MA304 Assignment

Student Registration Number: 2322761

Using various data visualization tools, this report aims to analyse a policing data set from Colchester, England in 2023, an excerpt of which can be seen below:

```
crime_data <- read.csv("crime23.csv")
head(crime_data)
```

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

General Overview

Frequency of each category of crime

The data set gives information about 6878 observations on 12 different variables: the category of the crime, the persistent id (i.e. the unique identifier of that crime), the date of the crime, latitude, longitude, street ID, street name, context (extra information about the crime), ID of the crime (for the API), location type (i.e. whether it took place at a normal police force location or a British Transport location), location subtype, and the outcome status. To get a better picture of the crime data, a bar chart was plotted showing the different crime types and their relative frequencies.

```
library(dplyr)
```

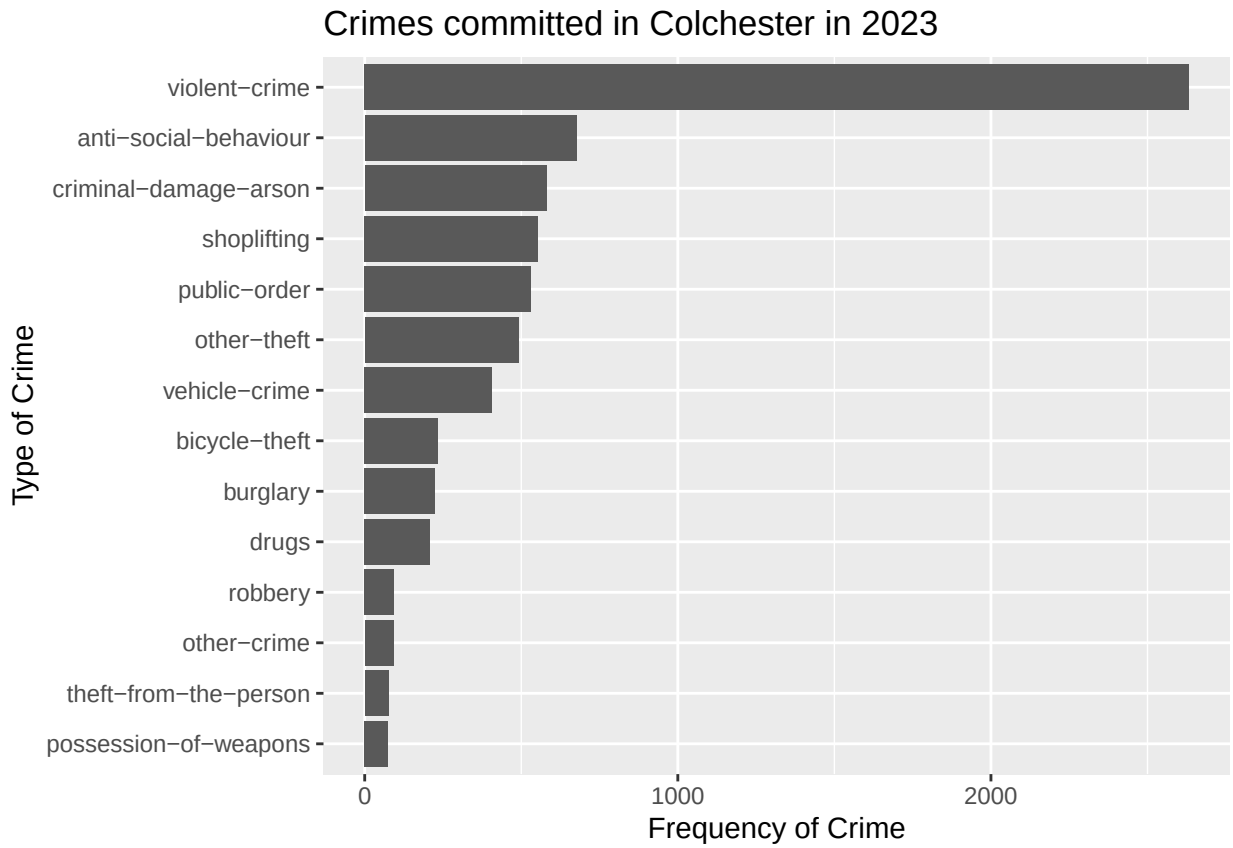
```
##
## Attaching package: 'dplyr'

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

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

```r
library(ggplot2)

