darkred", "white", "darkgreen"),
           title = "Correlation Matrix of Crime and Weather Variables")
```



Correlation analysis using corrplot and ggcorrplot shows a moderate positive link between temperature and crime (r = 0.42), supporting the idea that warmer weather may encourage activity—and thus crime.

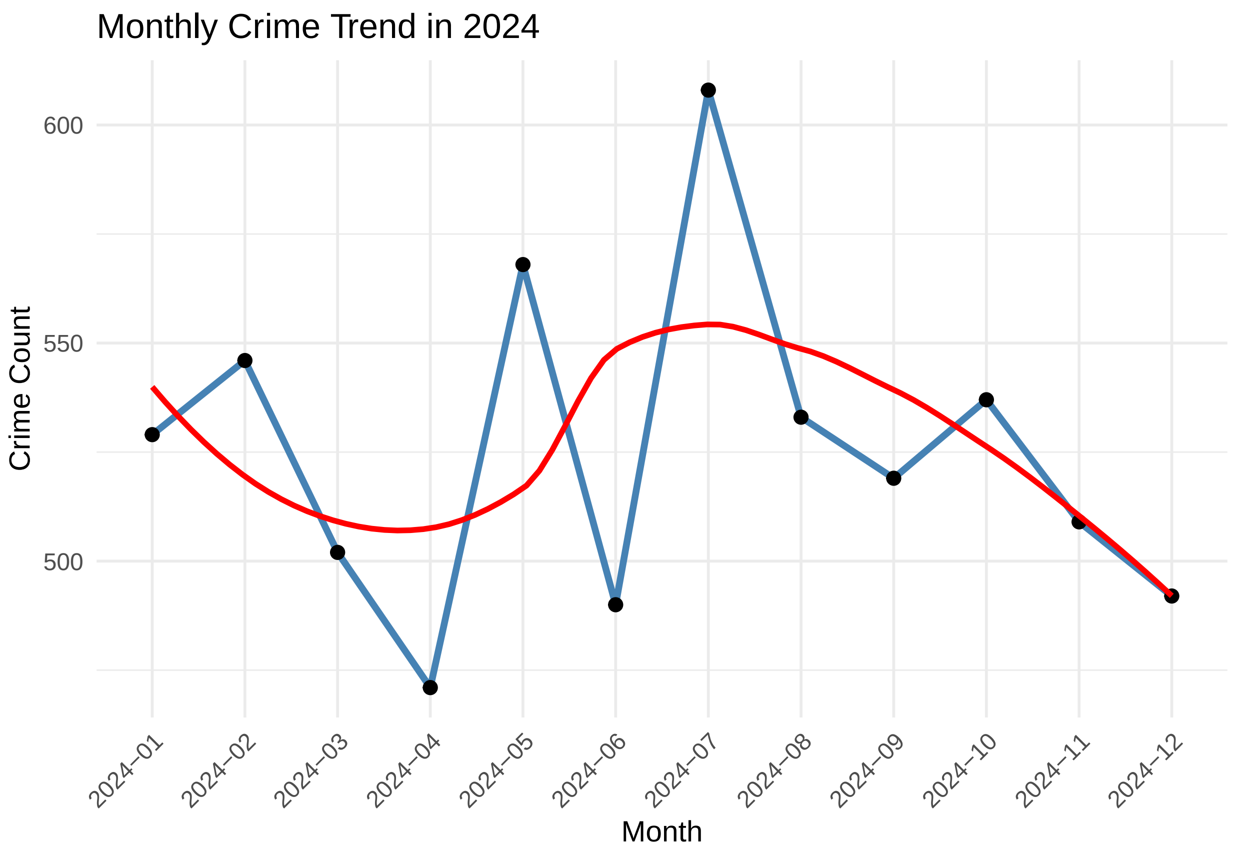
Rainfall also correlates positively with crime (r = 0.58), which may reflect specific periods of high activity despite rain. Wind speed, however, shows a moderate negative correlation (r = -0.45), suggesting it might deter crime by limiting outdoor movement.

