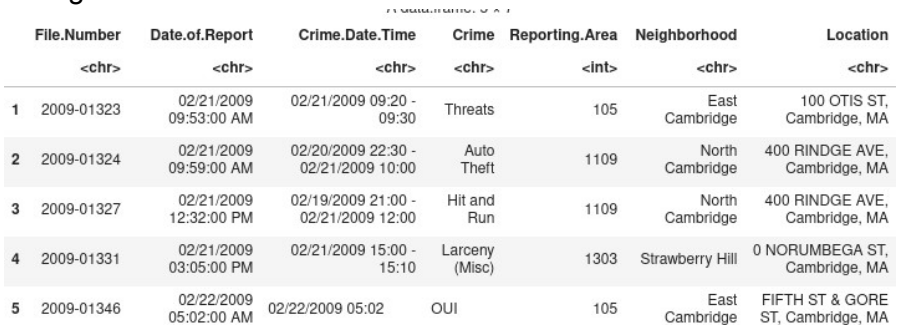
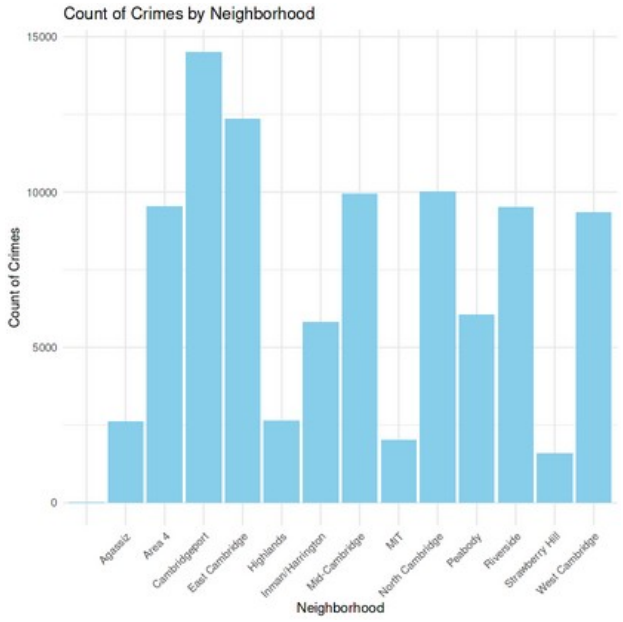
| | |
|---|---|
| Name | Aditya Khuman |
| UID | 2021300061 |
| Aim | Create basic charts using R programming language on dataset Crime or Police / Law and Order<br><br>Basic - Bar chart, Pie chart, Histogram, Time line chart, Scatter plot, Bubble plot<br><br>Write observations from each chart |
| Dataset | Cambridge Crime Data:<br>  |
| Charts andCode | Bar chart:<br># Load necessary libraries<br>library(tidyverse)<br><br># Bar chart for the count of crimes by neighborhood<br>ggplot(crime_data, aes(x = Neighborhood)) +<br>geom_bar(fill = "skyblue") +<br>  labs(title = "Count of Crimes by Neighborhood", x = "Neighborhood", y = "Count of Crimes") +<br>  theme_minimal() +<br>  theme(axis.text.x = element_text(angle = 45, hjust = 1)) |

| | File.Number | Date.of.Report | Crime.Date.Time | Crime | Reporting.Area | Neighborhood | Location |
|---|---|---|---|---|---|---|---|
| | <chr> | <chr> | <chr> | <chr> | <int> | <chr> | <chr> |
| 1 | 2009-01323 | 02/21/2009 09:53:00 AM | 02/21/2009 09:20 - 09:30 | Threats | 105 | East Cambridge | 100 OTIS ST, Cambridge, MA |
| 2 | 2009-01324 | 02/21/2009 09:59:00 AM | 02/20/2009 22:30 - 02/21/2009 10:00 | Auto Theft | 1109 | North Cambridge | 400 RINDGE AVE, Cambridge, MA |
| 3 | 2009-01327 | 02/21/2009 12:32:00 PM | 02/19/2009 21:00 - 02/21/2009 12:00 | Hit and Run | 1109 | North Cambridge | 400 RINDGE AVE, Cambridge, MA |
| 4 | 2009-01331 | 02/21/2009 03:05:00 PM | 02/21/2009 15:00 - 15:10 | Larceny (Misc) | 1303 | Strawberry Hill | 0 NORUMBEGA ST, Cambridge, MA |
| 5 | 2009-01346 | 02/22/2009 05:02:00 AM | 02/22/2009 05:02 | OUI | 105 | East Cambridge | FIFTH ST & GORE ST, Cambridge, MA |