crime_data %>%
  count(category) %>%
  ggplot(aes(x = reorder(category, n),
             y = n)) +
  geom_col() + labs(x = "Type of Crime",
                    y = "Frequency of Crime",
                    title = "Crimes committed in Colchester in 2023") + coord_flip()
```



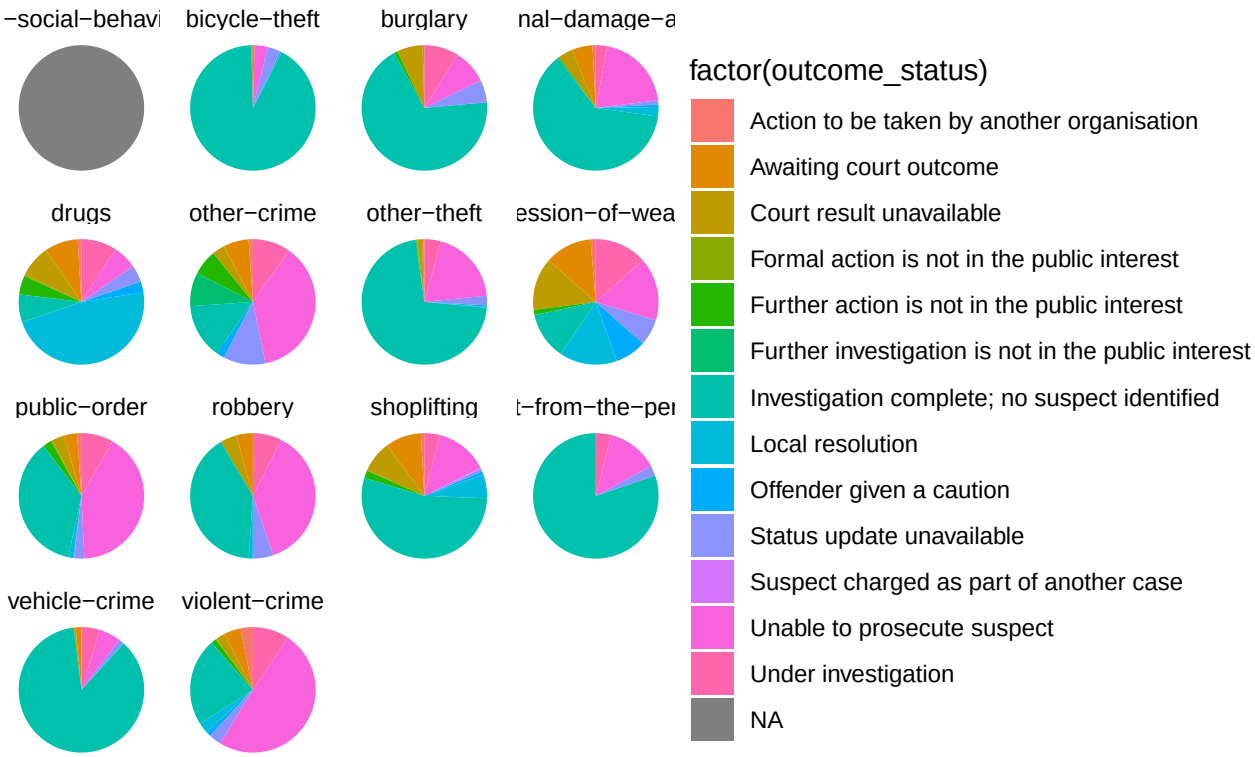
The bar chart shows that, by far, the most common type of crime was violent crime, with approximately 2600 incidents that year - a large contrast in comparison to the next most common type of crime, anti-social behavior, which had approximately 700 incidents reported that year. This could be due to the fact that

violent crime encompasses a range of offenses, from common assault and Actual Bodily Harm, all the way to murder (1), and so it would be interesting to see the frequencies for each category which falls within violent crime to get a better idea of what is meant exactly by 'violent crime'.

Proportion of outcomes

```
#create pie charts

ggplot(crime_data, aes(x="",  fill = factor(outcome_status))) +
  geom_bar(stat = "count", width = 1, position =  "fill") +
  coord_polar("y", start=0) +
  theme_void() +
  facet_wrap(~ category, nrow = 4)
```



The pie charts show the proportion of the different outcomes for each type of crime. They show that the categories burglary, arson, vehicle crime, bicycle theft, theft from the person, and "other theft", are most likely to complete the investigation without finding any suspect. This could be due to the fact that these categories all involve the perpetrator leaving the site of the offence immediately afterwards, meaning it is a lot harder to find the offender. For bicycle theft, theft from the person and vehicle crime in particular, where the damage is relatively minimal, lack of time and police resources could also contribute to being unable to identify the suspect as it would be too costly to find the perpetrators for such small crimes.

Unfortunately, despite having the highest frequency of the crimes, the pie charts also show that violent crime has the highest proportion of the category 'unable to prosecute suspect', with around 50% of all violent crime

resulting in this outcome. This could be due to the fact that a lot of offences that fall under this category (e.g. common assault, Actual Bodily Harm, Grievous Bodily Harm) rely on hearsay if the incident had occurred somewhere without cameras, and therefore may be insufficient evidence to prosecute the suspect.

Frequency of location types

```
attach(crime_data)

#calculate frequencies for each location type
Force_freq <- sum(crime_data$location_type == "Force")
BTP_freq <- sum(crime_data$location_type == "BTP")

#create data frame
Location_type_df <- data.frame(
  Location_Type = c("Force", "BTP"),
  Frequency = c(Force_freq, BTP_freq))

knitr::kable(Location_type_df,align = "cccc",caption="Table showing the frequencies of each location ty
```

Table 1: Table showing the frequencies of each location type.

Location_Type	Frequency
Force	6854
BTP	24

The frequency table shows the number of crimes that took place on a normal police force location (Force) and the number that took place on a British Transport location. As one would expect, there is a much higher number that took place on a normal police force location, as generally this tends to cover a much larger area. A more detailed analysis on the locations and the relationships to other variables in the crime data is covered below.

Location and Crime Data

References

1.The Crown Prosecution Service. Violent Crime [Internet]. CPS. [cited 2024 Apr 18]. Available from: https://www.cps.gov.uk/crime-info/violent-crime

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, https://CRAN.R-project.org/package=dplyr.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

ChatGPT 3.5. OpenAI; 2022. Available from: https://openai.com/chatgpt