Time Series Plot with Smoothing

The time series analysis, enhanced with smoothing techniques, provides a clear view of how crime counts and weather variables evolve month by month throughout 2024.

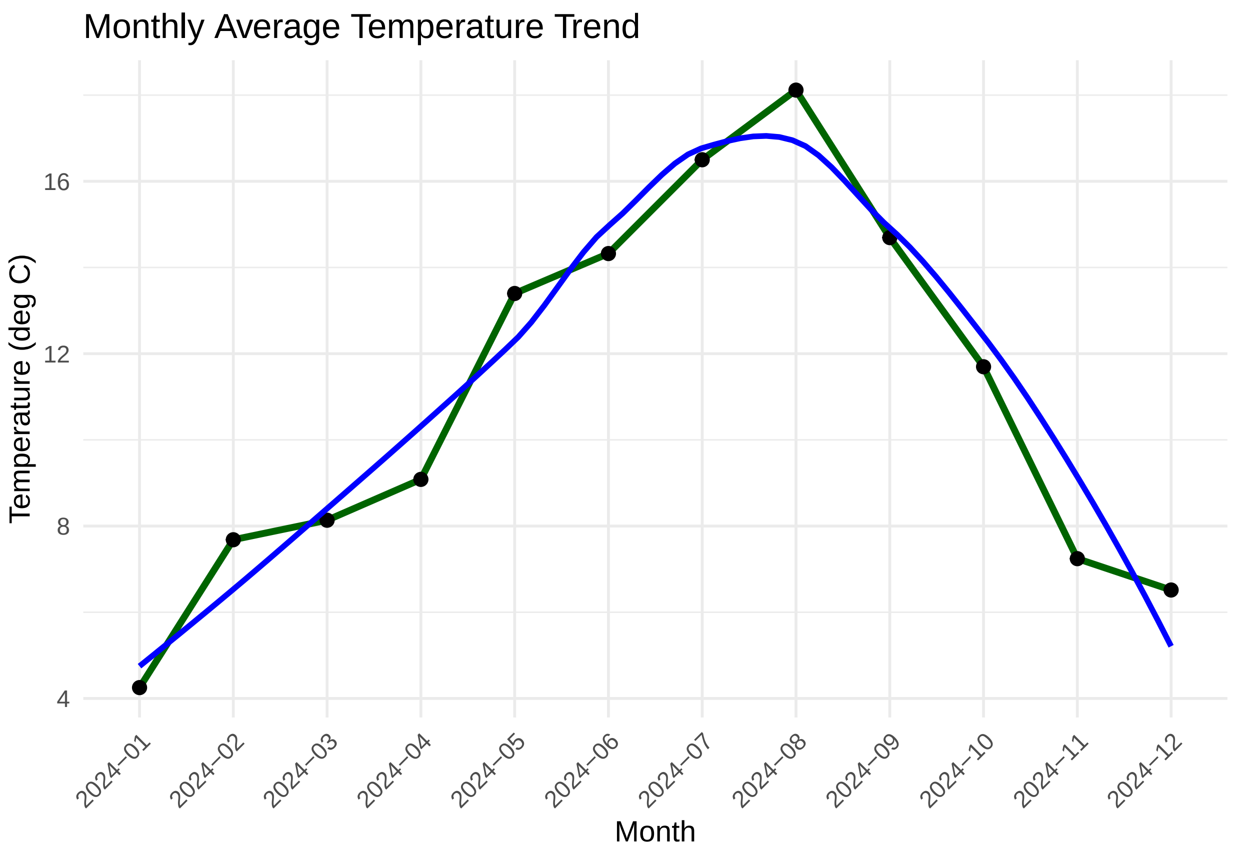
```r
# Counting Crime Over Time
ggplot(merged_monthly, aes(x = month, y = crime_count, group = 1)) +
  geom_line(color = "steelblue", linewidth = 1.2) +
  geom_point(color = "black", size = 2) +
  geom_smooth(se = FALSE, color = "red", method = "loess") +
  labs(title = "Monthly Crime Trend in 2024",
       x = "Month", y = "Crime Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

