Crime & Climate Data Analysis

Author : Nagadithya Bathala

Introduction

Now a days crime rate is increasing in many parts of the world. Analyzing the crimes based on the data available gives insights into crime rate, location where crime rate is more and types of crimes that took place in specific street or location. In many occasions crime rate depends on the temperature and climate conditions. Analyzing these data and finding the relationship between crime patterns and climate conditions will have impact on the society and environment. It also helps in providing the public safety.

Objective

The objective of this data analysis report is to explore the crime and temperature dataset and find insights in the data. Different exploratory data analysis(EDA) techniques were used to find the relationship between the variables. The aim is find out the crime patterns, trends and the correlation between the variables. Additionally, the analysis was performed to check if there is any relationship exists between the crime rate and different temperature variables. This report also includes exploring if change in climate conditions influence the crime rate and the frequency of different crime categories.

Exploratory Data Analysis

Loding required libraries

```
library(dplyr)
library(plotly)
library(ggplot2)
library(lubridate)
library(xts)
library(forecast)
library(kableExtra)
library(leaflet)
library(ggcorrplot)
library(reshape2)
```

These are all the libraries that were used to perform exploratory data analysis on the crime and temperature datasets.

Importing Crime dataset

```
#Reading dataset
crime_df<-read.csv("crime23.csv", header=T)
#Dimension of the dataset
dim(crime_df)
```

```
## [1] 6878    12
```

```
#Column names
names(crime_df)
```

```
##  [1] "category"        "persistent_id"   "date"            "lat"
##  [5] "long"            "street_id"       "street_name"     "context"
##  [9] "id"              "location_type"   "location_subtype" "outcome_status"
```

```
#Head of the dataset
head(crime_df)
```

	category <chr>	persistent_id <chr>	date <chr>	lat <dbl>	long <dbl>	street_id <int>
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

6 rows | 1-7 of 13 columns

The crime dataset was imported into the r work space using readcsv function. The dimensions of the dataset is checked and it has 12 columns and 6878 rows of data. The features of the dataset were explored and it has information about crime category, date, street name and geo-spacial coordinates of the each crime.

Importing Climate dataset

```
#Reading the dataset
climate_df<-read.csv("temp2023.csv")
#Dimension of the dataset
dim(climate_df)
```

```
## [1] 365  18
```

```
#Column names
names(climate_df)
```

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

```
#Head of the dataset
head(climate_df)
```

	station_ID <int>	Date <chr>	TemperatureCAvg <dbl>	TemperatureCMax <dbl>	TemperatureCMin <dbl>	TdA... <dbl>	Hr... <dbl>	Windkmh... <chr>
1	3590	2023-12-31	8.7	10.6	4.4	7.2	89.6	S
2	3590	2023-12-30	6.6	9.7	4.4	4.2	85.5	WSW
3	3590	2023-12-29	9.9	11.4	6.9	6.0	77.2	SW
4	3590	2023-12-28	9.9	11.5	4.0	7.5	84.6	SSW
5	3590	2023-12-27	5.8	10.6	3.9	3.7	86.4	SW
6	3590	2023-12-26	9.8	12.7	6.3	7.6	86.9	WSW

6 rows | 1-9 of 19 columns

The temperature dataset was imported into the r work space using readcsv function. The dimensions of the dataset is checked and it has 18 columns and 365 rows of data. The features of the dataset were explored and it has information about temperature, humidity, wind speed, sea level pressure and total cloudiness of each day of the year 2023 in Colchester.

```
#Checking for null values in crime dataset
sum(is.na(crime_df))
```

```
## [1] 7555
```

```
#Checking for null values in temperature dataset
sum(is.na(climate_df))
```

```
## [1] 851
```

Deleting unwanted columns

```
#Deleting columns
crime_df <- select(crime_df, -persistent_id, -context, -location_subtype, -id)
climate_df <- select(climate_df, -station_ID, -PreselevHp, -SnowDepcm)
```

The crime and temperature datasets were explored and it is found that some columns has mostly null values. These columns were deleted as they are not useful for analysis. The datasets has features such as persistent id, id and station id which are not useful for analysis. So, these features were removed from the dataset.

Replacing NA values in the outcome_status column

```
crime_df$outcome_status[is.na(crime_df$outcome_status)] <- "Information Not Available"
```

The crime data set has a feature outcome status. It represents the investigation status of each crime that took place in Colchester in 2023. It has 677 null values. To use this variable in the analysis, the null values were replaced with 'Information Not Available'.

Table

```
table_ <- round( 100*prop.table(table(crime_df$location_type)),2)
print('Table - Percentage of crimes in each location type')
```

```
## [1] "Table - Percentage of crimes in each location type"
```

```
table_
```

```
##
##    BTP Force
##   0.35 99.65
```

The crime dataset contains a location type variable which represents the type of the location in which the crimes took place. The location type is categorized into two types i.e., Force and BTP location types. Force location type indicates a normal police force location and BTP location type indicates a British Transport Police location. Here the number of crimes and type of crimes occurred in both the location types were analyzed. From the table above it is observed that most of the crimes occurred in force location type.

Two Way Table to visualize crimes in each category by month

```
#Converting date variable to year month format
crime_df_ <- crime_df
crime_df_$date <- ym(crime_df_$date)

#Ordering data based on date
crime_df_$date <- format(crime_df_$date, "%b")
crime_df_$date <- factor(crime_df_$date, levels = c(as.list(month.abb)))

#Two way table
two_way_table_ <- table(crime_df_$category, crime_df_$date)
print('Two way table - Number of crimes in each crime category by month in year 2023 in Colchester')
```

```
## [1] "Two way table - Number of crimes in each crime category by month in year 2023 in Colchester"
```

```
kable(two_way_table_)
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
anti-social-behaviour	46	49	21	53	67	52	76	71	90	68	39	45
bicycle-theft	20	14	19	16	16	14	15	21	37	26	27	10
burglary	17	22	14	22	15	26	14	20	18	31	11	15
criminal-damage-arson	59	37	52	63	64	42	42	33	47	45	53	44
drugs	14	17	21	21	22	15	17	7	25	19	13	17
other-crime	7	5	6	15	3	11	12	9	7	6	5	6
other-theft	48	37	35	38	42	41	51	41	34	49	37	38
possession-of-weapons	3	3	11	5	7	3	8	5	8	6	8	7
public-order	45	42	58	51	37	36	40	41	45	52	45	40
robbery	8	7	8	7	7	17	6	5	8	9	5	7
shoplifting	76	31	51	40	51	59	33	57	33	43	39	41
theft-from-the-person	6	7	12	7	5	6	9	5	7	3	4	5
vehicle-crime	65	15	21	29	24	45	25	16	20	26	56	64
violent-crime	237	181	226	207	226	196	236	219	263	209	221	212

The dataset shows that different types of crimes took place in Colchester in 2023. Here two way table is used to see number of crimes happened in each month in 2023 by each category. The above two shows information different types of crimes that occurred in each month and also the frequency of those crimes.

Visualizing crime categories using Bar plot

```
#Sorting the categories by number of crimes
crime_df_copy <- crime_df
crime_df_copy$category <- factor(crime_df$category, levels = names(sort(table(crime_df$category), decreasi
ng = TRUE)))

#Creating bar plot object using ggplot
crimes_gg_bar_plot <- ggplot(crime_df_copy ,aes(x=category)) +
  geom_bar(fill=heat.colors(n=length(unique(crime_df$category))), alpha = 0.9) +
  labs(x="Crime Type", y="Number of Crimes") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

#Interactive bar plot
crimes.plotly_bar <- ggplotly(crimes_gg_bar_plot, height=450, width=800) %>% layout( hoverlabel = list(bgc
olor = "cyan", font = list(color = "black")))

#Customizing the title with layout
crimes.plotly_bar <- layout(crimes.plotly_bar, title = list(text = "Number of Crimes in Colchester by Cate
gory in 2023", x = 0.5, y = 0.99))

#Plotting the graph
crimes.plotly_bar
```
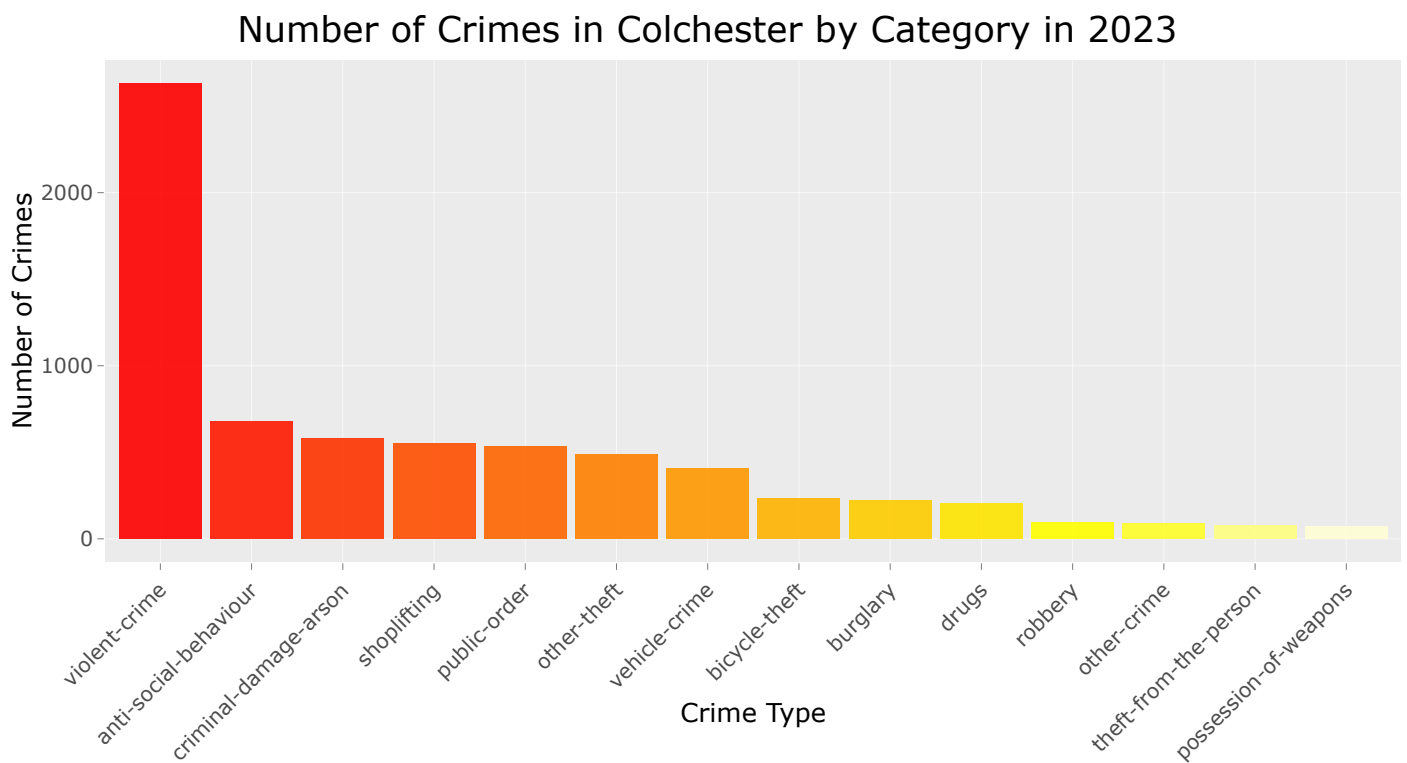
Number of Crimes in Colchester by Category in 2023

There are different types of crimes that took place in Colchester in 2023. Total number of crimes that occurred in 2023 are 6878. A bar plot was used here to see the frequency of each crime over the year. It is observed from the plot that the violent-crimes happened most often compared to other crimes. Violent crimes took place 2633 times in year 2023. The second most type of crimes that occurred are anti-social-behaviour crimes which are less compared to violent crimes. There are 677 instances where anti social behaviour crimes took place. The records indicate that there are very less occasions where possession-of-weapons crimes occurred.

Visualizing top 2 crimes in colchester using time series plots

```
##Violent crimes

#Filtering data based on category
violent_crime_df <- crime_df[crime_df$category=="violent-crime",]

#Grouping data by date
violent.crime.ts.df <- violent_crime_df %>% group_by(date) %>% summarise(total_violent_crimes_ =table(dat
e))

#Preparing data for time series analysis
v.dates_ <- ym(violent.crime.ts.df$date)
violent.crimes_ <- as.numeric(violent.crime.ts.df$total_violent_crimes_)
violent.ts.crimes_ <- xts(data.frame(Crimes=violent.crimes_), v.dates_)

#Plotting using autoplot
violent_ts_plot <- autoplot(violent.ts.crimes_)+geom_point(data = violent.crime.ts.df, aes(x = v.dates_, y
= violent.crimes_, text = paste("Month : ", format(v.dates_, "%m"), "<br>Violent Crimes: ", violent.crimes
_)), color = "brown", size = 3) + labs(title = "Violent Crimes in Colchester by month in 2023", x = "Year
2023", y = "Number of Violent Crimes") + theme_bw()

#Converting autoplot into interactive plot
plotly_violent_ts <- ggplotly(violent_ts_plot, tooltip = "text", height = 450, width = 800) %>%layout( hov
erlabel = list(bgcolor = "cyan", font = list(color = "black")))
#Customizing the title with layout
plotly_violent_ts <- layout(plotly_violent_ts, title = list(text = "Violent Crimes in Colchester by month
in 2023", x = 0.5, y = 0.95))


##Anti Social Crimes

#Filtering data based on category
antisocial_crime_df <- crime_df[crime_df$category=="anti-social-behaviour",]

#Grouping data by date
antisocial.crime.ts.df <- antisocial_crime_df %>% group_by(date) %>% summarise(total_antisocial_crimes_ =t
able(date))

#Preparing data for time series analysis
as.dates_ <- ym(antisocial.crime.ts.df$date)
antisocial.crimes_ <- as.numeric(antisocial.crime.ts.df$total_antisocial_crimes_)
antisocial.ts.crimes_ <- xts(data.frame(Crimes=antisocial.crimes_), as.dates_)

#Plotting using autoplot
antisocial_ts_plot <- autoplot(antisocial.ts.crimes_)+geom_point(data = antisocial.crime.ts.df, aes(x = a
s.dates_, y = antisocial.crimes_, text = paste("Month : ", format(as.dates_, "%m"), "<br>Anti Social Crime
s: ", antisocial.crimes_)), color = "brown", size = 3) + labs(title = "Anti Social Crimes in Colchester by
month in 2023", x = "Year 2023", y = "Number of Anti Social Crimes") + theme_bw()

#Converting autoplot into interactive plot
plotly_antisocial_ts <- ggplotly(antisocial_ts_plot, tooltip = "text", height = 450, width = 800) %>%layou
t( hoverlabel = list(bgcolor = "cyan", font = list(color = "black")))

