```
# Install R essentials
!apt-get install r-base
# Install the IR kernel to use R in Colab
!pip install rpy2
```

```
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
r-base is already the newest version (4.4.1-1.2204.0).
0 upgraded, 0 newly installed, 0 to remove and 49 not upgraded.
Requirement already satisfied: rpy2 in /usr/local/lib/python3.10/dist-packages (3.4.2)
Requirement already satisfied: cffi>=1. Run cell (Ctrl+Enter)                ython3.10/dist-packages (from rpy2) (1.17.1)
Requirement already satisfied: jinja2 i cell executed since last change     .10/dist-packages (from rpy2) (3.1.4)
Requirement already satisfied: pytz in                                      0/dist-packages (from rpy2) (2024.1)
Requirement already satisfied: tzlocal  executed by Basuri Bhujade          3.10/dist-packages (from rpy2) (5.2)
Requirement already satisfied: pycparse 4:45 PM (0 minutes ago)             on3.10/dist-packages (from cffi>=1.10.0->rpy2) (2.22)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from jinja2->rpy2) (2.1.5)
```

```
%load_ext rpy2.ipython
```

```
%%R
# Install the required libraries
install.packages("ggplot2")  # For plotting
install.packages("dplyr")    # For data manipulation
```

```
%%R
# Load the required libraries
library(ggplot2)
library(dplyr)
```

```
WARNING:rpy2.rinterface_lib.callbacks:R[write to console]:
Attaching package: ‘dplyr’


WARNING:rpy2.rinterface_lib.callbacks:R[write to console]: The following objects are masked from ‘package:stats’:

    filter, lag

WARNING:rpy2.rinterface_lib.callbacks:R following objects are masked from ‘package:base’:

    intersect, setdiff, setequal, union
```

Run cell (Ctrl+Enter)
cell executed since last change

executed by Basuri Bhujade
4:45 PM (0 minutes ago)
executed in 4.358s

```
%%R
# Load the dataset from the file path
crime_data <- read.csv("/content/2010.csv")
# Preview the first few rows of the dataset
head(crime_data)
```