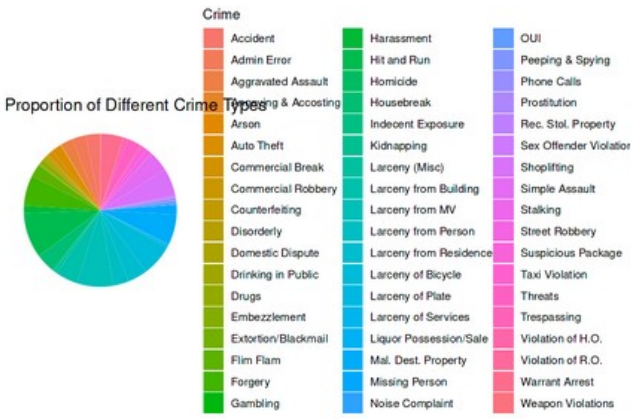

**Pie chart:**

```
# Pie chart of crime types
crime_type_counts <- crime_data
%>%group_by(Crime) %>%
summarise(count = n())

ggplot(crime_type_counts, aes(x = "", y = count, fill = Crime)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  labs(title = "Proportion of Different Crime Types", x = "", y = "") +
  theme_void() +
  theme(legend.position = "right")
```

**Observation**:
Cambridgeport has the highest number of crimes among the neighborhoods,followed closely by Inman/Harrington. Both neighborhoods show crime counts exceeding 12,000, making them significant areas of concern.

**Pie chart:**

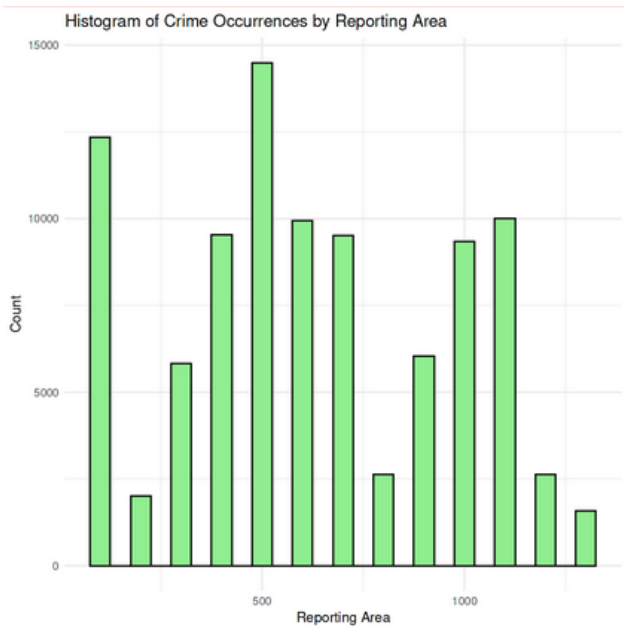



**Observation**:
This is showing the pie chart of proportion of different crimes. It includes various types of crimes.

**Histogram**:

```
# Histogram of crime occurrences by Reporting Area
ggplot(crime_data, aes(x = Reporting.Area)) +
geom_histogram(binwidth = 50, fill = "lightgreen", color = "black") +
labs(title = "Histogram of Crime Occurrences by Reporting Area", x =
"Reporting Area", y = "Count") +
  theme_minimal()
```

Conclusion: in this practical, using R, using crime dataset, i drawn basic charts

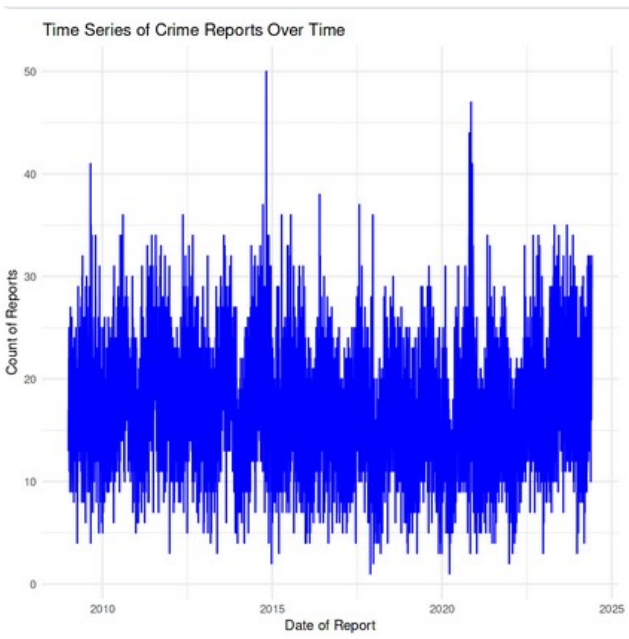



**Observation**:
The most significant peak is around the reporting area in the 400–500 range, with counts exceeding 14,000. This indicates a concentration of crimes in these reporting zones.
Another noticeable peak is seen around the reporting area in the 100–200range, with crime counts close to 12,000, suggesting these zones are alsoprominent for criminal activity.