#Customizing the title with layout
plotly_antisocial_ts <- layout(plotly_antisocial_ts, title = list(text = "Anti Social Crimes in Colchester
by month in 2023", x = 0.5, y = 0.95))
```
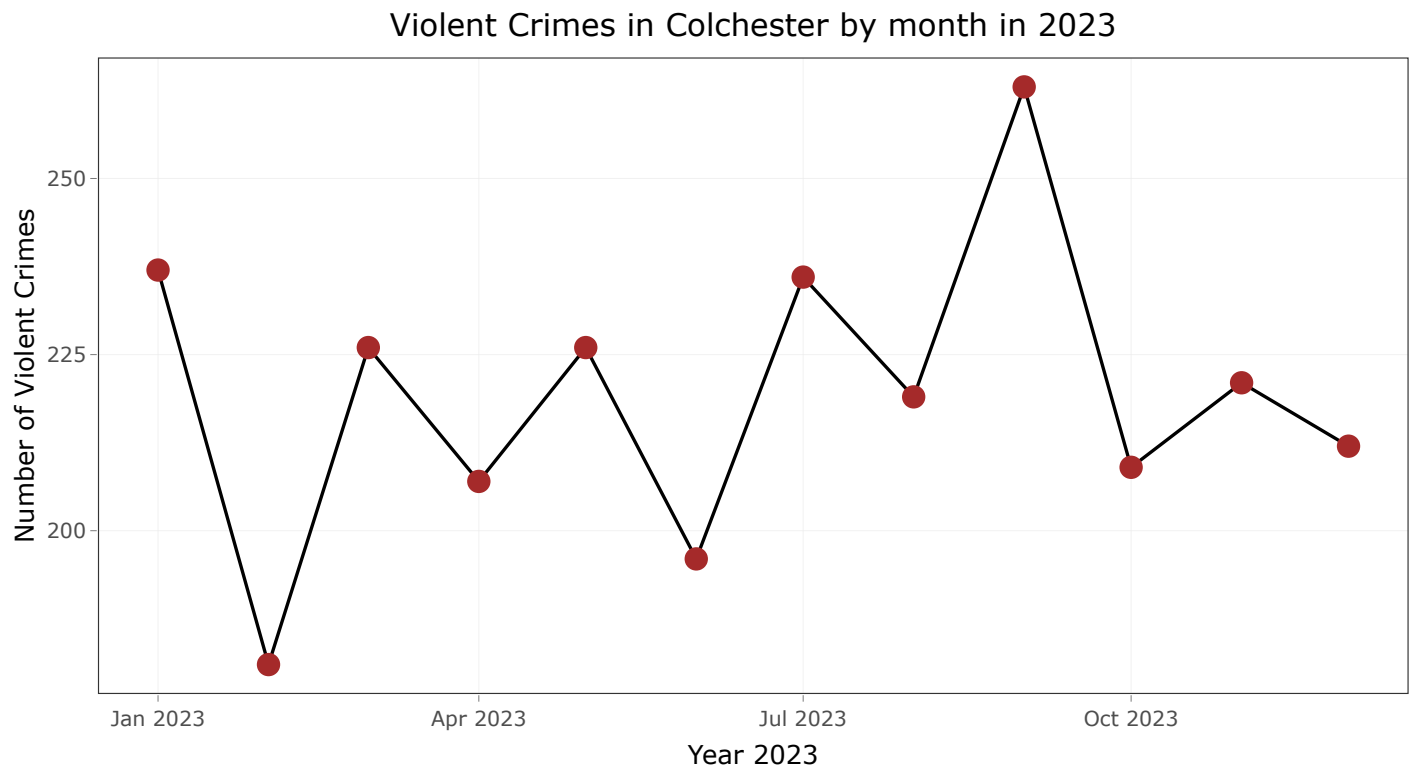
It is observed from the above bar plot that violent crimes and anti social crimes were the top 2 crimes that took place in Colchester in 2023. Here a time series plot is used to check frequency of the top 2 crimes in each month over the year.
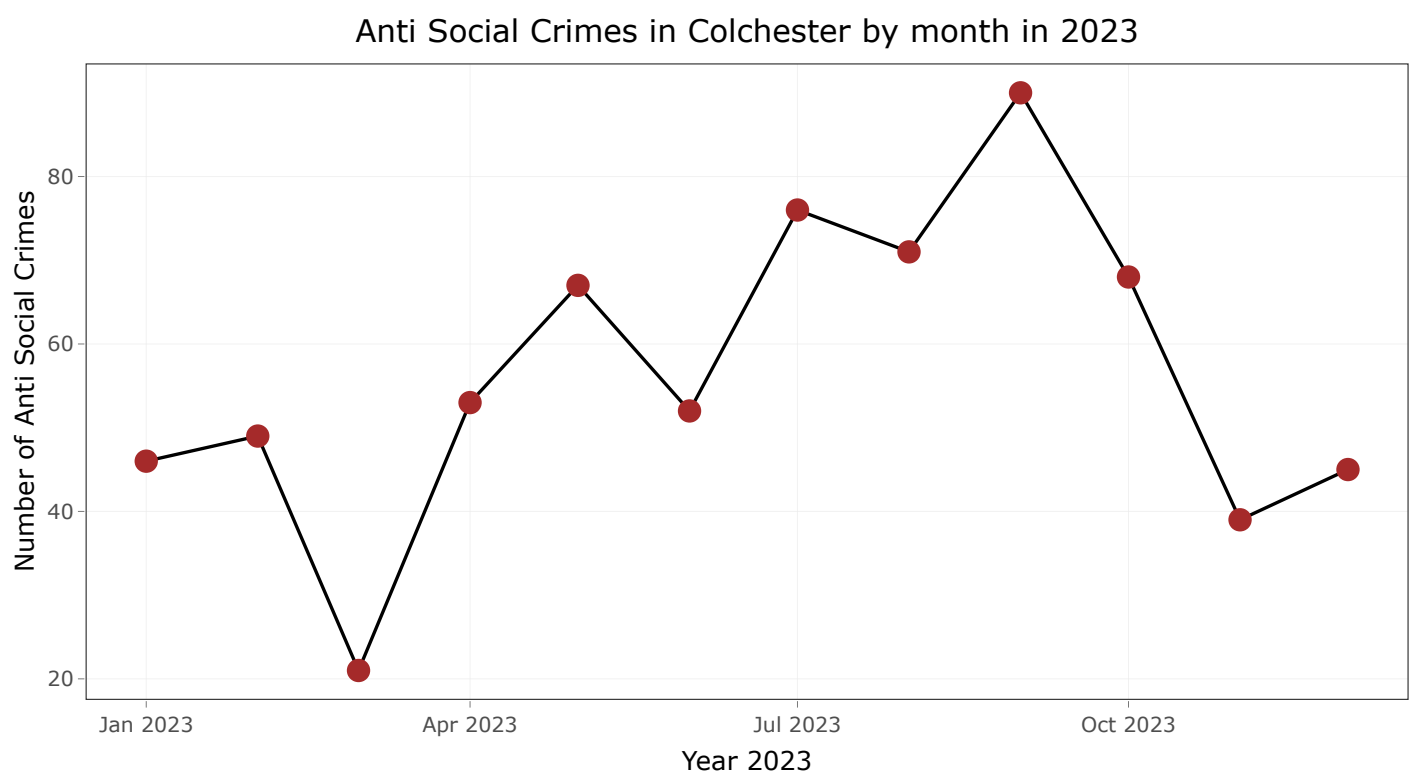
Violent crimes

```
#Plotting the time series plot
plotly_violent_ts
```

Violent Crimes in Colchester by month in 2023

A time series plot was used here to check the frequency of violent crimes over the year. It is observed from the plot that the number of violent crimes in the month of September are more compared to others. There are 263 instances in september where violent crime occurred. Another observation from the above time series plot is that, the violent crimes occurred more than 200 time in every month of year 2023. This indicates that violent crimes took place more often in Colchester in 2023.

Anti Social crimes

```
#Plotting the time series plot
plotly_antisocial_ts
```

Anti Social Crimes in Colchester by month in 2023

A time series plot was used here to check the frequency of anti social crimes over the year. It is observed from the plot that, the number of violent crimes have increased over the year till September and then reduced towards the end of the year. The number of anti social crimes that occurred in September were more compared to other months. There are 90 instances in September where anti social crime occurred. The least number of anti social crimes happened in the month of march. Another observation from the above time series plot is that, the anti social crimes reduced significantly in the month of march and stared increasing drastically. This indicates most of the anti social crimes in 2023 took place in between April and October months.

Visualizing crimes in colchester based on location type

```
#Table showing percentage of crimes in each location type
table_ <- round( 100*prop.table(table(crime_df$location_type)),2)
print('Table - Percentage of crimes in each location type')
```

```
## [1] "Table - Percentage of crimes in each location type"
```

```
table_
```

```
##
##   BTP Force
##  0.35 99.65
```

The crime dataset contains a location type variable which represents the type of the location in which the crimes took place. The location type is categorized into two types i.e., Force and BTP location types. Force location type indicates a normal police force location and BTP location type indicates a British Transport Police location. Here the the number of crimes and type of crimes occurred in both the location types were analyzed. From the table above it is observed that most of the crimes occurred in force location type.

```
## Crimes in Force location type

#Sorting the categories by number of crimes
force_crime_df_ <- crime_df[crime_df$location_type=="Force",]
force_crime_df_$category <- factor(force_crime_df_$category, levels = names(sort(table(force_crime_df_$cat
egory), decreasing = TRUE)))

#Creating bar plot object using ggplot
force_gg_bar_plot <- ggplot(force_crime_df_ ,aes(x=category)) +
  geom_bar(fill=heat.colors(n=length(unique(force_crime_df_$category)))) +
  scale_fill_gradient(low = "blue", high = "red") +
  labs(title = "Number of Crimes in Colchester by Category in Force location type", x="Crime Type", y="Num
ber of Crimes") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

#Interactive bar plot
force.plotly_bar <- ggplotly(force_gg_bar_plot, height=450, width=800) %>% layout( hoverlabel = list(bgcol
or = "cyan", font = list(color = "black")))

#Customizing the title with layout
force.plotly_bar <- layout(force.plotly_bar, title = list(text = "Number of Crimes in Colchester by Catego
ry in Force location type", x = 0.5, y = 0.96))


## Crimes in BTP location type

#Sorting the categories by number of crimes
btp_crime_df_ <- crime_df[crime_df$location_type=="BTP",]
btp_crime_df_$category <- factor(btp_crime_df_$category, levels = names(sort(table(btp_crime_df_$categor
y), decreasing = TRUE)))

#Creating bar plot object using ggplot
btp_gg_bar_plot <- ggplot(btp_crime_df_ ,aes(x=category)) +
  geom_bar(fill=heat.colors(n=length(unique(btp_crime_df_$category)))) +
  labs(title = "Number of Crimes in Colchester by Category in BTP location type", x="Crime Type", y="Numbe
r of Crimes") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

#Interactive bar plot
btp.plotly_bar <- ggplotly(btp_gg_bar_plot, height=450, width=800) %>% layout( hoverlabel = list(bgcolor =
"cyan", font = list(color = "black")))

#Customizing the title with layout
btp.plotly_bar <- layout(btp.plotly_bar, title = list(text = "Number of Crimes in Colchester by Category i
n BTP location type", x = 0.5, y = 0.96))
```
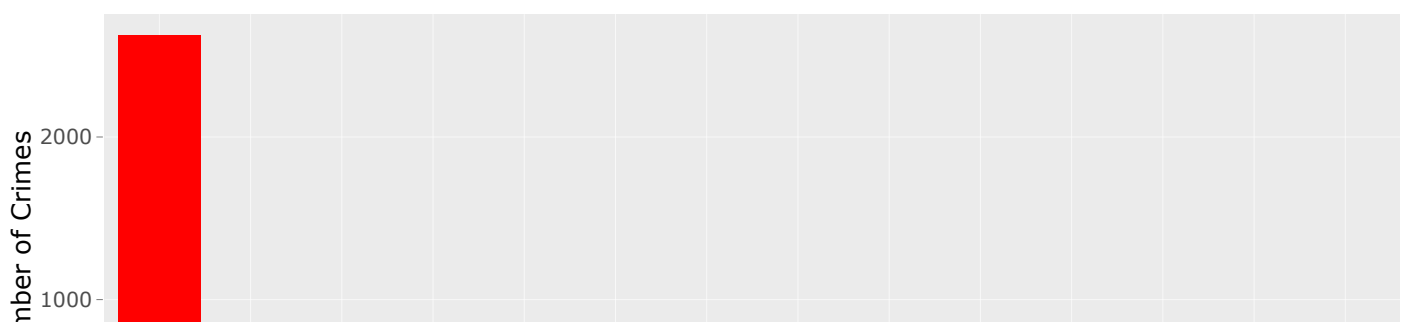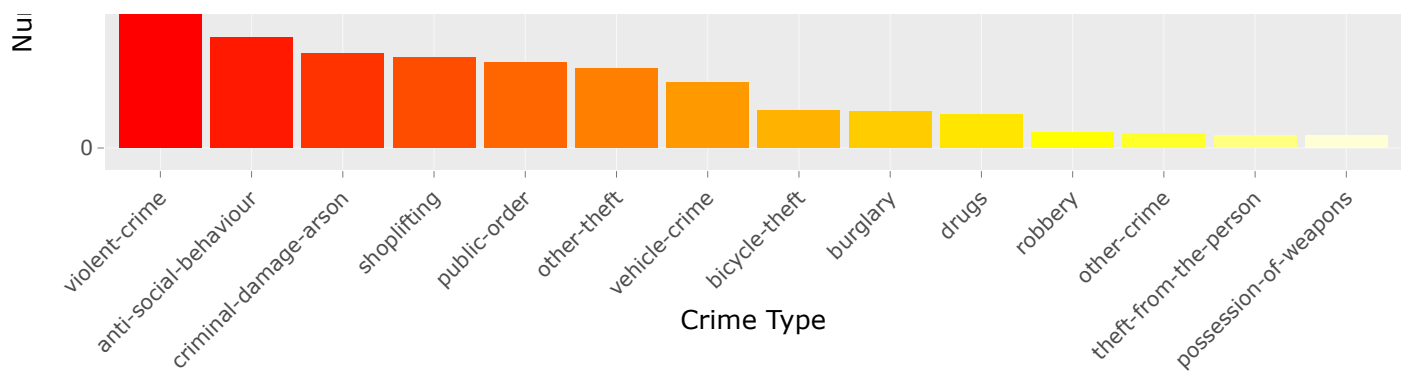
Crimes in Force location type

```
#Plotting the graph
force.plotly_bar
```

Number of Crimes in Colchester by Category in Force location type
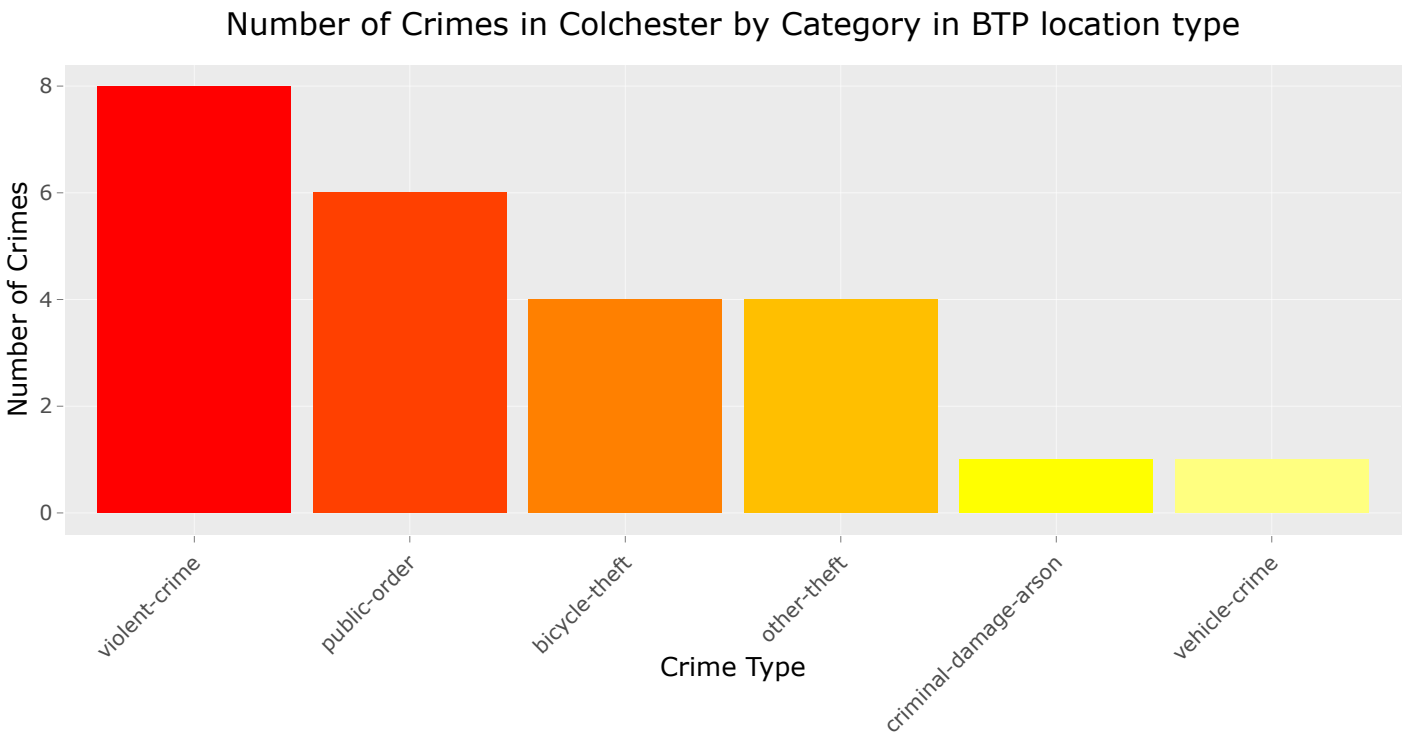
A bar plot is used to analyze the crimes that happened in Colchester in Force location type. It is observed from the graph that violent crimes and anti social behaviour crimes are the top 2 crimes that occurred in colchester in force location type. There were more than 2600 instances where violent crime took place. Anti social crimes, criminal damage arson, shoplifting and vehicle theft took place around 500 times each.

Crimes in BTP location type

```
#Plotting the graph
btp.plotly_bar
```

Number of Crimes in Colchester by Category in BTP location type

Here a bar plot is used to visualize the crimes that occurred in BTP location type in Colchester in 2023. It is observed from the plot that 5 types of crimes took place in BTP location type. Violent crimes and public order crimes are the top 2 crimes that occurred in BTP location type. Violent crimes and public order crimes are the top 2 crimes that occurred in the BTP location type.

Visualizing number of crimes in colchester in a Map using leaflet

```
#Grouping crimes based on latitude, longitude and street name
crimes.map.df <- crime_df %>%
  group_by(lat,long,street_name) %>%
  summarise(crimes_ =table(long))

#Plotting map using leaflet
crimes.map_ <- crimes.map.df %>%
  leaflet() %>%
  addTiles() %>%
  setView(0.900117, 51.88924, zoom = 14) %>%
  addCircleMarkers(radius=~crimes_*0.13, color = "darkblue",
    popup = ~paste(crimes.map.df$street_name, "<br>Crimes : ",crimes_))

#Adding title to the map
crimes.map_ <- addControl(
  crimes.map_,
  position = "topright",
  title <- "Colchester crimes locator",
  )