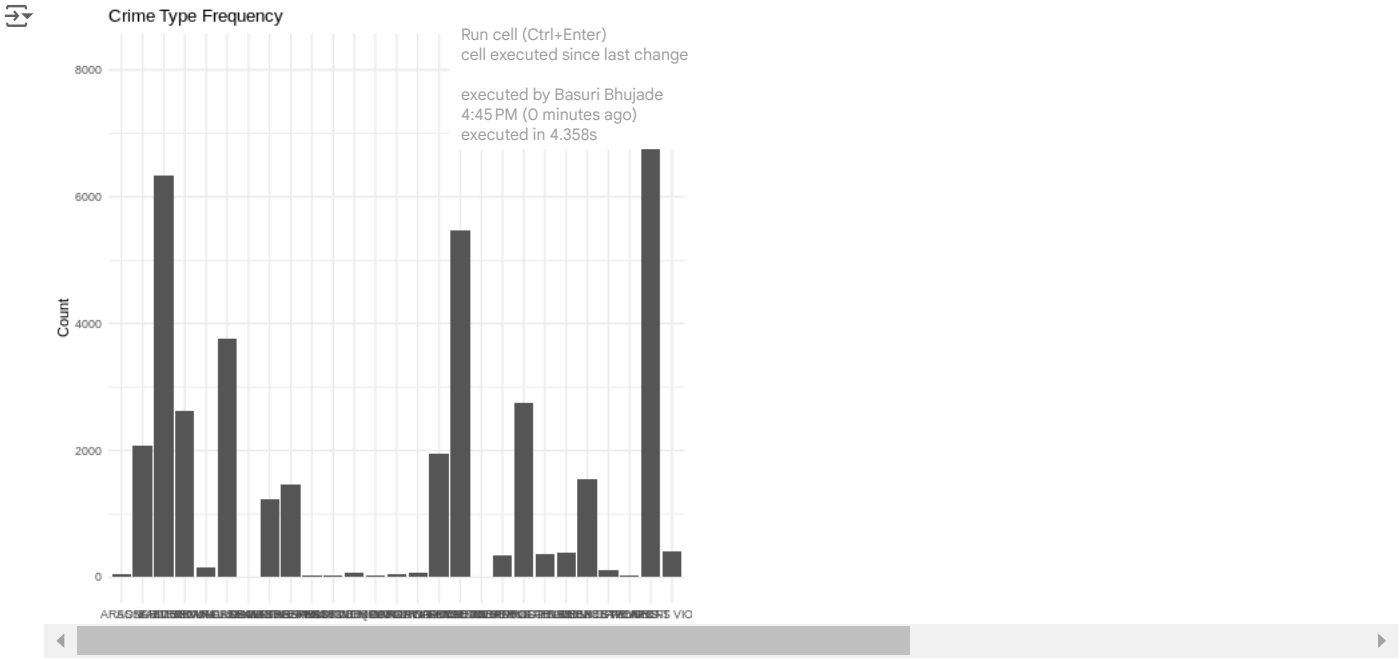
```
         ID Case.Number                   Date                   Block IUCR
1 11039140    JA371686 01/01/2010 12:00:00 AM   055XX W FARRAGUT AVE 1753
2 10342825    HY533211 01/01/2010 12:00:00 AM   056XX W EASTWOOD AVE 1752
3 10938629    JA251783 01/01/2010 12:00:00 AM 043XX N MONTICELLO AVE 1753
4 11262896    JB194570 01/01/2010 12:00:00 AM    057XX N ROCKWELL ST 0266
5 11875312    JC490052 01/01/2010 12:00:00 AM        017XX W 48TH ST 1752
6 11033112    JA366109 01/01/2010 12:00:00 AM        047XX S WOOD ST 1562
                 Primary.Type                  Description Location.Description
1 OFFENSE INVOLVING CHILDREN SEX ASSLT OF CHILD BY FAM MBR            RESIDENCE
2 OFFENSE INVOLVING CHILDREN AGG CRIM SEX ABUSE FAM MEMBER            RESIDENCE
3 OFFENSE INVOLVING CHILDREN SEX ASSLT OF CHILD BY FAM MBR            RESIDENCE
4        CRIM SEXUAL ASSAULT                    PREDATORY                OTHER
5 OFFENSE INVOLVING CHILDREN AGG CRIM SEX ABUSE FAM MEMBER           APARTMENT
6               SEX OFFENSE     AGG CRIMINAL SEXUAL ABUSE            RESIDENCE
  Arrest Domestic Beat District Ward Community.Area FBI.Code X.Coordinate
1  false    false 1623       16   45             11       02           NA
2  false     true 1622       16   45             15       20      1137897
3  false    false 1723       17   33             16       02           NA
4   true    false 2011       20   40              2       02           NA
5  false    false  931        9   15             61       17           NA
6  false     true  931        9   15             61       17      1165145
  Y.Coordinate Year            Updated.On Latitude Longitude
1           NA 2010 08/02/2017 03:54:51 PM       NA        NA
2      1930354 2010 02/10/2018 03:50:01 PM 41.96503 -87.76838
3           NA 2010 05/09/2017 03:54:21 PM       NA        NA
4           NA 2010 10/22/2019 04:01:10 PM       NA        NA
5           NA 2010 10/29/2019 03:51:02 PM       NA        NA
6      1873300 2010 06/28/2019 04:19:07 PM 41.80793 -87.66981
                     Location
1
2 (41.96503178, -87.768375694)
3
4
5
6 (41.807934431, -87.66981324)
```

```
%%R
# Perform Data preprocessing
crime_data <- crime_data %>% na.omit()
```

```
%%R
# Check the column names of the data
colnames(crime_data)
```

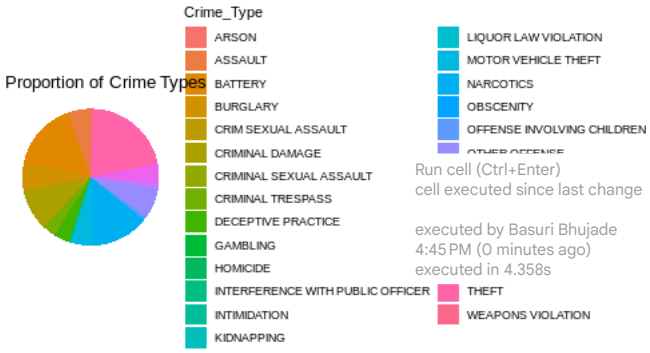
```
 [1] "ID"                  "Case.Number"          "Date"
 [4] "Block"               "IUCR"                 "Primary.Type"
 [7] "Description"         "Location.Description" "Arrest"
[10] "Domestic"            "Beat"                 "District"
[13] "Ward"                "Community.Area"       "FBI.Code"
[16] "X.Coordinate"        "Y.Coordinate"         "Year"
[19] "Updated.On"          "Latitude"             "Longitude"
[22] "Location"
```

```
%%R
library(ggplot2)

# Bar chart of Primary Type of crime
ggplot(crime_data, aes(x = `Primary.Type`)) +
    geom_bar() +
    ggtitle("Crime Type Frequency") +
    xlab("Primary Type of Crime") +
    ylab("Count") +
    theme_minimal()
```