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



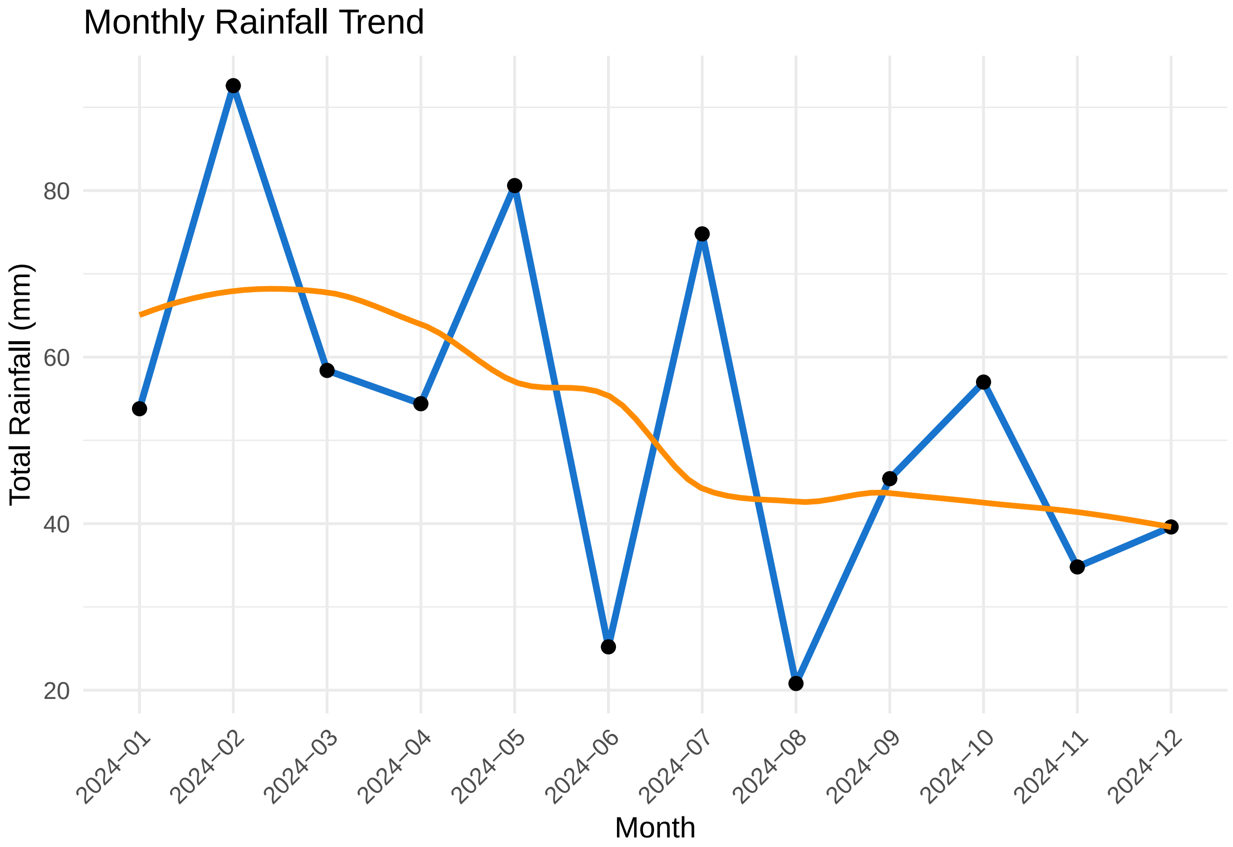
```r
# Average Temperature Over Time
ggplot(merged_monthly, aes(x = month, y = avg_temp, group = 1)) +
  geom_line(color = "darkgreen", linewidth = 1.2) +
  geom_point(color = "black", size = 2) +
  geom_smooth(se = FALSE, color = "blue", method = "loess") +
  labs(title = "Monthly Average Temperature Trend",
       x = "Month", y = "Temperature (deg C)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

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



```r
# Rainfall Over Time
ggplot(merged_monthly, aes(x = month, y = total_rain, group = 1)) +
  geom_line(color = "dodgerblue3", linewidth = 1.2) +
  geom_point(color = "black", size = 2) +
  geom_smooth(se = FALSE, color = "darkorange", method = "loess") +
  labs(title = "Monthly Rainfall Trend",
       x = "Month", y = "Total Rainfall (mm)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

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



Time series plots show clear seasonal trends in crime and weather. Crime peaks in July and September, closely mirroring rising temperatures, which peak in August—supporting the idea that warmer weather may elevate crime due to increased activity.