#Visualizing map
crimes.map_
```
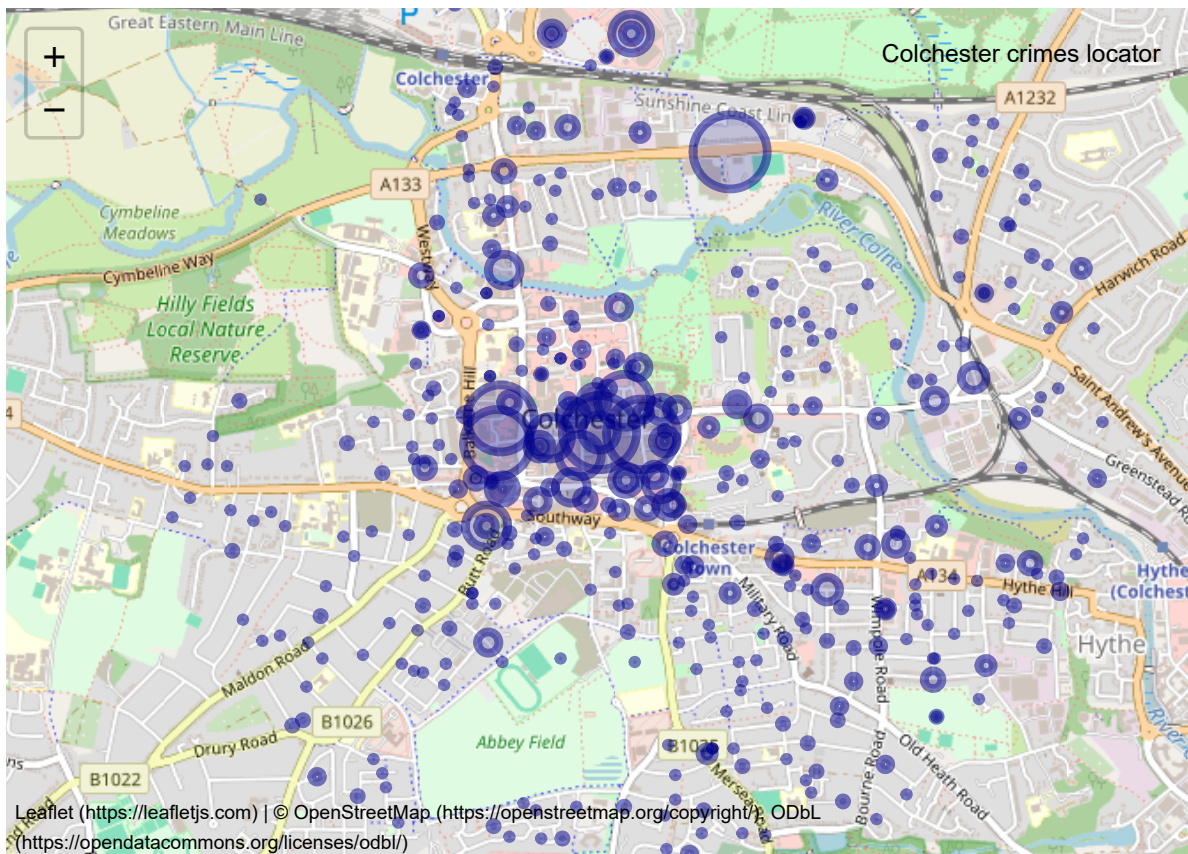

The number of crimes occurred in each street in Colchester in the year 2023 were visualized using open street map which is plotted using leaflet. The crimes were grouped based on the latitude, longitude and street name to get the count of crimes happened in specific street. The map above clearly shows number of crimes that took place in each street of Colchester. The size of the circles in the map indicates the count of crimes i.e., bigger the size of the circle more the number of crimes occurred in that specific area. From the map, it is observed that the crimes happened near Shopping Area are more compared to others. There are 328 instances of crimes that took place near shopping area. Some other areas in which the crime rate is more are Cowdray Avenue, St Nicholas Street, Balkerne Gardens, Church Street and George Street.

Visualizing crimes in top 2 streets where crime rate is higher

```
##On or near Shopping Area

#Filtering dataset based on street name
shopping_area_df_ <- crime_df[crime_df$street_name=="On or near Shopping Area",]

#Assigning labels and values to variables
shp_crime_freq_ <- table(shopping_area_df_$category)
shp_crime_names_ <- names(shp_crime_freq_)

#Plotting pie chart using plotly
shopping.area.pie_plot <- plot_ly(labels = shp_crime_names_, values = shp_crime_freq_, type = "pie", heigh
t = 450, width = 800)

#Customize the title with layout
shopping.area.pie_plot <- layout(shopping.area.pie_plot, title = list(text = "Types of Crimes took place n
ear Shopping Area", x = 0.5, y = 0.99))


##On or near Cowdray Avenue

#Filtering dataset based on street name
cowdray_avenue_df_ <- crime_df[crime_df$street_name=="On or near Cowdray Avenue",]

#Assigning labels and values to variables
cw_crime_freq_ <- table(cowdray_avenue_df_$category)
cw_crime_names_ <- names(cw_crime_freq_)

#Plotting pie chart using plotly
cowdray.avenue.pie_plot <- plot_ly(labels = cw_crime_names_, values = cw_crime_freq_, type = "pie", height
= 450, width = 800)

#Customize the title with layout
cowdray.avenue.pie_plot <- layout(cowdray.avenue.pie_plot, title = list(text = "Types of Crimes took place
near Cowdray Avenue", x = 0.5, y = 0.99))
```

It is observed from the map above that top 2 streets in which the crime rate is more are near Shopping Area and Cowdray Avenue. Here analysis was carried out using pie chart to check the types of crimes that took place in these specific streets and which crime happened most frequently.

On or near Shopping Area

```
#Plotting the pie chart
shopping.area.pie_plot
```
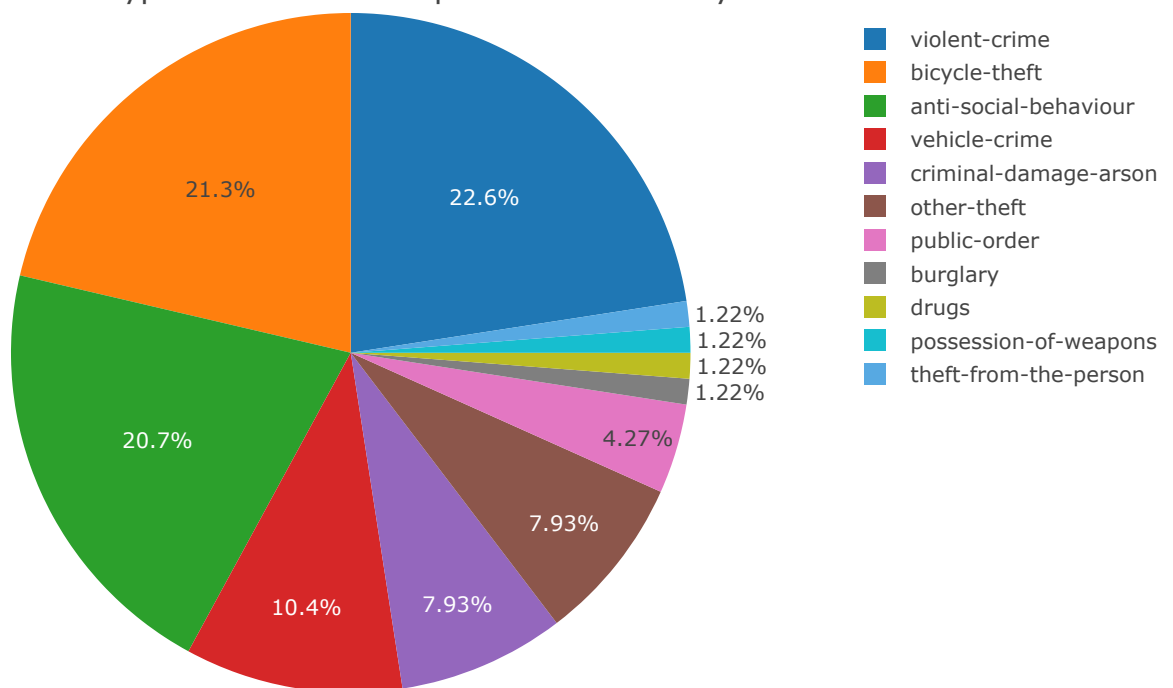

Types of Crimes took place near Shopping Area
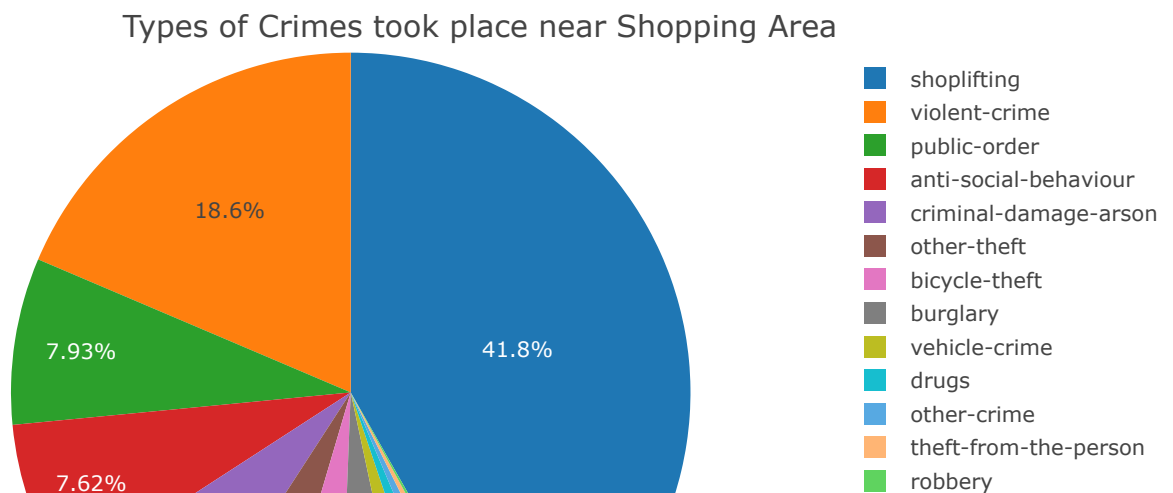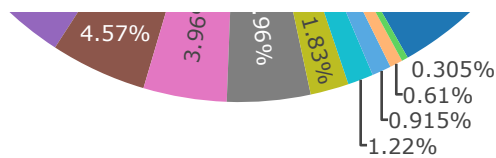
The crimes that took place near Shopping area are more compared to other streets. Here, a pie chart is used to visualize and understand the types of crimes that occurred in this street and the frequency of the each crime category. Form the pie chart above, it is clearly seen that around 42 percent of the total crimes that took place near the Shopping area are Shop Lifting crimes. The second most number of crimes that happened are Violent crimes which is around 19 percent of the total crimes that occurred in this street. Public Order crimes, Anti Social Behaviour crimes and Criminal Damage Arson crimes are around 7 percent of the total number of crimes.

On or near Cowdray Avenue

```
#Plotting the pie chart
cowdray.avenue.pie_plot
```

Types of Crimes took place near Cowdray Avenue

Cowdray Avenue has second most number of crimes occurred in colchester in 2023. Here, a pie chart is used to visualize and understand the types of crimes that occurred in this street and the frequency of the each crime category. Form the pie chart above, it is clearly seen that Violent crimes happened around 23 percent of the total crimes that took place near the Shopping. The second most number of crimes that happened are Bicycle theft crimes and Anti Social Behaviour crimes which are around 21 percent each of the total crimes that occurred in this street. Vehicle crimes and Criminal Damage Arson crimes are next in the list which are of around 10 and 8 percent of the total number of crimes.

Visualizing total number of crimes in colchester by month in 2023 using Time Series Plot

```
#Grouping data by date
crimes.ts.df <- crime_df %>% group_by(date) %>% summarise(total_crimes_ =table(date))

#Preparing data for time series analysis
crime_dates_ <- ym(crimes.ts.df$date)
crimes_ <- as.numeric(crimes.ts.df$total_crimes_)
crime_ts_ <- xts(data.frame(Crimes=crimes_), crime_dates_)

#Plotting using autoplot
crime_ts_plot_ <- autoplot(crime_ts_, facet=NULL)+geom_point(data = crimes.ts.df, aes(x = crime_dates_, y
= crimes_, text = paste("Month : ", format(crime_dates_, "%m"), "<br>Crimes: ", crimes_)), color = "brow
n", size = 3) + labs(title = "Number of Crimes in Colchester by Month in 2023", x = "Year 2023", y = "Numb
er of Crimes") + theme_bw()

#Converting autoplot into interactive plot
crime_plotly_ts <- ggplotly(crime_ts_plot_,tooltip = "text") %>%layout( hoverlabel = list(bgcolor = "cya
n", font = list(color = "black")))

#Customizing the title with layout
crime_plotly_ts <- layout(crime_plotly_ts, title = list(text = "Number of Crimes in Colchester by Month in
2023", x = 0.5, y = 0.97))

#Plotting the time series graph
crime_plotly_ts
```
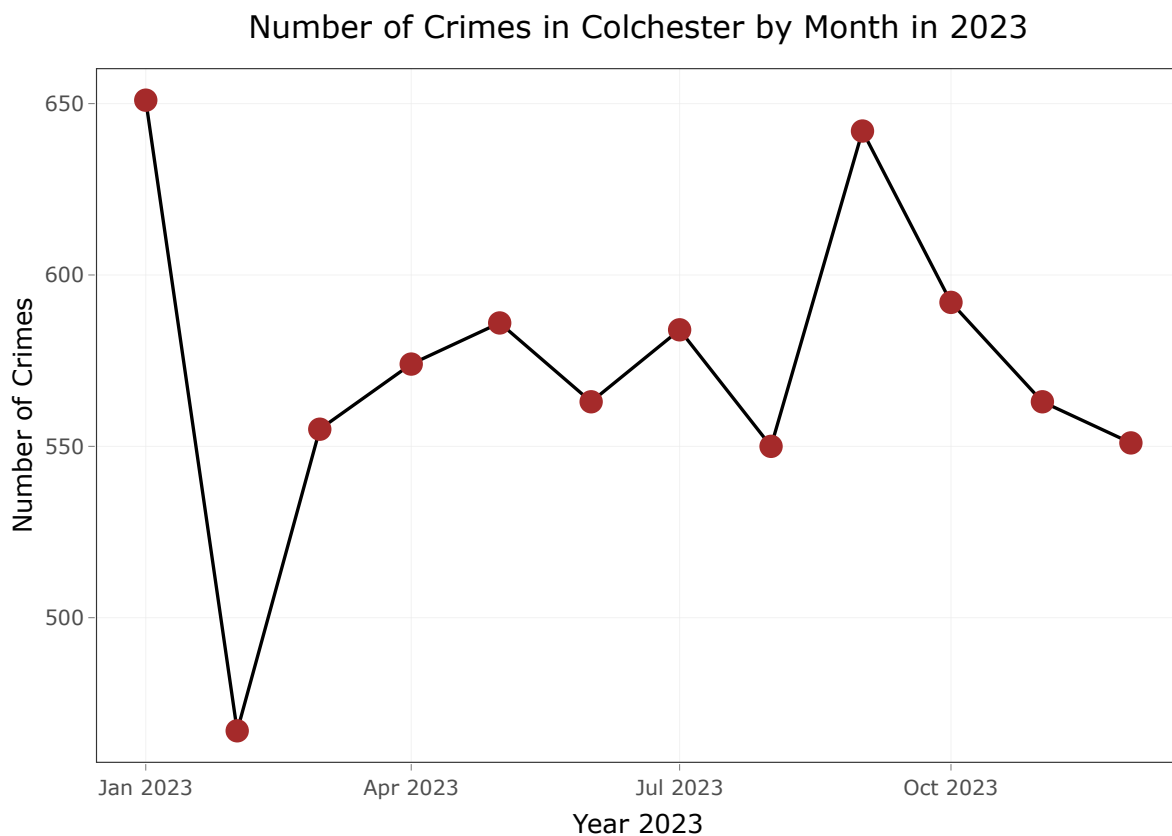

Number of Crimes in Colchester by Month in 2023

In this report, types of crimes and their frequencies were analyzed till now. Here, a time series plot is used to visualize the total number of crimes that occurred in each month over the year 2023 in Colchester. To do this analysis, the crimes were grouped by month to get the number of crimes that took place in each month. From the time series plot, it is observed that there is a sudden dip in the crime rate in the month of February and then started increasing over the year till September. The crime rate is more in the months of January and September compared to other months. The least number crimes occurred in the month of February. Another observation was there is a sudden raise in crime rate in the months of March and September. Another observation was the number of crimes occurred in each month were in the range of 450 to 650 in each month.

Visualizing the Crimes recored in January and September months in colchester in year 2023

```
##Bar plot of January month crimes

#Sorting the categories by number of crimes
january_crime_df_ <- crime_df[crime_df$date == "2023-01" ,]
january_crime_df_$category <- factor(january_crime_df_$category, levels = names(sort(table(january_crime_d
f_$category), decreasing = TRUE)))

