Name	Hatim Sawai
UID No.	2021300108
Course	Advanced Data Visualization

Experiment 4

Aim	Create basic charts using R programming language on dataset Crime or Police / Law and Order 1. Basic - Bar chart, Pie chart, Histogram, Time line chart, Scatter plot, Bubble plot 2. Write observations from each chart

1. Importing Libraries

In [1]:
```
import pandas as pd
import numpy as np
```

2. Loading Dataset

In [2]:
```
df = pd.read_csv('../Datasets/crime.csv', encoding='ISO-8859-1')
df.head()
```

Out[2]:

	incident_id	offense_id	offense_code	offense_code_extension	offense_type_id	c
0	202268791	202268791299900	2999	0	criminal-mischief-other	
1	2021387586	2021387586299900	2999	0	criminal-mischief-other	
2	2020641486	2020641486299900	2999	0	criminal-mischief-other	
3	2018612468	2018612468299900	2999	0	criminal-mischief-other	
4	2020293614	2020293614299900	2999	0	criminal-mischief-other	

3. Data Preprocessing

In [6]:
```
df.isnull().sum()
```

Out[6]:
```
incident_id             0
offense_id              0
offense_code            0
offense_type_id         0
offense_category_id     0
first_occurrence_date   0
reported_date           0
incident_address        0
geo_x                   0
geo_y                   0
district_id             0
neighborhood_id         0
is_crime                0
is_traffic              0
victim_count            0
dtype: int64
```

In [4]:
```
# drop columns: offense_code_extension, last_occurrence_date, geo_lon, geo_lat, pre
df.drop(['offense_code_extension', 'last_occurrence_date', 'geo_lon', 'geo_lat', 'p

# drop rows with missing values
df.dropna(inplace=True)
```

In [5]:
```
print(df.shape)
df.head()
```

```
(370666, 15)
```

Out[5]:

	incident_id	offense_id	offense_code	offense_type_id	offense_category_id	first_
0	202268791	202268791299900	2999	criminal-mischief-other	public-disorder	2/10
1	2021387586	2021387586299900	2999	criminal-mischief-other	public-disorder	7/7
2	2020641486	2020641486299900	2999	criminal-mischief-other	public-disorder	1
3	2018612468	2018612468299900	2999	criminal-mischief-other	public-disorder	9/6
4	2020293614	2020293614299900	2999	criminal-mischief-other	public-disorder	5/8

In [10]:
```
# Save the file with UTF-8 encoding
df.to_csv("../Datasets/crime_cleaned.csv", encoding="utf-8", index=False)
```

4. R Plots

Following plots are created using R:

- **Bar Plot** - To show the Number of crimes per offense category

- **Pie Chart** - To show the distribution of crimes in the dataset
- **Histogram** - To show the victims count by year
- **Scatter Plot** - To show the no. of crimes by neighborhood
- **Bubble Plot** - To show the no. of victims by crime type
- **Timeline Plot** - To show the no. of crimes by year and month

All plots can be found in the Plots Directory.

5. Conclusion

In this experiment, I learned how to create different types of plots using R. I have created 6 different plots to visualize the dataset using the ggplot2 library in R. I also used the lubridate library to work with dates. I have created a bar plot, pie chart, histogram, scatter plot, bubble plot, and timeline plot. These plots help us to understand the dataset better and find insights from it.