Rainfall peaks in February and May but doesn't appear to impact crime levels significantly, suggesting its influence may be minimal or time-dependent.

Temperature follows a smooth seasonal curve, offering context for comparing crime patterns. These trends highlight the value of weather data in anticipating crime surges and improving resource planning for public safety.

8 Spatial Visualization

We used the leaflet package to build an interactive map based on geographic coordinates from the crime dataset, offering a clearer view of how crime is distributed across Colchester. To make the analysis more engaging and easier to explore, two of the most relevant visualizations were also converted into interactive formats using Plotly.

8.1: Interactive Crime Map with Leaflet

Each red dot represents an individual crime location, plotted using latitude and longitude. Users can hover to view crime type and outcome status, offering detailed, location-specific insights.

This geospatial visualization reveals clear clustering of criminal activity in the town center, especially around the High Street and surrounding urban areas. Outskirts such as the southern residential zones show far fewer incidents. Such clustering supports targeted policing and resource deployment by highlighting high-crime zones visually and interactively.

```r
# Load the leaflet map using cleaned crime data
leaflet(data = crime_clean) %>%
  addTiles() %>%  # Adding default OpenStreetMap tiles
  addCircleMarkers(
    lng = ~long,  # Adding Longitude from the dataset
    lat = ~lat,   # Adding Latitude from the dataset
    popup = ~paste("Crime:", category, "<br>",  # Popup info: crime category
                   "Outcome:", outcome_status), # and outcome status
    radius = 2,           # Small circle markers for clarity
    color = "red",        # Red color for markers
    fillOpacity = 0.7     # Adding slight transparency to reduce overlap visibility
  ) %>%
  addScaleBar(position = "bottomleft") %>%  # Adding a scale bar to bottom-left
  setView(                                   # Center the map view
    lng = mean(crime_clean$long, na.rm = TRUE),  # Mean longitude
    lat = mean(crime_clean$lat, na.rm = TRUE),   # Mean latitude
    zoom = 12                                # Default zoom level
  )
```

```r
# The above map can look cluttered when many crimes are located close together.
# Let's now visualize the same map using **marker clustering** to handle overlap.

leaflet(data = crime_clean) %>%
  addTiles() %>%  # Adding default OpenStreetMap tiles
  addMarkers(
    lng = ~long,  # Longitude from the dataset
    lat = ~lat,   # Latitude from the dataset
    clusterOptions = markerClusterOptions(),  # Enabling automatic clustering
    popup = ~paste("Crime:", category, "<br>",  # Popup info
                   "Outcome:", outcome_status)
  ) %>%
  setView(
    lng = mean(crime_clean$long, na.rm = TRUE),
    lat = mean(crime_clean$lat, na.rm = TRUE),
    zoom = 12
  )
```

8.2: Interactive Plot with Plotly

The first is a scatter plot showing the relationship between crime count and average temperature. In its interactive form, users can hover over individual points to view exact values, zoom into specific temperature ranges, and visually explore the positive correlation between warmer conditions and higher crime rates.

```r
# Interactive Scatter Plot: Crime vs. Temperature

library(plotly)

# Creating plotly object from ggplot
p1 <- ggplot(merged_monthly, aes(x = avg_temp, y = crime_count)) +
  geom_point(color = "darkred", size = 3) +
  geom_smooth(method = "lm", se = FALSE, color = "black") +
  labs(title = "Interactive: Crime Count vs. Avg Temperature",
       x = "Average Temperature (deg C)", y = "Crime Count") +
  theme_minimal()

# Converting to plotly interactive plot
ggplotly(p1)
```

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

```
# Also, interactive Line Plot of Crime Trend

p2 <- ggplot(merged_monthly, aes(x = month, y = crime_count, group = 1)) +
  geom_line(color = "steelblue", linewidth = 1.2) +
  geom_point(color = "black", size = 2) +
  labs(title = "Interactive: Monthly Crime Trend", x = "Month", y = "Crime Count") +
  theme_minimal()

ggplotly(p2)
```

The second plot is a time series of monthly crime counts for 2024. This interactive line graph allows users to trace seasonal fluctuations, clearly highlighting July as the peak month—corresponding with the highest temperatures.