#Bar plot using ggplot
gg_bar_january_ <- ggplot(january_crime_df_ ,aes(x=category)) +
  geom_bar(fill=heat.colors(n=length(unique(january_crime_df_$category)))) +
  labs(title = "Crimes in Colchester in January 2023 by each Category", x="Category", y="Number of Crime
s") +
  coord_flip() + theme(axis.text.x = element_text(angle = 45, hjust = 1))

#Converting ggplot into interactive plot
plotly_bar_january <- ggplotly(gg_bar_january_, height=450, width=800)%>% layout( hoverlabel = list(bgcolo
r = "cyan", font = list(color = "black")))

#Customizing the title with layout
plotly_bar_january <- layout(plotly_bar_january, title = list(text = "Crimes in Colchester in January 2023
by each Category", x = 0.5, y = 0.95))

##Bar plot of September month crimes

#Sorting the categories by number of crimes
september_crime_df_ <- crime_df[crime_df$date == "2023-07" ,]
september_crime_df_$category <- factor(september_crime_df_$category, levels = names(sort(table(september_c
rime_df_$category), decreasing = TRUE)))

#Bar plot using ggplot
gg_bar_september_ <- ggplot(september_crime_df_ ,aes(x=category)) +
  geom_bar(fill=heat.colors(n=length(unique(september_crime_df_$category)))) +
  labs(title = "Crimes in Colchester in September 2023 by each Category", x="Category", y="Number of Crime
s") +
  coord_flip() + theme(axis.text.x = element_text(angle = 45, hjust = 1))

#Converting ggplot into interactive plot
plotly_bar_september <- ggplotly(gg_bar_september_, height=450, width=800)%>% layout( hoverlabel = list(bg
color = "cyan", font = list(color = "black")))

#Customizing the title with layout
plotly_bar_september <- layout(plotly_bar_september, title = list(text = "Crimes in Colchester in Septembe
r 2023 by each Category", x = 0.5, y = 0.95))
```

It is seen in the above time series plot that the number of crimes that occurred in the months of January and September are more compared to other months. Here the further analysis is carried out to see types of crimes that occurred in these 2 months and the frequency of each crime type.
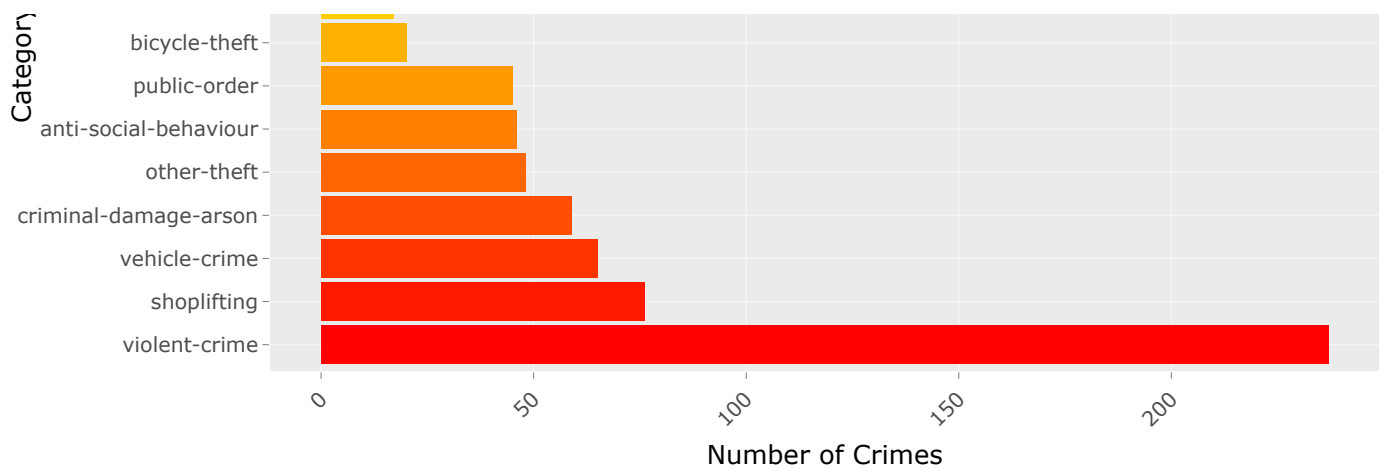
January month crimes

```
#Plotting the graph
plotly_bar_january
```

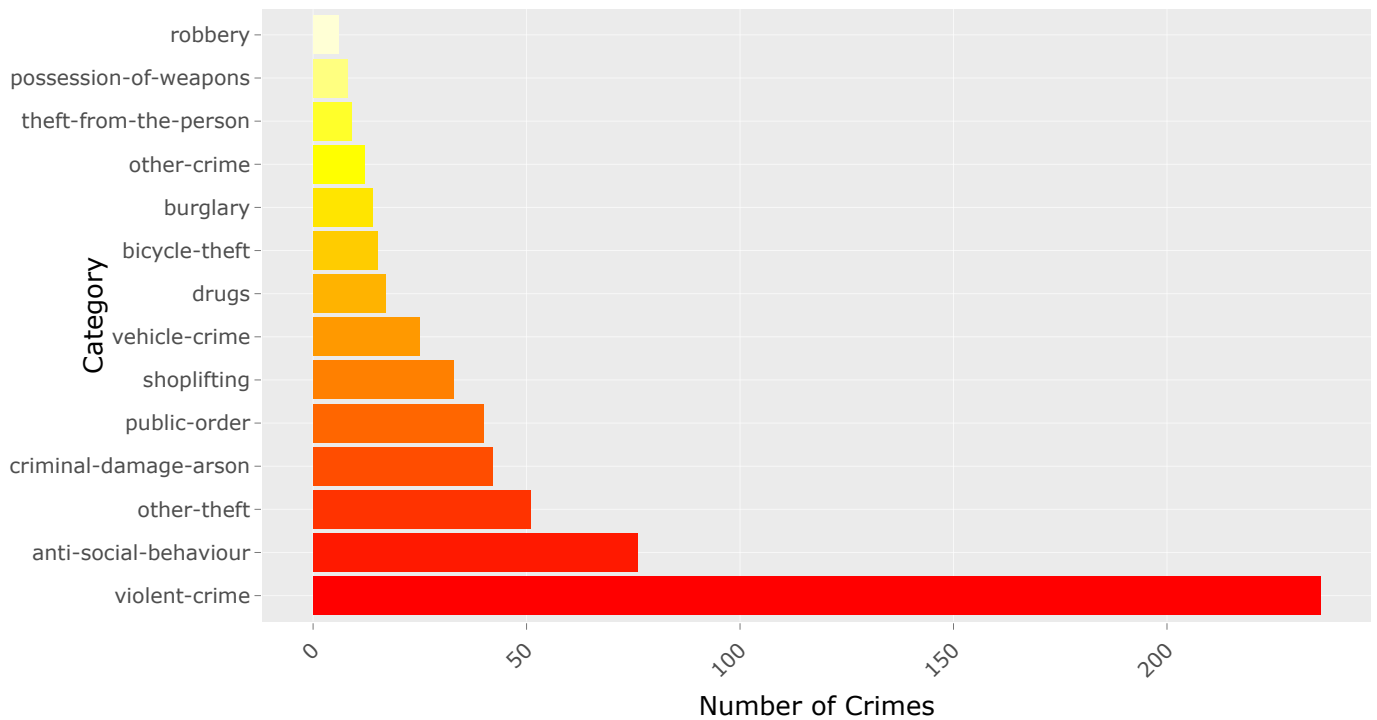## Crimes in Colchester in January 2023 by each Category

The types of crimes that took place in the month of January and the frequency of each crime is visualized using bar plot. It is seen from the graph that Violent crimes occurred in 237 instances which is higher compared to other crimes. The second most happened crime in January are Shop Lifting which took place 76 times. Crimes such as Vehicle crime, Criminal Damage occurred around 60 times each.

September month crimes

```
#Plotting the graph
plotly_bar_september
```

Crimes in Colchester in September 2023 by each Category

The types of crimes that took place in the month of September and the frequency of each crime is visualized using bar plot. It is seen from the graph that Violent crimes occurred in 236 instances which is higher compared to other crimes. The second most happened crime in September are Anti Social Behaviour crimes which took place 76 times. Crimes such as Criminal Damage and Other thefts occurred around 50 times each.

## Visualizing the outcome status of crimes using Pie plot

```
#Assigning labels and values to variables
crime_freq_ <- table(crime_df$outcome_status)
crime_names_ <- names(crime_freq_)

#Plotting the pie chart using plotly
outcome_status_pie_plot <- plot_ly(labels = crime_names_, values = crime_freq_, type = "pie", height = 45
0, width = 800)

#Customizing the title with layout
outcome_status_pie_plot <- layout(outcome_status_pie_plot, title = list(text = "Violent Crimes in Colchest
er by month in 2023", x = 0.5, y = 0.99))

#Plotting the pie chart
outcome_status_pie_plot
```


Violent Crimes in Colchester by month in 2023
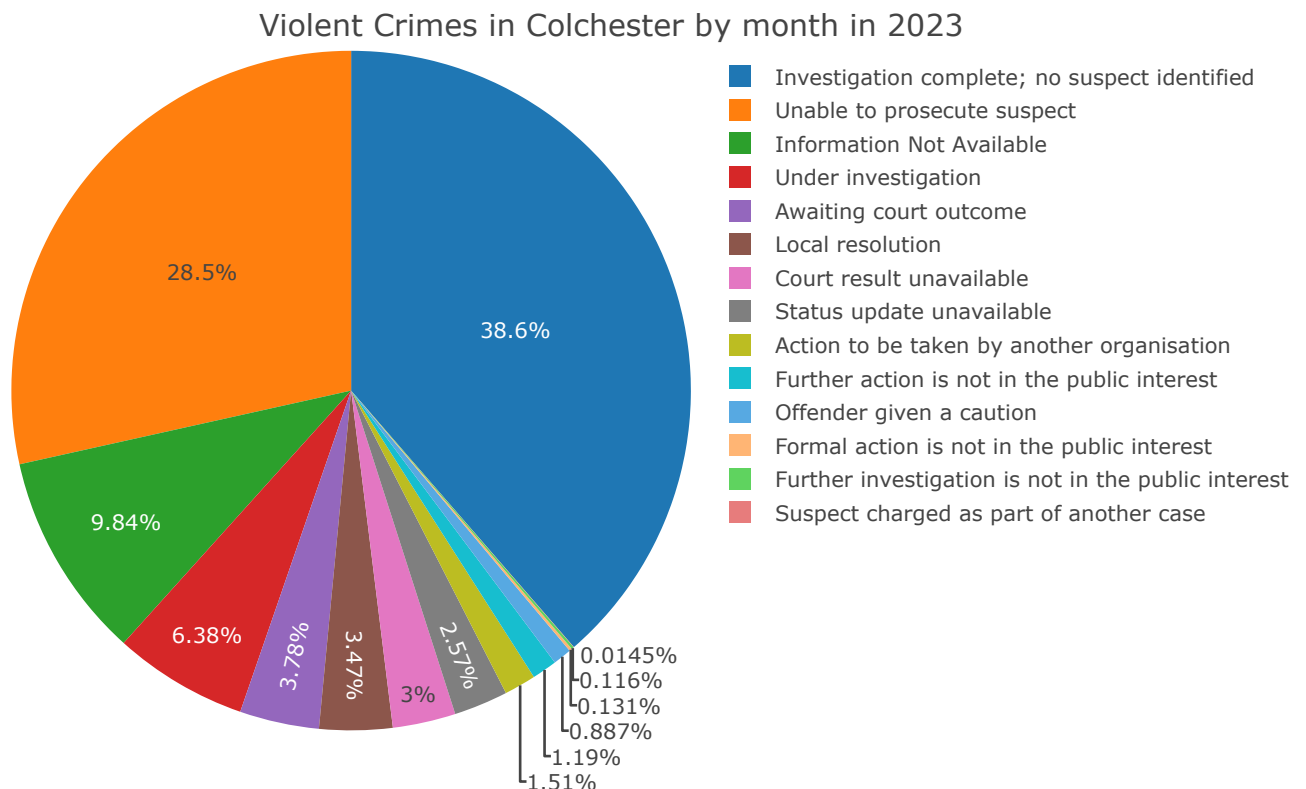
The dataset has a column outcome status which indicates the status of the crime investigation. A pie chart was used to check the different outcome status of the crimes that took place in Colchester in 2023. The pie chart above shows that for around 39 percent of the total crimes, the investigation was completed and no suspect was identified in these cases. For another 29 percent of the total crimes the investigation status was given as Unable to prosecute suspect. The outcome status was not available for around 10 percent of the total crimes and around 7 percent of the crimes have outcome status as Local resolution.

Number of crimes by outcome status as investigation completed

```
#Filtering the data by outcome status
invst_cmplt_crime_df_ <- crime_df[crime_df$outcome_status == "Investigation complete; no suspect identifie
d" ,]

#Sorting the categories by number of crimes
invst_cmplt_crime_df_$category <- factor(invst_cmplt_crime_df_$category,
                                         levels= names(sort(table(invst_cmplt_crime_df_$category), decreas
ing = TRUE)))
#Bar plot using ggplot
gg_bar_invst_cmplt_ <- ggplot(invst_cmplt_crime_df_ , aes(x=category)) +
  geom_bar(fill=heat.colors(n=length(unique(invst_cmplt_crime_df_$category))), alpha = 0.9) +
  labs(title = "Investigation Completed Crimes by Category", x="Types of Crime", y="Number of Crimes") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

#Converting ggplot into interactive plot
plotly_bar_invst_cmplt <- ggplotly(gg_bar_invst_cmplt_, height=450, width=800)%>% layout( hoverlabel = lis
t(bgcolor = "cyan", font = list(color = "black")))

#Customizing the title with layout
plotly_bar_invst_cmplt <- layout(plotly_bar_invst_cmplt, title = list(text = "Investigation Completed Crim
es by Category", x = 0.5, y = 0.95))

#Plotting the graph
plotly_bar_invst_cmplt
```
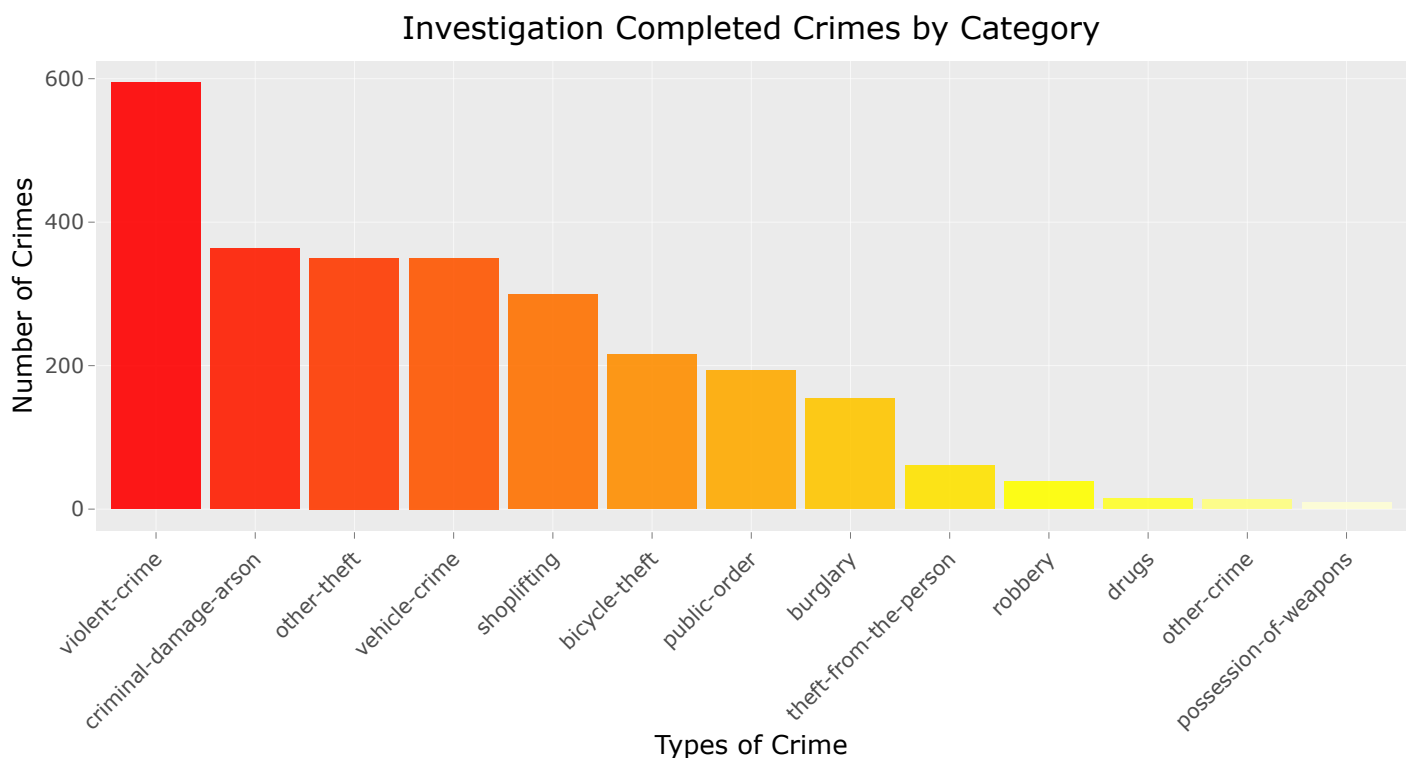
Investigation Completed Crimes by Category

It is observed from the above pie chart that the outcome status was given as Investigation Completed for around 40 percent of the total crimes. A bar chart was used here to check the types of crimes for which the investigation is completed. It is seen from the above bar plot that in 595 violent crimes, 363 criminal damage crimes, 350 vehicle crimes and 299 shop lifting crimes the investigation was completed and no suspect was identified.

Analysis on Climate Dataset

Visualizing data distribution of average temperature using histogram