**Time Series:**

```
# Convert 'Date.of.Report' to Date type
crime_data$Date.of.Report <- as.Date(crime_data$Date.of.Report, format = "%m/%d/%Y")

# Time series plot of crime reports over time
ggplot(crime_data, aes(x = Date.of.Report)) +
geom_line(stat = "count", color = "blue") +
  labs(title = "Time Series of Crime Reports Over Time", x = "Date of Report", y = "Count of Reports") +
  theme_minimal()
```

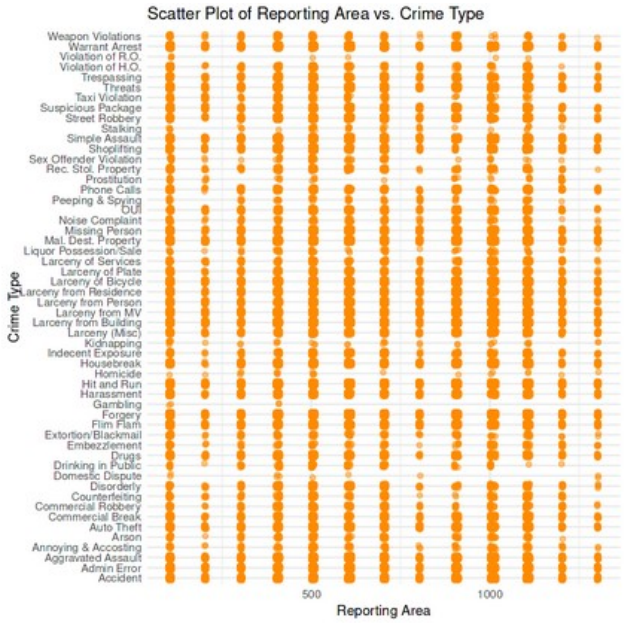



+ Code + Markdown

Observation: We can see that from 2010 to 2024, crime is reported alot. Some times we can see alot of crime reports.

Scatter plot:

```
# Scatter plot of Reporting Area vs. Crime
ggplot(crime_data, aes(x = Reporting.Area, y = Crime)) +
  geom_jitter(width = 0.2, height = 0.2, alpha = 0.5, color = "darkorange") +
labs(title = "Scatter Plot of Reporting Area vs. Crime Type", x = "Reporting Area", y = "Crime Type") +
  theme_minimal()
```




**Observation**:
We can see that on average in all areas all types of crimes are reported.
2.

```
# Load necessary libraries
library(tidyverse)

# Extract hour from 'Crime.Date.Time'
crime_data$Crime_Hour <-
as.numeric(format(strptime(crime_data$Crime.Date.Time, "%m/%d/%Y
%H:%M"), "%H"))

# Scatter plot of crime occurrences by hour of the day
ggplot(crime_data, aes(x = Crime_Hour, y = Crime)) +
geom_jitter(width = 0.2, height = 0.2, alpha = 0.5, color = "purple") +
labs(title = "Scatter Plot of Crime Occurrences by Hour of the Day",
           x = "Hour of the Day", y = "Crime Type") +
  theme_minimal()
```


Scatter Plot of Crime Occurrences by Hour of the Day
Crime Type
Weapon Violations
Warrant Arrest
Violation of R.O.
Violation of H.O.
Trespassing
Threats
Taxi Violation
Suspicious Package
Street Robbery
Stalking
Simple Assault
Shoplifting
Sex Offender Violation
Rec. Stol. Property
Prostitution
Phone Calls
Peeping & Spying
DUI
Noise Complaint
Missing Person
Mal. Dest. Property
Liquor Possession/Sale
Larceny of Services
Larceny of Plate
Larceny of Bicycle
Larceny from Residence
Larceny from Person
Larceny from MV
Larceny from Building
Larceny (Misc)
Kidnapping
Indecent Exposure
Housebreak
Homicide
Hit and Run
Harassment
Gambling
Forgery
Flim Flam
Extortion/Blackmail
Embezzlement
Drugs
Drinking in Public
Domestic Dispute
Disorderly
Counterfeiting
Commercial Robbery
Commercial Break
Auto Theft
Arson
Annoying & Accosting
Aggravated Assault
Admin Error
Accident