9. Insights and Interpretation

Through this analysis of Colchester's 2024 crime and weather data, several patterns emerged that offer valuable insights into the relationship between environmental conditions, seasonal trends, and public safety. By linking crime frequency with monthly weather patterns—particularly average temperature, rainfall, and wind speed—and visualizing outcomes geographically and interactively, we have gained a clearer understanding of how external factors may influence crime rates and their spatial distribution.

9.1 Does Crime Increase with Temperature?

The data strongly suggests that crime in Colchester increases with warmer weather. This trend is supported both visually and statistically. From the correlation matrix, average temperature and crime count share a moderate positive correlation of +0.42, meaning that as temperatures rise, so do incidents of crime .

Looking at the monthly breakdown, May recorded the highest number of crimes at 568 incidents, with an average temperature of 13.4°C, followed closely by July (608 incidents) and August (533 incidents), where temperatures reached 16.5°C and 18.1°C respectively . These findings are consistent with established criminological theories that associate warm weather with increased outdoor activity, social interaction, and hence, more opportunities for conflict or opportunistic crimes such as theft.

Conversely, cooler months like January (529 crimes, 4.2°C) and February (546 crimes, 7.7°C) reported comparatively lower crime rates. While the difference isn't drastic in all cases, the upward slope in the time series plot visually confirms a seasonal rise and fall that aligns with temperature changes .

9.2 Some Crimes are Seasonal

Yes—specific crime categories exhibit clear seasonal trends. Based on frequency tables and visual outputs, anti-social behaviour, violent crimes, and bicycle theft were all more common during the spring and summer months, especially from May to August.

For example:

1. Anti-social behaviour and public order offences peaked in July, coinciding with Colchester's warmest months.
2. Bicycle thefts also rose notably during summer, likely due to increased cycling activity in fair weather. A closer inspection of the bar plots shows these categories surging just as temperatures hit their annual highs.

On the other hand, burglary, vehicle crime, and criminal damage showed a more uniform distribution across the year, suggesting these crimes are less dependent on seasonal variables and more influenced by other factors such as opportunity, socioeconomic conditions, or routine household activity.

These insights support the hypothesis that weather-sensitive crimes—those that occur more often in public spaces or require public presence—peak during warmer, more sociable months.

9.3 What Can We Learn from Crime Locations?

Geospatial mapping using Leaflet revealed distinct crime clusters within Colchester, especially concentrated in the town centre, including areas near the High Street and Hythe. These hotspots were particularly linked to crimes like shoplifting, anti-social behaviour, and public order offences—activities that are more likely to occur in high-footfall commercial and leisure zones .

In contrast, residential and peripheral areas exhibited fewer total crimes but were relatively more prone to burglary and property damage, which tend to occur in quieter neighborhoods with less surveillance and foot traffic.

The spatial analysis thus highlights how environmental context shapes crime patterns:

1. Urban hubs attract people-based crimes
2. Outskirts are more vulnerable to property-based offences

These findings reinforce the importance of context-aware policing, where foot patrols and preventive measures are adapted based on local population dynamics and urban design.

9.4 What Role Did Rainfall and Wind Play?

Rainfall, somewhat unexpectedly, also showed a positive correlation with crime (r = +0.58)—even stronger than the correlation with temperature. This is contrary to the common assumption that bad weather discourages outdoor movement and thereby reduces crime . One explanation might be that Colchester experienced high crime during certain high-rainfall months like February (92.6 mm rainfall) and May (80.6 mm rainfall), suggesting that specific incidents or crime types may not be deterred by precipitation, or that rain coincided with other social events or holidays.