```
#Histogram plot using ggplot
avgtemp_gg_hist_ <- ggplot(climate_df,aes(x=TemperatureCAvg)) +
  geom_histogram(color = "black", fill = heat.colors(n=27), binwidth = 1, alpha = 0.9) +
  labs(title = "Average Temperature in Colchester in 2023", x = "Avg Temperature(C)", y = "Frequency") +
  theme_classic()

#Converting ggplot into interactive plot
avgtemp_hist_plotly_ <- ggplotly(avgtemp_gg_hist_, height = 450, width = 800)

#Customizing the hover label appearance
avgtemp_hist_plotly_ <- avgtemp_hist_plotly_ %>% layout( hoverlabel = list(bgcolor = "cyan", font = list(c
olor = "black")))

#Customizing the title with layout
avgtemp_hist_plotly_ <- layout(avgtemp_hist_plotly_, title = list(text = "Average Temperature in Colcheste
r in 2023", x = 0.5, y = 0.95))

#Plotting the graph
avgtemp_hist_plotly_
```
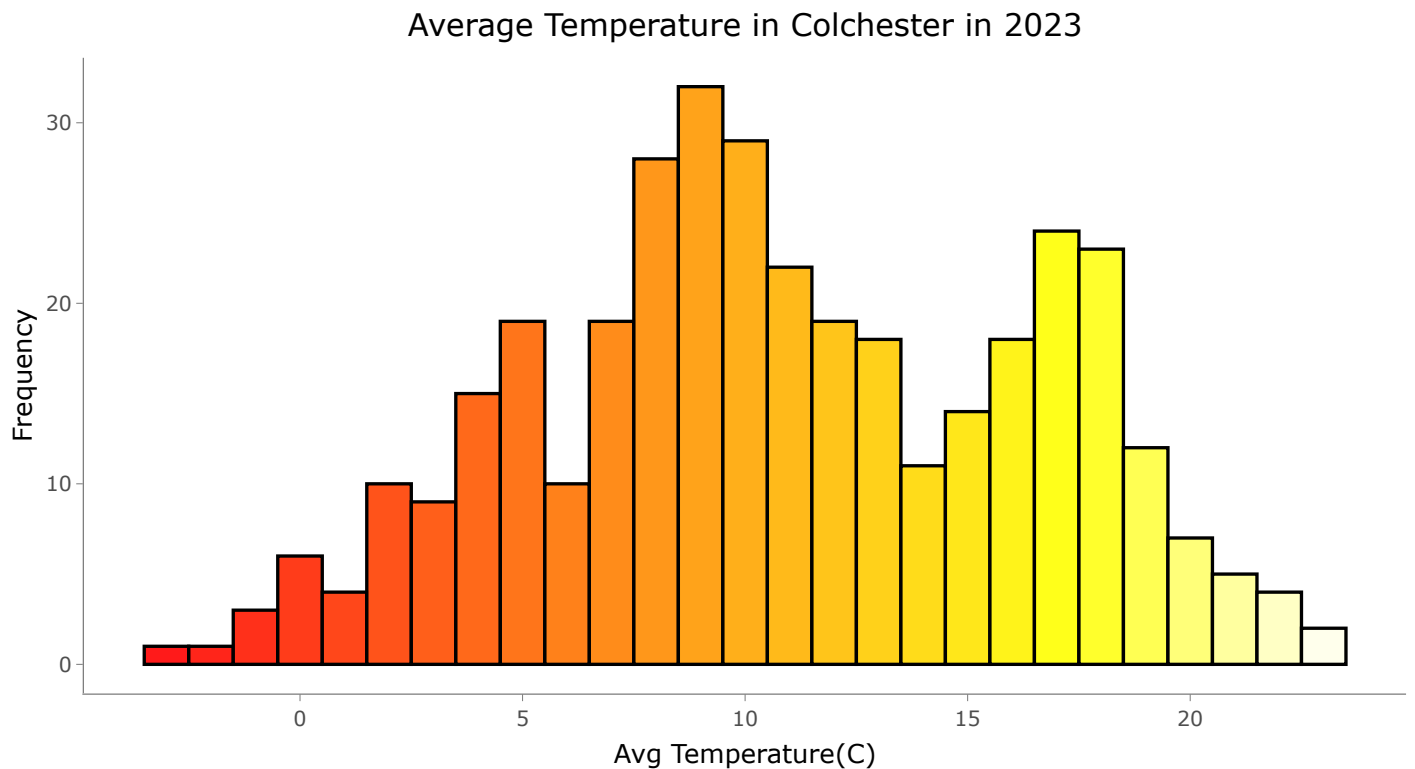

The temperature dataset has a Average temperature column which indicates the average temperature in Colchester on each day in 2023. Here a histogram plot was used to check the average temperature distribution over the year. It is seen from the plot that on 32 days the average temperature was 9 degrees and on 29 days it was 10 degrees. It is also observed that the average temperature over the year ranges from -3 degrees to 23 degree Celsius. On most of the days the average temperature was in between 7 to 13 degree Celsius. Another key observation was only on 4 days in 2023, the temperature has gone down below 0 degree Celsius.

Time series plot to visualize Min Temp and Max Temp over the year

```
#Preparing data for time series plot
temp_df_ <- climate_df[,c("TemperatureCMin","TemperatureCMax")]
temp_dates_ <- ymd(climate_df$Date)

#Time series data
ts.temp_ <- xts(temp_df_, temp_dates_)

#Modifying the series names
colnames(ts.temp_) <- c("Min Temp(C)", "Max Temp(C)")

#Creating Time Series object using autoplot
temp_ts_plot_ <- autoplot(ts.temp_, facet=NULL) +
  xlab("Dates") +
  ylab("Temperature(C)") +
  theme_bw() +
  aes(text = paste("Date : ", Index, "<br>Temp(C) : ", round(Value,2)))

#Interactive plot using ggplotly
plotly_temp_ts <- ggplotly(temp_ts_plot_, tooltip = "text", height = 450, width = 800) %>%layout( hoverlab
el = list(bgcolor = "cyan", font = list(color = "black")))

#Customizing the title using layout
plotly_temp_ts <- layout(plotly_temp_ts, title = list(text = "Min Temp and Max Temp in 2023", x = 0.5, y =
0.99))

#Plotting the time series plot
plotly_temp_ts
```
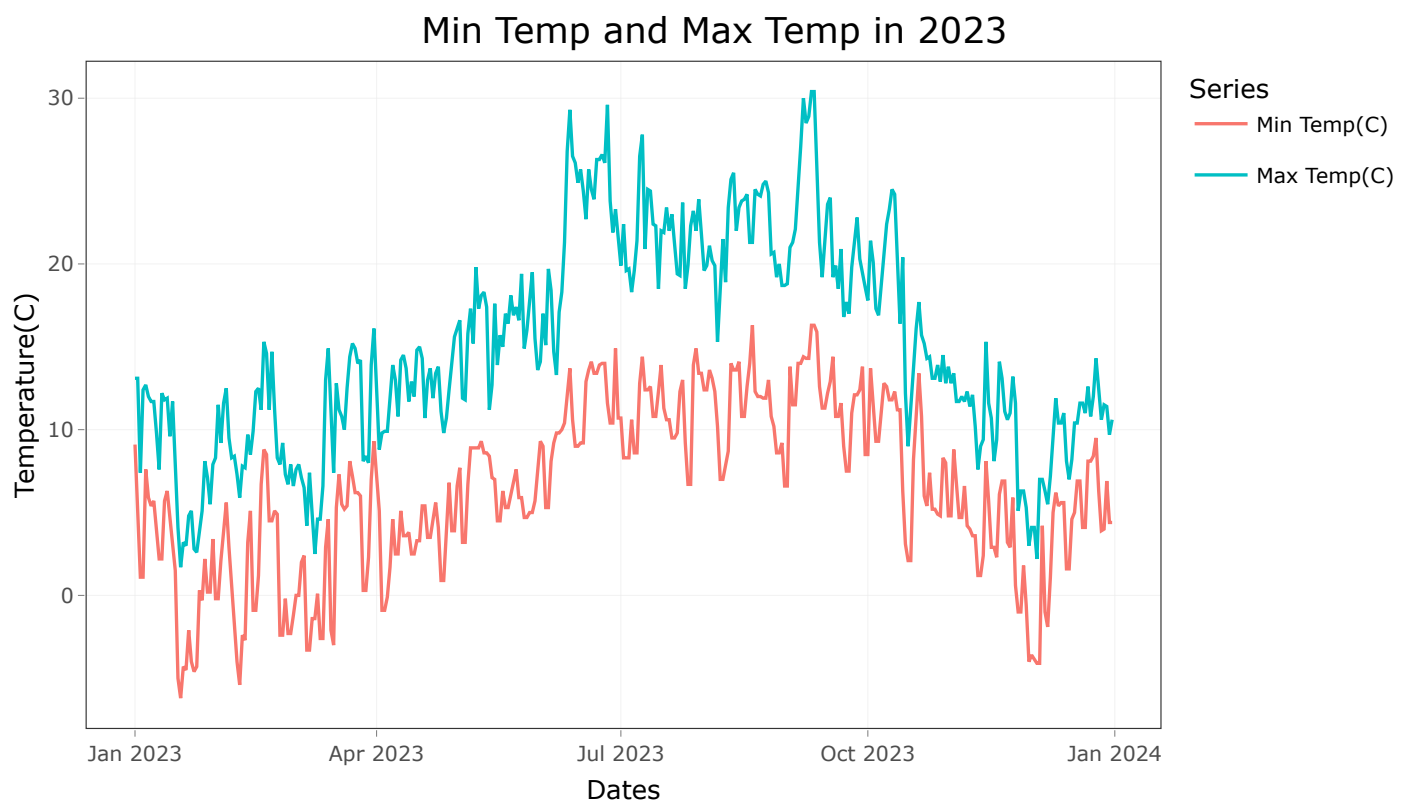
```
##Monthly Average of min temperature and max tempearture

#Preparing data for time series plot
temp_df_ma_ <- climate_df[,c("TemperatureCMin","TemperatureCMax")]
temp_dates_ <- ymd(climate_df$Date)

#Monthly Moving Average smoothing
temp_df_ma_$TemperatureCMin <- forecast::ma(temp_df_ma_$TemperatureCMin, 30)
temp_df_ma_$TemperatureCMax <- forecast::ma(temp_df_ma_$TemperatureCMax, 30)

#Time series data
ts.temp_ma_ <- xts(temp_df_ma_, temp_dates_)

#Modifying the series names
colnames(ts.temp_ma_) <- c("Min Temp<br>Monthly Avg(C)", "Max Temp<br>Monthly Avg(C)")

#Creating Time Series object using autoplot
temp_ma_ts_plot_ <- autoplot(ts.temp_ma_, facet=NULL) +
  xlab("Dates") +
  ylab("Temperature(C)") +
  theme_bw() +
  aes(text = paste("Date : ", Index, "<br>Temp(C) : ", round(Value,2)))

#Interactive plot using ggplotly
plotly_temp_ma_ts <- ggplotly(temp_ma_ts_plot_, tooltip = "text", height = 450, width = 800) %>%layout( ho
verlabel = list(bgcolor = "cyan", font = list(color = "black")))

#Customizing the title using layout
plotly_temp_ma_ts <- layout(plotly_temp_ma_ts, title = list(text = "Monthly Average of Min Temp and Max Te
mp in 2023", x = 0.5, y = 0.99))

#Plotting the time series plot
plotly_temp_ma_ts
```
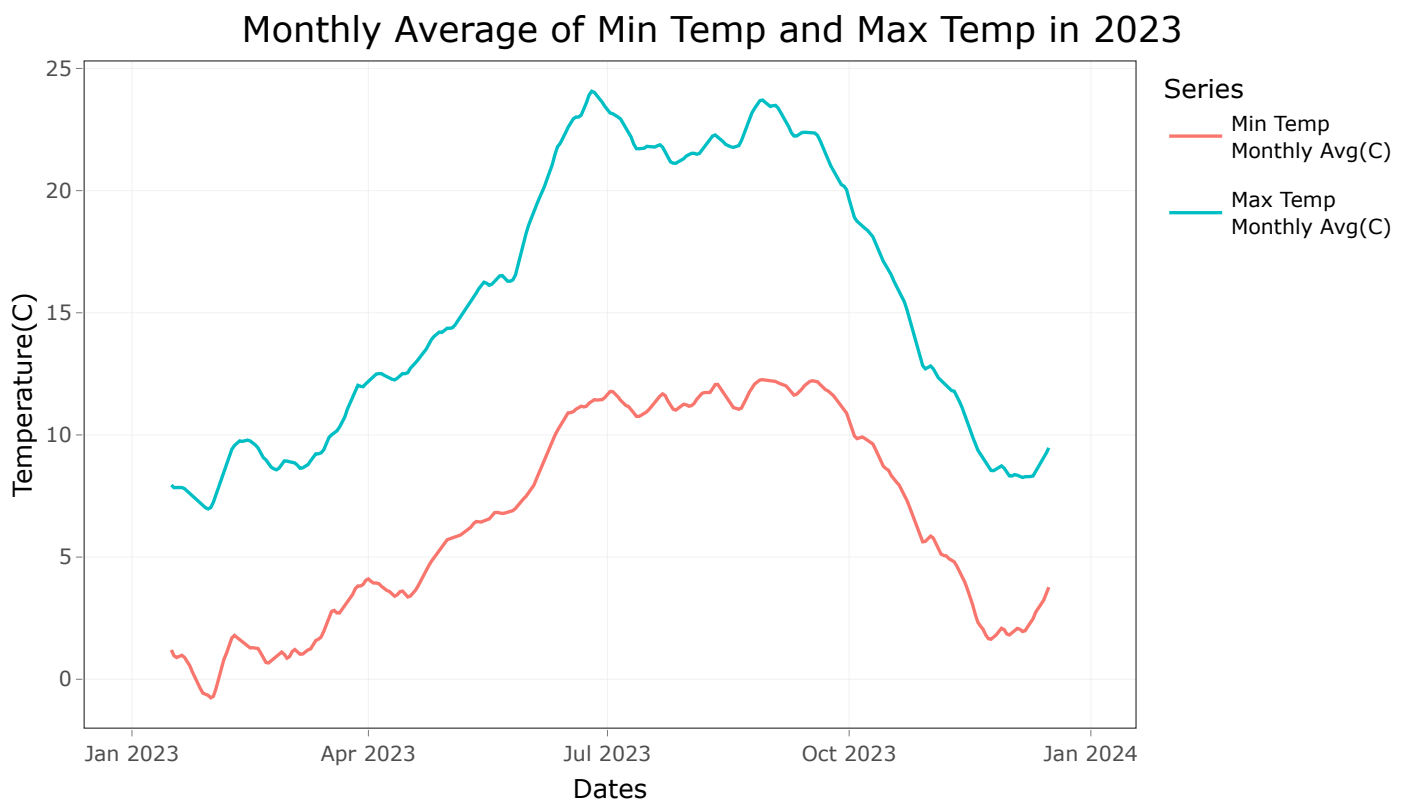
Monthly Average of Min Temp and Max Temp in 2023

The minimum and maximum temperature in 2023 is visualized using time series plot to check how the temperature was varied throughout the year. Here monthly moving average smoothing was used for better improve the visualization. From the above time series plot is observed that the temperature least in the months of January, February, November and December which is

a winter season in UK. The temperature started increasing from the month of March till September and then started decreasing again. The minimum temperature recorded in 2023 was -6.2 degree celsius and the maximum temperature recorded was 30.4 degree celsius.

Visulaizing the relationship between Wind Speed and Wind Gust using scatter plot

```
#Creating Scatter plot object using ggplot
wind_gg_plot<- ggplot(climate_df,aes(x = WindkmhInt, y = WindkmhGust, color = factor(WindkmhInt))) +
  geom_point(alpha = 0.9)+
  labs(title = "Scatter plot of Wind Speed vs Wind Gust", x="Wind Speed", y="Wind Gust") +
  theme_classic()

#Removing legend
wind_gg_plot <- wind_gg_plot + guides(color=FALSE)

#Interactive plot using ggplotly
wind_plotly_ <- ggplotly(wind_gg_plot,height = 450,width = 600)

#Modifying hover text in interactive plot
labels = paste("Wind Speed :", climate_df$WindkmhInt, "<br>Wind Gust :", climate_df$WindkmhGust)
wind_plotly_ <- style(wind_plotly_, text= labels)

#Customizing the title using layout
wind_plotly_ <- layout(wind_plotly_, title = list(text = "Scatter plot of Wind Speed vs Wind Gust", x = 0.
5, y = 0.95))