Observation : This chart shows the frequency of different crime types. You can see which crimes are most common.

```
%%R
library(ggplot2)

# Prepare data for the pie chart by counting the occurrences of each crime type
crime_count <- as.data.frame(table(crime_data$`Primary.Type`))
colnames(crime_count) <- c("Crime_Type", "Count")

# Create a pie chart
ggplot(crime_count, aes(x = "", y = Count, fill = Crime_Type)) +
    geom_bar(stat = "identity", width = 1) +
    coord_polar("y", start = 0) +
    ggtitle("Proportion of Crime Types") +
    theme_void() +
    theme(legend.position = "right")
```

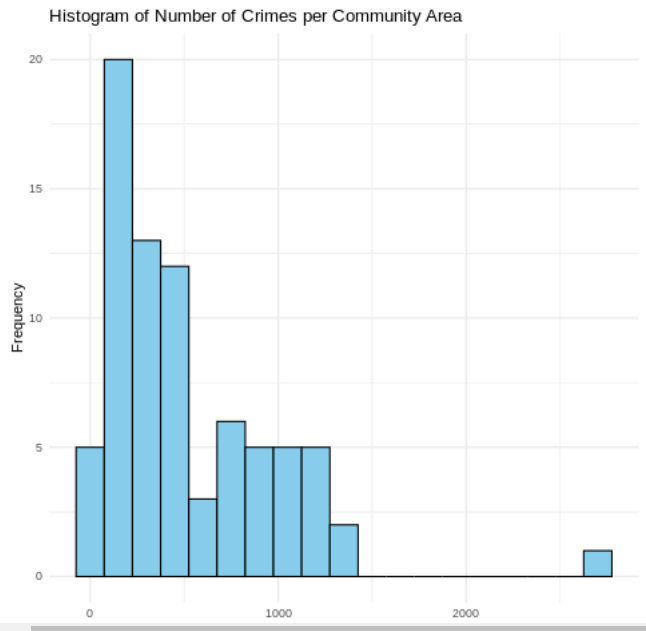


Run cell (Ctrl+Enter)
cell executed since last change

executed by Basuri Bhujade
4:45 PM (0 minutes ago)
executed in 4.358s

Observation : This chart shows the proportion of each crime type relative to the total number of crimes.

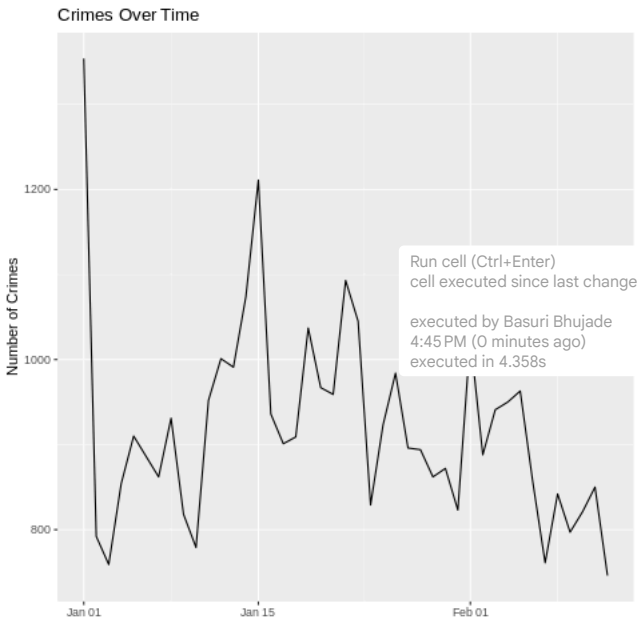
```
%%R
library(ggplot2)
library(dplyr)

# Count crimes per Community.Area
crime_area_count <- crime_data %>%
  group_by(Community.Area) %>%
  summarize(Count = n())

# Create histogram
ggplot(crime_area_count, aes(x = Count)) +
  geom_histogram(binwidth = 150, fill = "skyblue", color = "black") +
  labs(x = "Number of Crimes", y = "Frequency", title = "Histogram of Number of Crimes per Community Area") +
  theme_minimal()
```