Wind speed showed the opposite effect: a negative correlation with crime (r = -0.45). Windier months such as January and April coincided with lower crime counts, which may reflect the discomfort or visibility disruption that discourages outdoor or opportunistic criminal behavior.

These weather patterns demonstrate that while temperature is a strong driver, rainfall and wind can also influence public behavior and safety, albeit in more complex or situational ways.

9.5 What Did the Interactive Plots Reveal?

The use of plotly added depth to the visual storytelling. The interactive scatter plot between crime and temperature allowed users to hover over data points and see precise values, which helped illustrate how months with 14°C+ temperatures consistently saw over 500 crimes.

The interactive time series plot further emphasized these peaks, with July and May standing out as the highest crime months. The dynamic view made seasonal crime trends immediately apparent to both technical and non-technical audiences, and could be a valuable tool for public presentations or community engagement.

9.6 Why Do These Insights Matter?

The findings have practical implications for crime prevention and resource allocation in Colchester:

1. Seasonal Preparedness: Since crime rises in warmer months, police presence and community outreach could be increased between May and August. Resources like mobile units or outreach vans can be deployed more heavily during this period.

2. Targeted Awareness Campaigns: With bicycle theft peaking in summer, local councils could run awareness campaigns promoting bike locks and parking safety between June and September.

3. Area-Specific Policing: Knowing that the town centre is a hotspot for shoplifting, while residential outskirts face burglary, police could adjust patrol routes and install surveillance in risk-prone zones accordingly.

4. Urban Planning and Lighting: In areas where property crimes are frequent, improved street lighting, CCTV coverage, and neighbourhood watch programs could deter criminal activity.

5. Use of Weather Forecasts: Integrating weather data into policing systems could enable predictive patrol scheduling, especially when hot, dry weather is expected.

10. Conclusion

We explored the relationship between crime patterns and weather conditions in Colchester throughout 2024. By combining street-level crime data with daily meteorological records and applying a variety of data visualisation techniques, several key insights emerged.

10.1 The Crime situation in Colchester

The analysis revealed that crime in Colchester generally increased during the warmer months of the year. Notably, May and July stood out with significantly higher crime counts, supporting the idea that pleasant weather leads to more outdoor activity—and possibly more opportunities for crime to occur.

Certain types of crimes appeared to follow a clear seasonal pattern. Anti-social behaviour and bicycle theft were particularly common during the summer months. This likely reflects increased social interaction, outdoor gatherings, and the higher use of bicycles during this time, which makes them easier targets for theft.

Geographically, crime was not evenly spread across the town. It tended to cluster in specific locations, with central Colchester emerging as a hotspot. This area experienced more public-facing crimes such as shoplifting, possibly due to its busier streets, higher foot traffic, and concentration of retail businesses.

Interestingly, weather also seemed to influence crime levels in another way—rainfall may have a dampening effect on criminal activity. In the colder months, when rain was more frequent, overall crime appeared to decrease. This could be because fewer people were outside, reducing the likelihood of crimes happening in public spaces.

One unexpected finding was the sharp dip in crime during April, despite moderate temperatures. This anomaly may reflect external influences (e.g., school holidays, public events, or targeted policing campaigns), pointing to the importance of considering socio-political events alongside environmental data.

Also surprising was that burglary and criminal damage did not show strong seasonal shifts, suggesting they are driven by opportunity rather than weather or public activity levels.

10.2 Possible Follow-Up Analysis

To build on the findings from this project, several directions could be explored in future research for deeper and more precise insights.

1. Use more advanced models (like regression) to understand which weather factors matter most.
2. Study what times of day crimes happen to see if there's a daily pattern.
3. Add demographic data to check if crime levels are related to things like income or housing.
4. Use social media or news to see how people feel about safety and compare that with real crime numbers.

11. References

1. https://ukpolice.njtierney.com/reference/ukp_crime.html
2. https://bczernecki.github.io/climate/reference/meteo_ogimet.html
3. tidyverse, leaflet, ggplot2, ggcorrplot, etc.