#Plotting the scatter plot
wind_plotly_
```
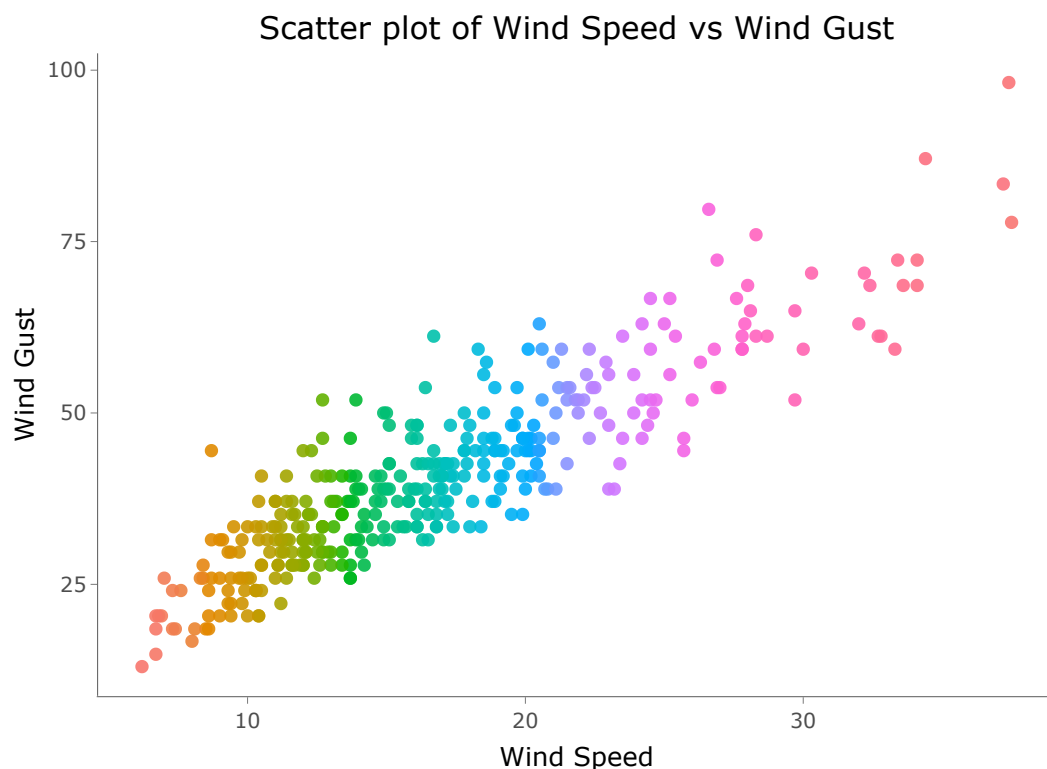

Scatter plot of Wind Speed vs Wind Gust

The temperature data contains the columns such as wind speed and wind gust which indicates the speed of the wind and a sudden increase in the speed of the wind. Here a scatter plot was used to see the relationship between the wind speed and wind gust. It is observed from the graph that there is a linear relationship between the wind speed and wind gust which indicates that increase in one of these two will increase the other as well. The graph shows the points are slopping upwards which shows there is a positive relationship between the parameters. So, it is understood that, with change in one of wind speed or wind gust the other will change in the same direction.

Box plot to check humidity over the months in 2023

```r
#Preparing data
humidity_df <- climate_df
humidity_df$Date <- as.character(humidity_df$Date)
humidity_df$Date <- as.Date(humidity_df$Date)
humidity_df$Date <- format(humidity_df$Date, "%y-%m")

#Box plot using ggplot
humidity_gg_box_ <- ggplot(humidity_df,aes(x=Date,y=HrAvg, fill=Date)) +
  geom_boxplot(alpha = 0.9) +
  scale_x_discrete(labels=c("Jan", "Feb", "Mar", "Apr", "May","Jun", "Jul", "Aug","Sep", "Oct", "Nov","De
c")) +
  labs(title = "Average Relative Humidity in Colchester by Month in 2023", x = "Year 2023", y = "Average R
elative Humidity") +
  theme_classic()

#Removing the legend
humidity_gg_box_ <- humidity_gg_box_ + guides(fill=FALSE)

#Interactive plot using ggplotly
humidity_plotly_ <- ggplotly(humidity_gg_box_,height = 450,width = 700)

#Customizing the title using layout
humidity_plotly_ <- layout(humidity_plotly_, title = list(text = "Average Relative Humidity in Colchester
by Month in 2023", x = 0.5, y = 0.97))

#Plotting box plot
humidity_plotly_
```
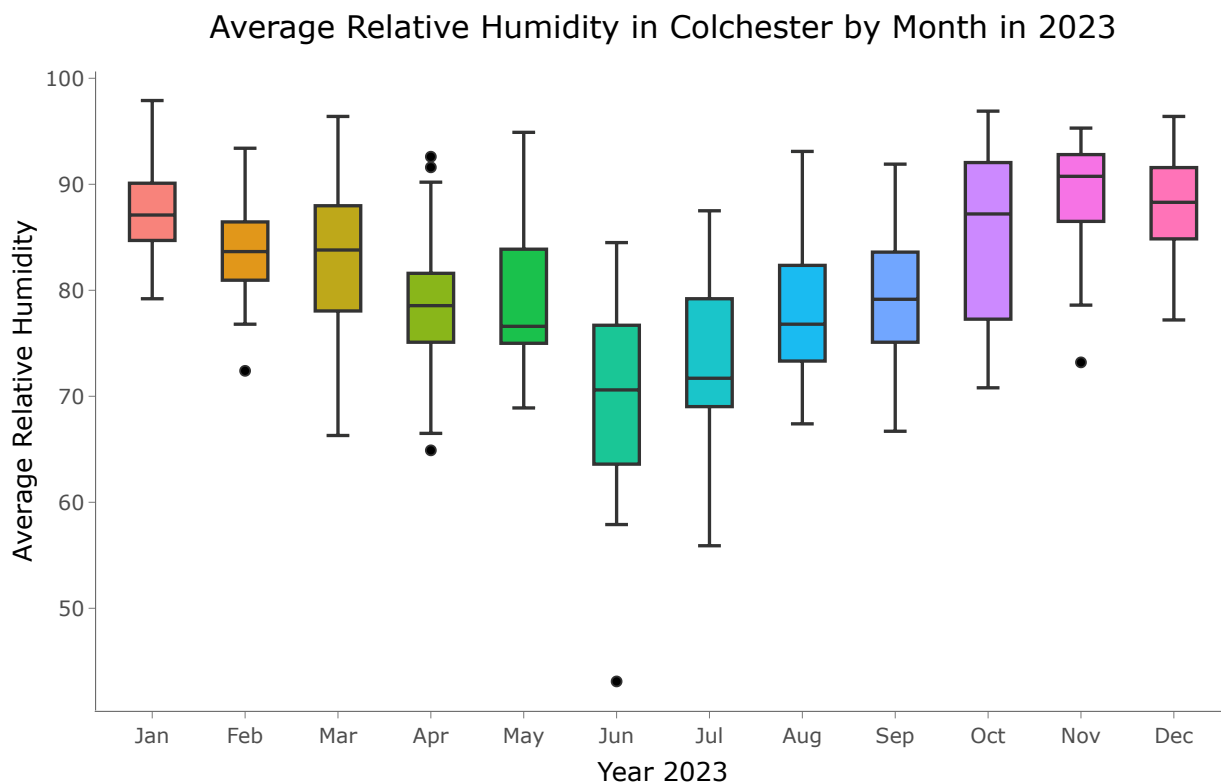
Average Relative Humidity in Colchester by Month in 2023

The temperature dataset contains average relative humidity column which is a common measure of humidity. It is used to measure how much water vapour is there in the air compared to how much there could be at that specific temperature. Here a box plot was used to visualize and understand the range of relative humidity in each month through the year 2023 in Colchester. From the above box plot, it was observed that the highest relative humidity was recorded in the month of January and lowest was recorded in the month of July. Another observation was most of the months recorded the relative humidity in the range of lower fence and upper fence. But, the box plots of relative humidity for some of the months indicates the the data has outliers which indicates on some days of the month recorded the relative humidity outside the range compared to other days of the month.

Correlation Plot to check the relationship between the variables in climate data

```
#Filtering climate data to plot correlation plot
cor_df <- climate_df[, c("TemperatureCAvg", "TdAvgC", "HrAvg", "WindkmhInt", "WindkmhGust", "PresslevHp",
"TotClOct", "VisKm")]

#Compute the correlation matrix
cor_mat_ <- round(cor(cor_df), 1)

#Reordering the correlation matrix using hierarchical clustering
cor_gg_plot_ <- ggcorrplot(cor_mat_, hc.order = TRUE, lab = TRUE, type = "lower") + labs(title = "Correlat
ion matrix of Climate data")

#Interactive correlation map using ggplotly
cor_plotly_ <- ggplotly(cor_gg_plot_,height = 450, width = 800)

#Customizing the title with layout
cor_plotly_ <- layout(cor_plotly_, title = list(text = "Correlation matrix of Climate data", x = 0.5, y =
0.97))

#Plotting the correlation map
cor_plotly_
```
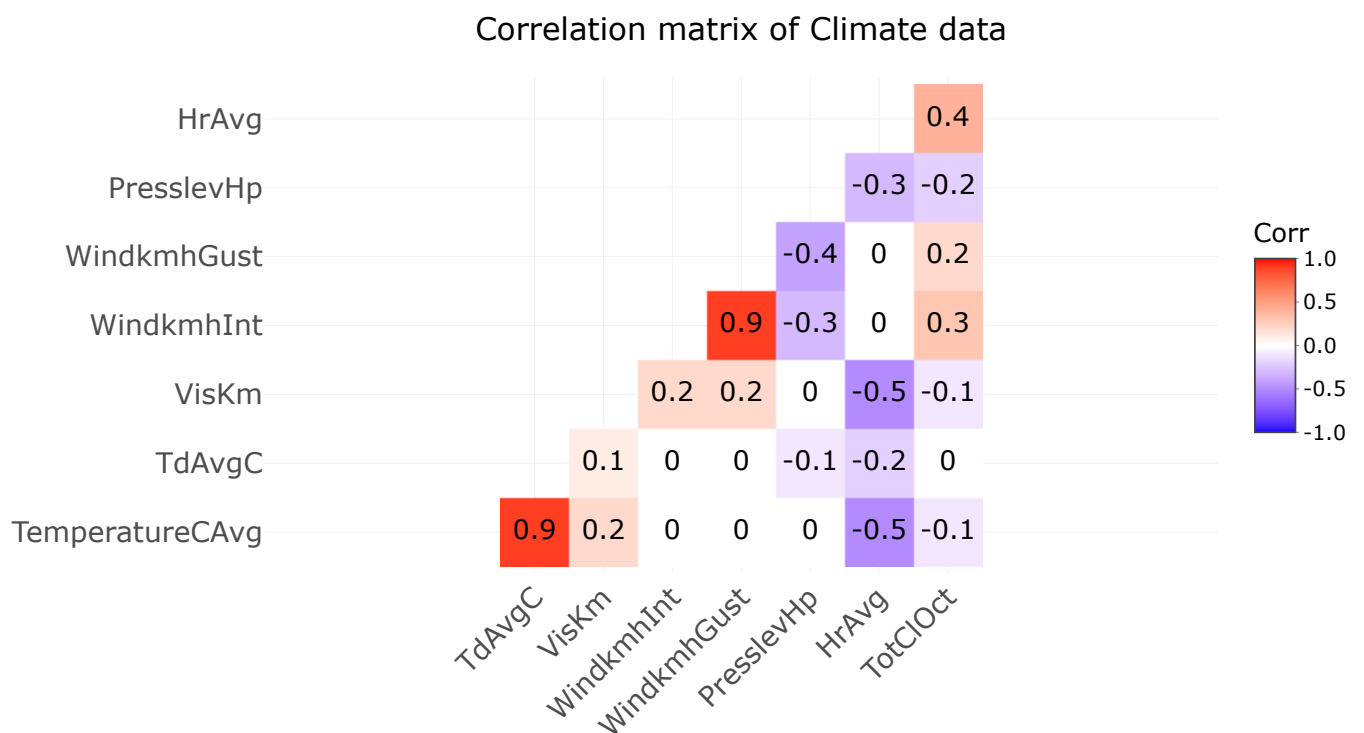

Correlation matrix of Climate data

The correlation map was used to check the relation between different variables present in the temperature data and number of crimes. It is observed from the above correlation matrix that the temperature was highly positively correlated with average dew point temperature and negatively correlated with relative humidity. It is also observed that relative humidity is negatively correlated with temperature and visibility. Another observation was there was no relationship between temperature and wind speed and sea level pressure.
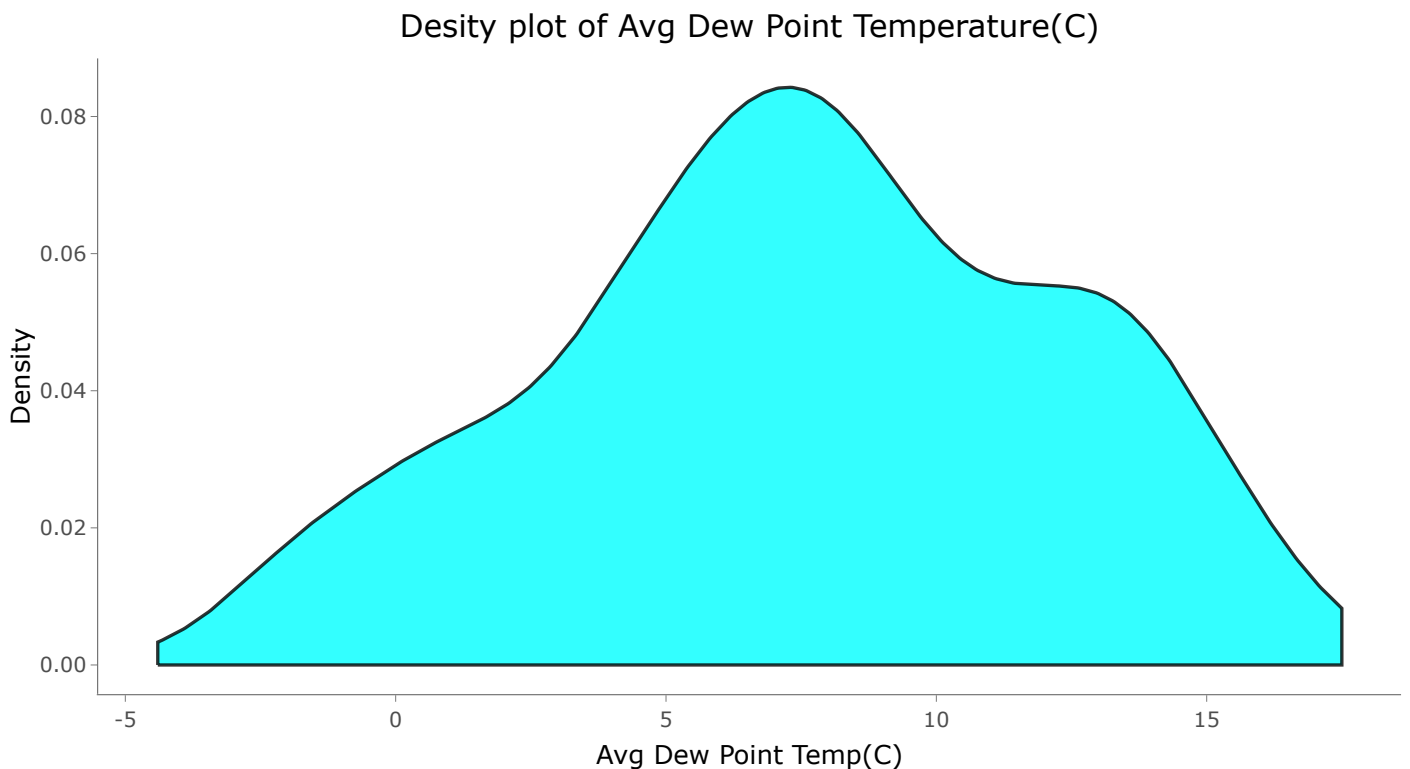
Density plot to check the Avg Dew Point Temperature distribution over the months in 2023

```
#Density plot using ggplot
avgdew_temp_gg_ <- ggplot(climate_df, aes(x=TdAvgC)) +
  geom_density(fill="cyan", alpha=0.8) +
  labs(title = "Desity plot of Avg Dew Point Temperature(C)", x ="Avg Dew Point Temp(C)", y = "Density") +
  theme_classic()

#Interactive plot using ggplotly
avgdew_temp_plotly_ <- ggplotly(avgdew_temp_gg_, height = 450, width = 800)

#Customizing the title using layout
avgdew_temp_plotly_ <- layout(avgdew_temp_plotly_, title = list(text = "Desity plot of Avg Dew Point Tempe
rature(C)", x = 0.5, y = 0.95))