Observation : Histogram of Number of Crimes per Community Area

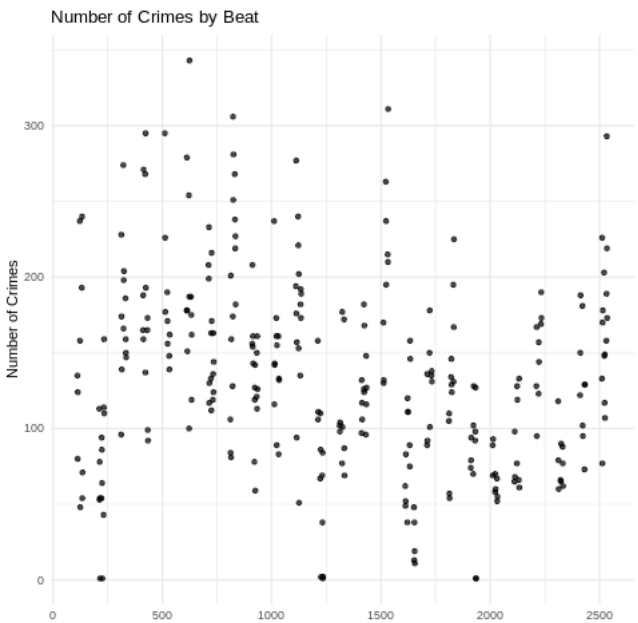
```
%%R
# Timeline of crimes over time
crime_data$Date <- as.Date(crime_data$Date, format="%m/%d/%Y %I:%M:%S %p")
ggplot(crime_data, aes(x = Date)) +
  geom_line(stat = "count") +
```

```
  labs(title = "Crimes Over Time", x = "Date", y = "Number of Crimes")
```



Observation : This timeline chart reveals trends in crime over time, showing periods of higher or lower crime rates.

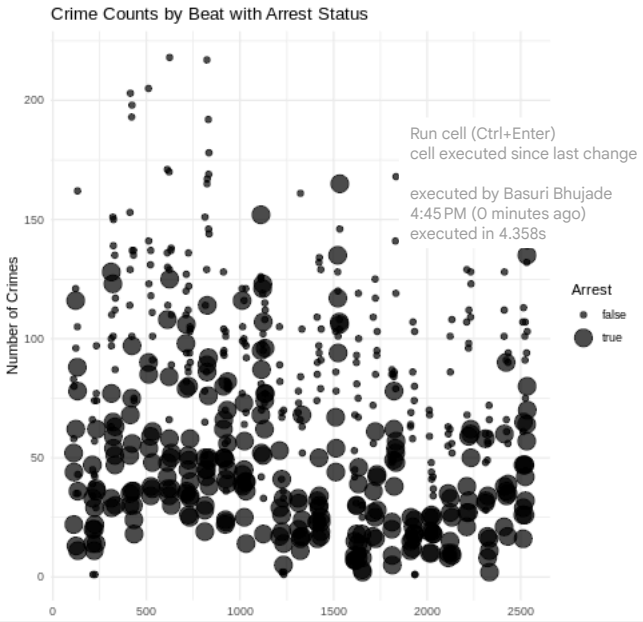
```
%%R
# Crime counts by beat
library(dplyr)
crime_counts_by_beat <- crime_data %>% count(Beat)
# Scatter plot
ggplot(crime_counts_by_beat, aes(x = Beat, y = n)) +
  geom_point(alpha = 0.7) +
  labs(title = "Number of Crimes by Beat", x = "Beat", y = "Number of Crimes") +
  theme_minimal()
```



Observation : This scatter plot helps identify which beats have higher crime counts. It can be useful for understanding crime distribution across different beats.

```
%%R
# Bubble plot of crime counts by beat with arrest status as bubble size
library(dplyr)
crime_by_beat_arrest <- crime_data %>%
  group_by(Beat, Arrest) %>%
  summarize(count = n(), .groups = 'drop')
# Convert Arrest to a factor for better visualization
```

```
crime_by_beat_arrest$Arrest <- as.factor(crime_by_beat_arrest$Arrest)
ggplot(crime_by_beat_arrest, aes(x = Beat, y = count, size = Arrest)) +
  geom_point(alpha = 0.7) +
  labs(title = "Crime Counts by Beat with Arrest Status",
       x = "Beat", y = "Number of Crimes", size = "Arrest") +
  theme_minimal()
```



Run cell (Ctrl+Enter)
cell executed since last change

executed by Basuri Bhujade
4:45 PM (0 minutes ago)
executed in 4.358s

Observation : This bubble plot helps visualize how the number of crimes varies by beat and whether arrests are more frequent in certain