#Plotting density plot
avgdew_temp_plotly_
```


Desity plot of Avg Dew Point Temperature(C)

In general, a density plot is used to understand the distribution of the continuous variable. Here the analysis was carried out using density plot to check the data distribution of the average dew point temperature measured in degree celsius over the year 2023 in Colchester. Average dew point temperature is the temperature at which the air becomes saturated with water. It is observed from the density plot that, many days in the year recorded dew point temperature in the range of 5 to 13 degree celsius. Another observation was the plot is skewed towards the right which indicates most of the days in the 2023 recorded moderate to high dew point temperature.

Visualizing the distribution of total cloudniess by month using box plot

```
#Preparing the data
cloud_df <- climate_df
cloud_df$Date <- as.Date(cloud_df$Date)

#Ordering data based on date
cloud_df$Date <- format(cloud_df$Date, "%b")
cloud_df$Month <- factor(cloud_df$Date, levels = c(as.list(month.abb)))

#Violin plot using ggplot
cloud_gg_violin_ <- ggplot(cloud_df,aes(x=Month, y=TotClOct, fill=Month)) +
  geom_violin(alpha = 0.9) +
  labs(title = "Total Cloudiness by month in Colchester in 2023", x = "Year 2023", y = "Total Cloudiness")
+
  stat_summary(fun = median, geom='point') +
  theme_bw()

#Converting ggplot into interactive plot
cloud_violin_plotly_ <- ggplotly(cloud_gg_violin_, height = 475, width = 800)

#Modifying hover text in interactive plot
#labels = paste("Month :", cloud_df$Date, "<br>TotalCloudOct :", cloud_df$TotClOct)
#cloud_violin_plotly_ <- style(cloud_violin_plotly_, text= labels)

#Modifying layout of the interactive plot
cloud_violin_plotly_ <- layout(cloud_violin_plotly_, title = list(text = "Total Cloudiness by month in Col
chester in 2023", x = 0.5, y = 0.95))

#Plotting the graph
cloud_violin_plotly_
```
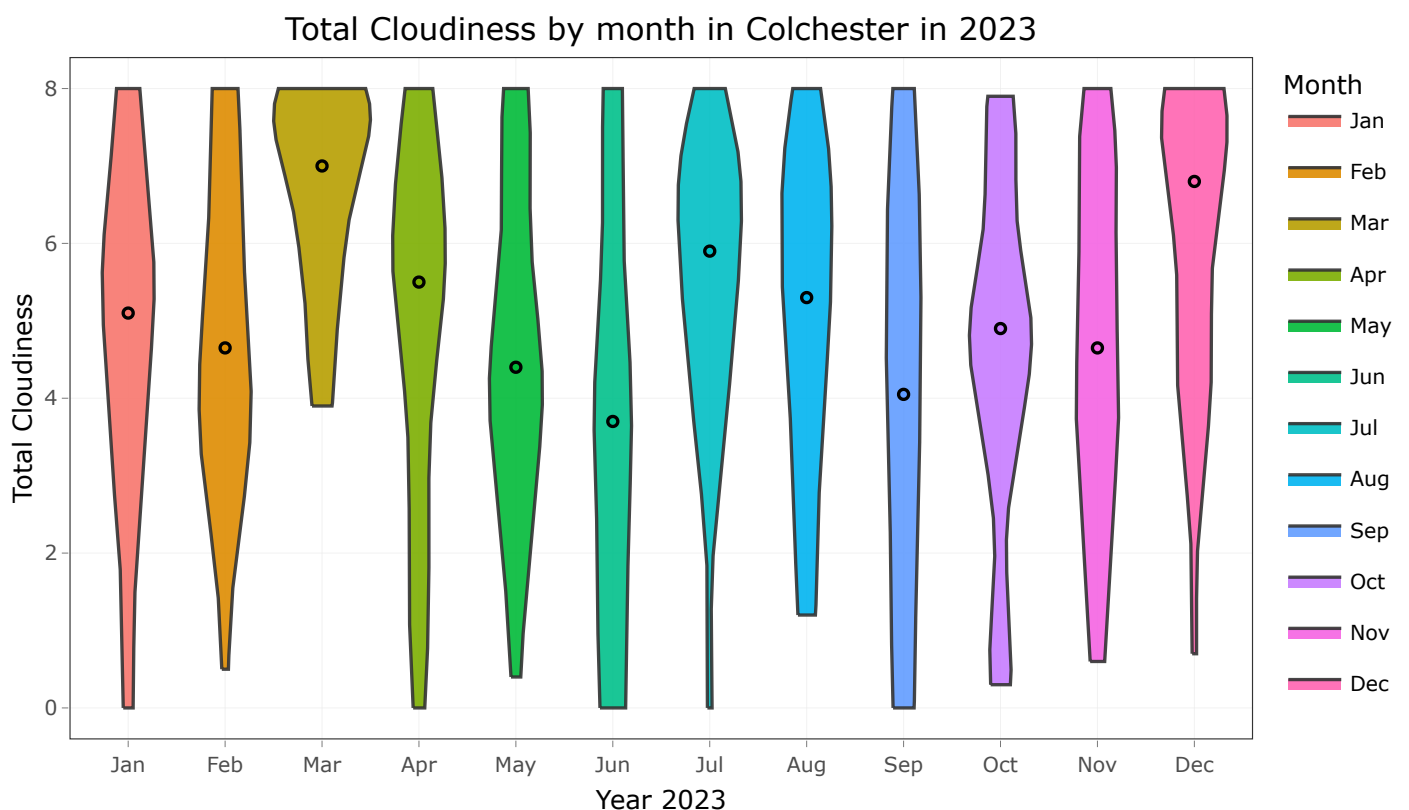

Total Cloudiness by month in Colchester in 2023

The temperature dataset contains the information about total cloudiness in Colchester in 2023. Here a violin plot is used to check the distribution of the total cloudiness by month. The width of the violin represents the density of a particular data point which means more the width of the violin more the density. It is observed from the above plot that the width of the violin plot is more which indicates that many days in the month of march recorded high total cloudiness. The point in each violin represents the median total cloudiness of that particular month. Another observation was in march month, the minimum total cloudiness was recorded on day in march month was 3.9 octants which is higher compared to other months.

Visualizing the relationship between tempearture, humidity and clodiness

```
#Creating variables
AverageTemperature <- climate_df$TemperatureCAvg
RealtiveHumidity <- climate_df$HrAvg
Cloudiness <- climate_df$TotClOct

#Plotting pair plot
pairs(cbind(AverageTemperature, Cloudiness, RealtiveHumidity),cex=0.6,pch=19, col = c('#c70630', 'royalblu
e', 'darkgreen'),
                 main = "Pair plot of Tempearture, Humidity and Clodiness")
```

Pair plot of Tempearture, Humidity and Clodiness

To understand
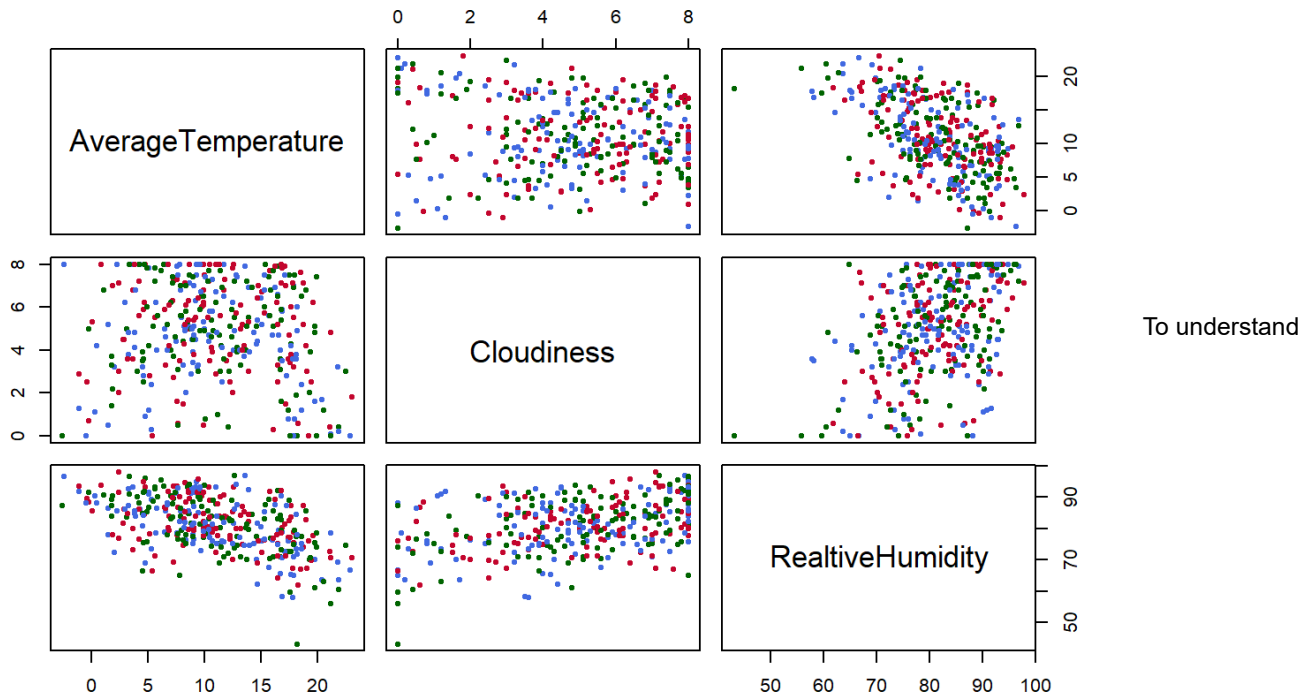
the relation between the temperature, humidity and cloudiness the analysis was performed using pair plot. It was observed from the plot that temperature is negatively correlated with humidity and humidity is positively correlated with total cloudiness. It was also observed that there is no relation between temperature are cloudiness.

Checking the relationship between Precipitation and Visibility using scatter plot

```
#Creating Scatter plot object using ggplot
precp_gg_plot<- ggplot(climate_df,aes(x = Precmm, y = VisKm, color = factor(VisKm))) +
  geom_point(alpha = 0.9)+
  labs(title = "Scatter plot of Precipitation vs Visibility", x="Precipitation(mm)", y="Visibility(km)") +
  theme_classic()

#Removing legend
precp_gg_plot <- precp_gg_plot + guides(color=FALSE)

#Interactive plot using ggplotly
percp_plotly_ <- ggplotly(precp_gg_plot,height = 450,width = 600)

#Modifying hover text in interactive plot
labels_ = paste("Precipitation(mm) :", climate_df$Precmm, "<br>Visibility(km) :", climate_df$VisKm)
percp_plotly_ <- style(percp_plotly_, text= labels_)

#Customizing the title using layout
percp_plotly_ <- layout(percp_plotly_, title = list(text = "Scatter plot of Precipitation vs Visibility",
x = 0.5, y = 0.95))

#Plotting the scatter plot
percp_plotly_
```
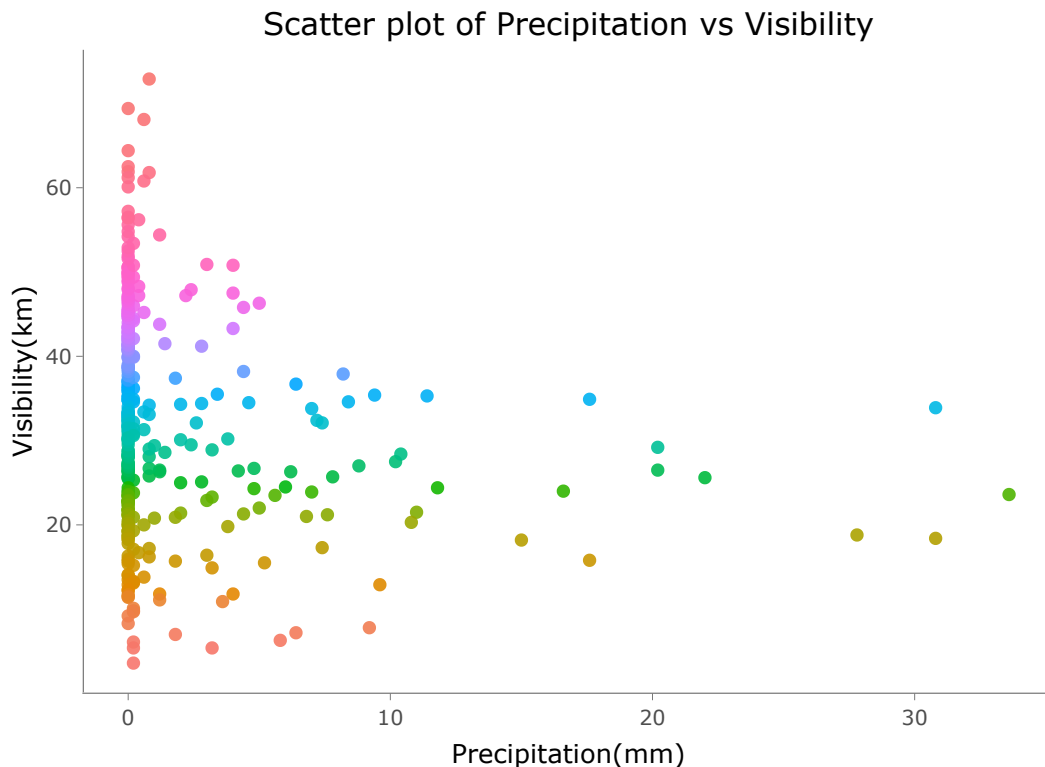

Scatter plot of Precipitation vs Visibility

Precipitation is a form of water and it includes rain, snow, sleet and hail. To understand the relationship between the precipitation and visibility analysis was carried out using scatter plot. It was observed from the plot that when precipitation is high then the visibility is less. So, precipitation is inversely proportional to visibility and it is negatively correlated to visibility.

Analysis to find out relation between crimes and temperature

In general crime rate and temperature are closely related. So, to find out the relation between them both the datasets were combined here to perform analysis. The temperature data has daily data and this data was grouped using date. Then the crime count and temperature data was combined using date to carry out further analysis using different data visualization methods.

Checking the relationship between crimes and temperature using time series plot

```
#Preparing temperature data
temp_ts_df <- climate_df
temp_ts_df$Date <- as.character(temp_ts_df$Date)
temp_ts_df$Date <- as.Date(temp_ts_df$Date)
temp_ts_df$Date <- format(temp_ts_df$Date, "%y-%m")
temp_ts_df_ <- temp_ts_df %>% group_by(Date) %>% summarise(AvgTemp =round(mean(TemperatureCAvg),2))

#Preparing data for time series plot
temp_dates_ <- ym(temp_ts_df_$Date)
temp_ts_ <- xts(data.frame(Temperature = temp_ts_df_$AvgTemp), temp_dates_)

#Timeseries plot using autoplot
temp_ts_plot_ <- autoplot(temp_ts_, facet=NULL)+geom_point(data = temp_ts_df_, aes(x = temp_dates_, y = Av
gTemp, text = paste("Month : ", format(temp_dates_, "%m"), "<br>Avg Temp(C) : ", AvgTemp)), color = "red",
size = 3, alpha = 0.9) + labs(title = "Average Temperature in Colchester by Month in 2023", x = "Year 202
3", y = "Avg Temp(C)") + theme_bw()

#Interactive plot using ggplotly
temp_plotly_ts <- ggplotly(temp_ts_plot_,tooltip = "text", height = 450, width = 800)

#Customizing the hoverlabel
temp_plotly_ts <- temp_plotly_ts %>% layout( hoverlabel = list(bgcolor = "cyan", font = list(color = "blac
k")))

#Customizing the title using layout
temp_plotly_ts <- layout(temp_plotly_ts, title = list(text = "Average Temperature in Colchester by Month i
n 2023", x = 0.5, y = 0.95))
```
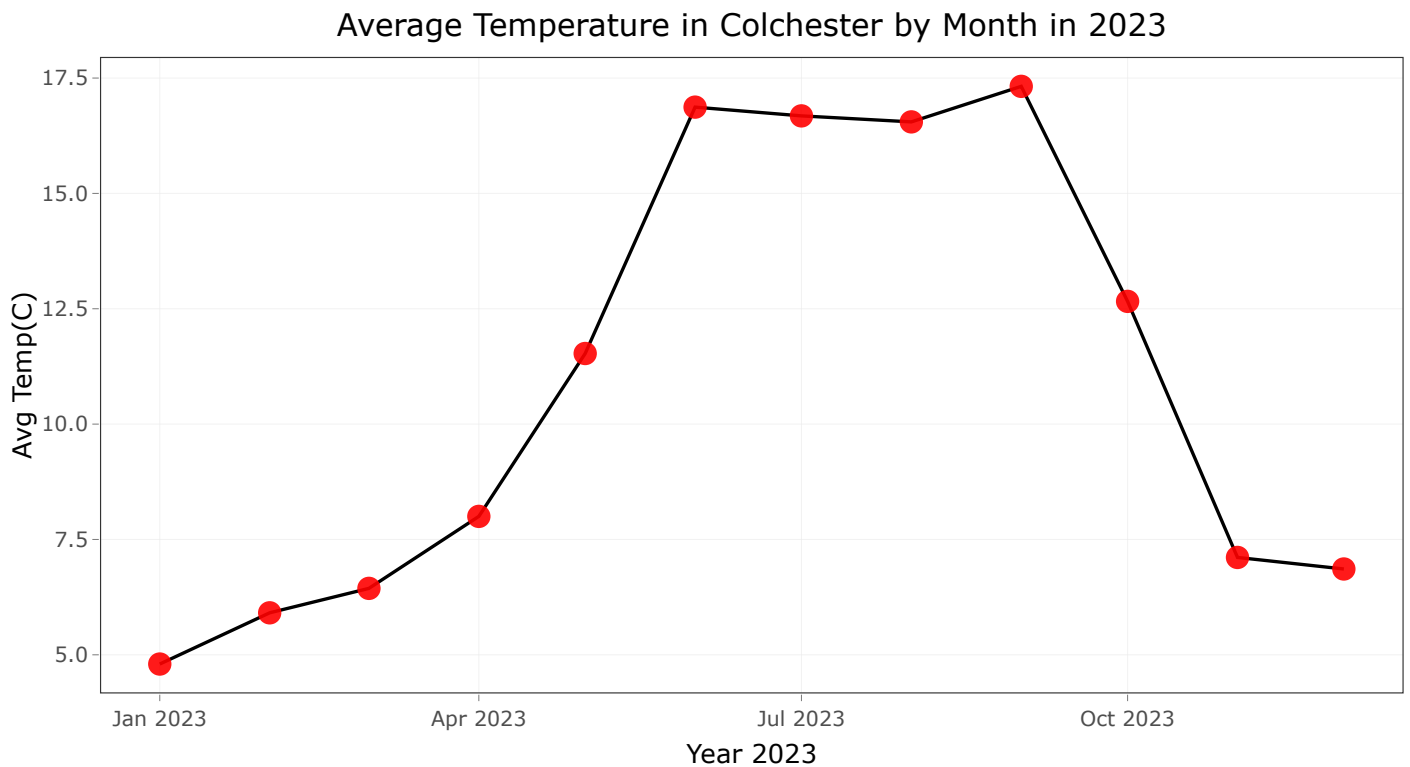
Temperature time series plot

```
#Plotting the time series interactive plots
temp_plotly_ts
```
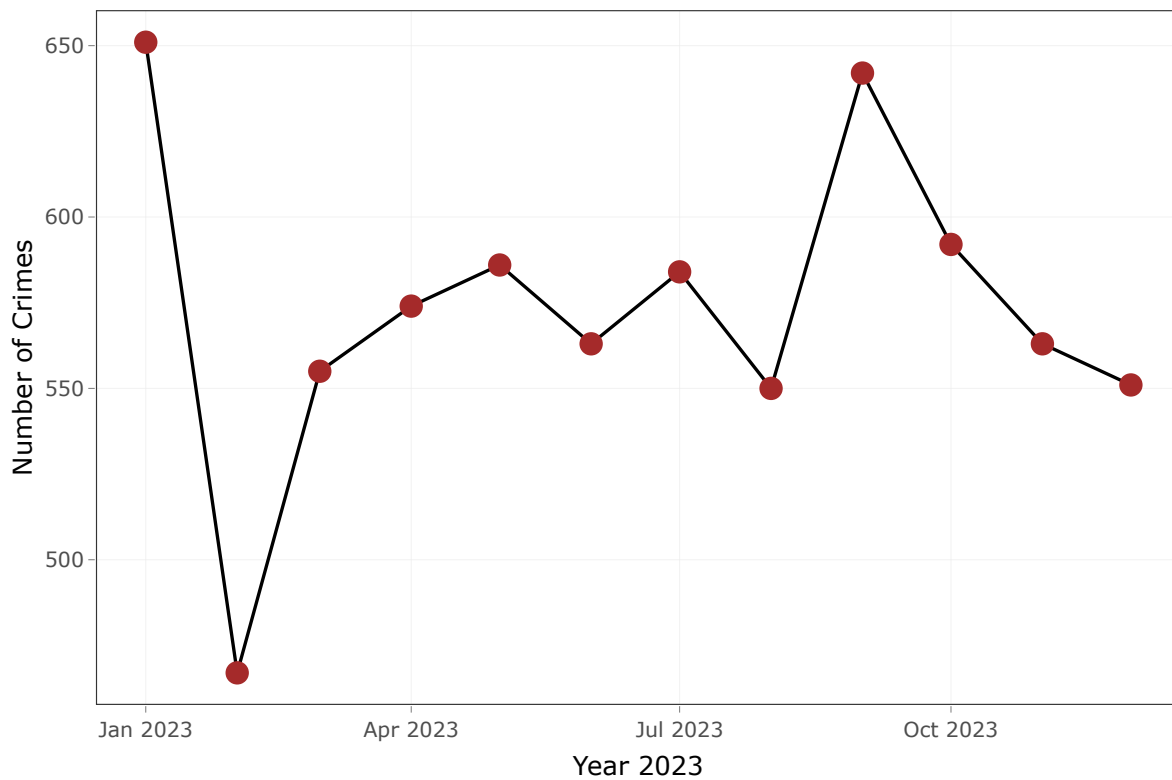

Average Temperature in Colchester by Month in 2023

Crimes time series plot

```
#Plotting the time series interactive plots
crime_plotly_ts
```

Number of Crimes in Colchester by Month in 2023

The analysis is performed here to see if there is any relationship between the temperature and the number of crimes that took place in Colchester in 2023. To carry out this analysis, time series plots were used to check the number of crimes by month and the average temperature in each month of year 2023. From the above time series plots, it was observed that the crime rate is more in the months of January and September. The observation from temperature time series plot was January month reported the least temperature and the temperature in the month of September is highest compared to other months. The key observation from both the time series plots was the crime rate was more in the months which recorded least and highest temperature i.e., in the months of January and September. Another observation was the crime rate started increasing from March and peaked in September where the temperature also increased from the month of March and continued till September.

Correlation Plot to check the relationship between the crimes and climate data

```
#Filtering climate data to plot correlation plot
climate_df_copy_ <- climate_df
corr_df <- climate_df_copy_[, c("TemperatureCAvg", "TdAvgC", "HrAvg", "WindkmhInt", "WindkmhGust", "Pressl
evHp", "TotClOct", "VisKm", "Date", "Precmm")]

#Grouping data by month
corr_df$Date <- as.Date(corr_df$Date)
corr_df$Date <- format(corr_df$Date, "%b")
corr_df$Date <- factor(corr_df$Date, levels = c(as.list(month.abb)))

corr_df_ <- corr_df%>% group_by(Date) %>%
  summarise(AvgTemp =round(mean(TemperatureCAvg),2), AvgDewTemp =round(mean(TdAvgC),2),
          AvgHumidity =round(mean(HrAvg),2), AvgWindSpd =round(mean(WindkmhInt),2),
          AvgWindGust =round(mean(WindkmhGust),2), AvgPress =round(mean(PresslevHp),2),
          AvgTotClOct =round(mean(TotClOct),2), AvgVis =round(mean(VisKm),2),
          AvgPercmm =round(mean(Precmm, na.rm = TRUE),2))
corr_df_$Crimes <- crimes.ts.df$total_crimes_

#Deleting date column
corr_df_ <- corr_df_[, -which(names(corr_df_) == 'Date')]

#Compute the correlation matrix
corr_mat_ <- round(cor(corr_df_), 1)

#Reordering the correlation matrix using hierarchical clustering
corr_gg_plot_ <- ggcorrplot(corr_mat_, hc.order = TRUE, lab = TRUE, type = "lower") + labs(title = "Correl
ation matrix of Climate data and Number of crimes")

#Interactive correlation map using ggplotly
corr_plotly_ <- ggplotly(corr_gg_plot_,height = 450, width = 800)

#Customizing the title with layout
corr_plotly_ <- layout(corr_plotly_, title = list(text = "Correlation matrix of Climate data and Number of
crimes", x = 0.5, y = 0.97))

#Plotting the correlation map
corr_plotly_
```
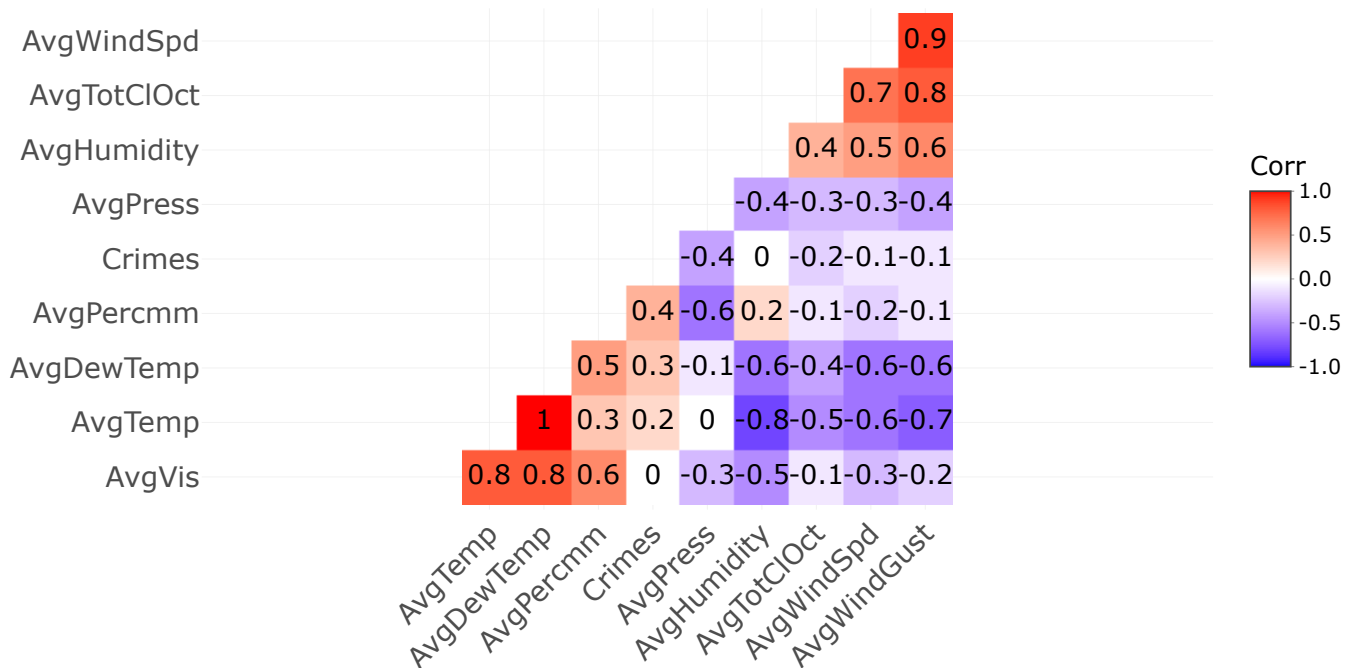

Correlation matrix of Climate data and Number of crimes

The correlation map was used to check the relation between different variables present in the temperature data and number of crimes. It is observed from the above correlation matrix that the temperature was highly positively correlated with average dew point temperature and negatively correlated with relative humidity. It is also observed that relative humidity is negatively correlated with temperature and visibility. Another observation was there was no relationship between temperature and wind speed and sea level pressure. The key observation from the above correlation matrix was number of crimes was positively correlated with precipitation. The positive correlation indicates change in one parameter changes the other in the same direction. So, there was a chance of raise in crime rate when precipitation increased. It was also observed that there was no correlation between crime rate and visibility, wind speed, wind gust, relative humidity.

Summary

In this report exploratory data analysis was carried out on two datasets namely crime dataset and temperature datasets which has data about crimes that happened in colchester over the year in 2023 and the climate condition in 2023. Initially the analysis was carried out on both the datasets individually to find out the relationship between the variables. Crime data has information about the different types of crime categories that occurred in Colchester. It has the information such as the date of the crime happened, street name, location co-ordinates, location type, street number and the outcome status of the crime. First the different crime categories and the frequency of the crimes were explored using a bar plot and it is found that the Violent crimes and Anti-social-behaviour crimes took place many times over the year. To further investigate about these top 2 crimes categories, a time series plot was used to check the frequency of these two crimes categories by each month over the year. It is observed that the every month in 2023 recorded around 200 violent crimes. In case of anti social behaviour crimes the highest crimes were recorded in the month of September. Then the analysis was carried to check the number of crimes that happened in different location types. It is found that around 99 percent of the crimes were occurred in Force location type and the remaining 1 percent in BTP location type. Bar plot was used to further check the crime categories and their frequency. It was observed that in the both location types violent crimes were the majority of the crimes and in BTP location type only 5 types of crimes took place whereas in force location type 14 types of crimes occurred. Then a map was used to locate the number of crimes that happened in each street. It was found that crime rate was more near shopping area and cowdray avenue compared to other streets. To further investigate the types of crimes that took place in the above mentioned streets pie chart was used and it is found that near shopping area most the crimes occurred were shop lifting crimes and near cowdray avenue violent crimes and bicycle thefts were the majority crimes that took place. Then a time series plot was used to find out the frequency of crimes in each month over the year and it was observed that the crime rate was more in the months of January and September. To see the types of crimes and its their frequency in these two months, bar plot was used and it was observed that the violent crimes were the majority crimes that happened in both the months. Further the outcome status of the crimes were visualized using pie chart and it was observed that for around 40 percent of the crimes, the investigation was completed and no suspect was identified. For around 30 percent of the crimes the investigation status was unable to prosecute the suspect.

Then the EDA was carried out on the temperature dataset. It has the information about the daily climate conditions in Colchester in 2023 such as temperature, relative humidity, dew point temperature, wind speed, wind direction, sea level pressure and visibility. First the data distribution of the average temperature was checked using histogram. It was observed from the graph that most of the days recorded an average temperature in the range of 7 to 15 degree celsius. Then the distribution of minimum temperature and maximum temperature in year 2023 were checked using time series plot. To check the monthly average smoothing was used and it is found that the highest temperature was recorded in the months of June and August and the lowest temperature was recorded in the months of January, February and December. Further to check the relationship between wind speed and wind gust a scatter plot was used and it was observed from the plot that there is a linear relationship between the wind speed and wind gust. In the plot the points are slopping upwards which shows there is a positive relationship between them. So, it is understood that, with change in one of wind speed or wind gust the other will change in the same direction. Further to find out the range of relative humidity in each month a box plot was used and it shows that relative humidity range is different in each month. It was observed that the highest relative humidity was recorded in the month of January and lowest was recorded in the month of July. The data distribution of the average dew point temperature was visualized using density plot. It is observed from the density plot that, many days in the year recorded dew point temperature in the range of 5 to 13 degree celsius and the plot is skewed towards the right which indicates most of the days in the 2023 recorded moderate to high dew point temperature. Further the analysis was carried out to check total cloudiness by month using violin plot. In violin plots the width of the violin represents the density of data points. The width of the march month violin is more which indicates the many days in march recorded high total cloudiness. It is observed that the minimum total cloudiness was recorded on day in march month was 3.9 octants which is higher compared to other months. The relation between the temperature, humidity and cloudiness was analysed using pair plot. It was observed from the plot that temperature is negatively correlated with humidity and humidity is positively correlated with total cloudiness. It was also observed that there is no relation between temperature are cloudiness. To understand the relationship between the precipitation and visibility analysis was carried out using scatter plot. It was observed from the plot that when precipitation is high then the visibility is less. So, precipitation is inversely proportional to visibility and it is negatively correlated to visibility.

The analysis was further carried out to check if there is any relationship exists between the crime rate and temperature. Different plots were used to explore if change in climate conditions affected the crime rate in Colchester. First a time series plot was used to check the relation between temperature and crime rate. It was observed that the crime rate is more in the months of January and September. The observation from temperature time series plot was January month reported the least temperature and the temperature in the month of September is higher compared to other months. The key observation from both the time series plots was the crime rate was more in the months which recorded least and highest temperature i.e., in the months of January and September. Another observation was the crime rate started increasing from March and peaked in September where the temperature also increased from the month of March and continued till September. To further understand the relationship between the number of crimes and temperature a correlation plot was used. The key observation from the correlation matrix was number of crimes was positively correlated with precipitation. The positive correlation indicates change in one parameter changes the other in the same direction. So, there was a chance of raise in crime rate when precipitation increased. It was also observed that there was no correlation between crime rate and visibility, wind speed, wind gust and relative humidity. To conclude, some of the weather parameters had impact on the crime rate. It was observed from the correlation plot that precipitation has significant impact on the crime rate in Colchester in 2023 and the other weather parameters had weak correlation with the crime rate which indicates there is no impact on the crime rate.

References

[1]. https://www.rdocumentation.org/packages/plotly/versions/4.10.4/topics/ggplotly
(https://www.rdocumentation.org/packages/plotly/versions/4.10.4/topics/ggplotly)

[2]. https://ggplot2.tidyverse.org/articles/ggplot2-specs.html (https://ggplot2.tidyverse.org/articles/ggplot2-specs.html)

[3]. https://www.rdocumentation.org/packages/leaflet/versions/2.2.2
(https://www.rdocumentation.org/packages/leaflet/versions/2.2.2)

[4]. https://plotly.com/r/pie-charts/ (https://plotly.com/r/pie-charts/)

[5]. https://www.rdocumentation.org/packages/corrplot/versions/0.92/topics/corrplot
(https://www.rdocumentation.org/packages/corrplot/versions/0.92/topics/corrplot)