5.1 VISUALIZE YOUR DATA

Data Visualization

Data Visualization is used to communicate information clearly and efficiently to users by the usage of information graphics such as tables and charts. It helps users in analyzing a large amount of data in a simpler way. It makes complex data more accessible, understandable, and usable.

> **Why is data visualization important?**

Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making slight adjustments.

> **Data visualization can also:**
>   - Identify areas that need attention or improvement.
>   - Clarify which factors influence customer behavior.
>   - Help you understand which products to place where.
>   - Predict sales volumes

➤ **various charts such as :**
    a) Histogram
    b) Column charts
    c) Box plot chart
    d) Pie charts
    e) Scatter plot

1. Histogram:
    A histogram is an accurate representation of the <u>distribution</u> of numerical data. It is an estimate of the <u>probability distribution</u> of a <u>continuous variable</u> (CORAL) and was first introduced by Karl Pearson.
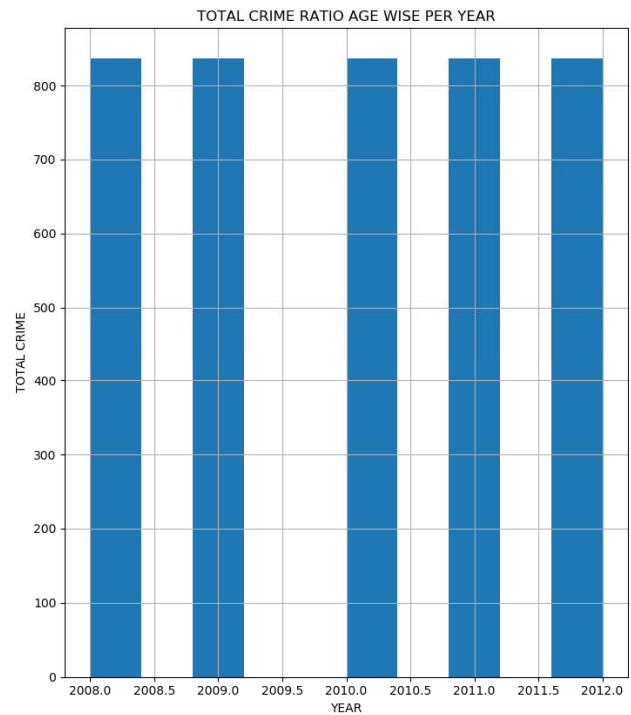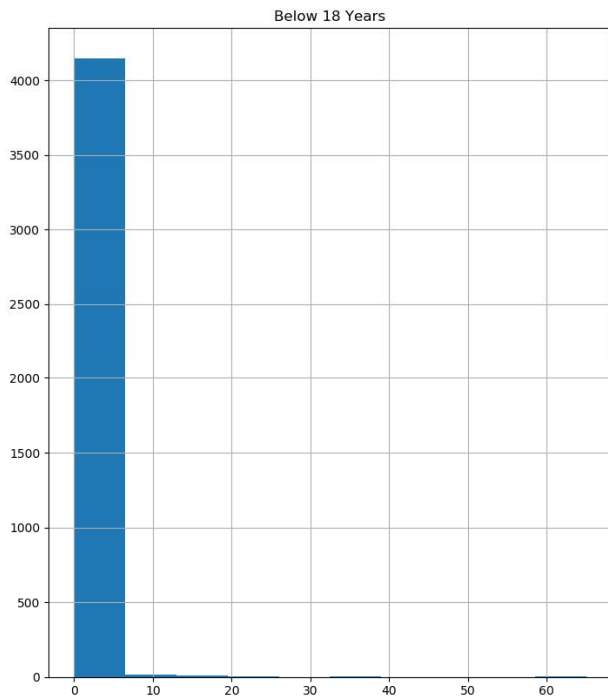
<u>Code:</u>

```python
import pandas as pd
import matplotlib.pyplot as plt


table = pd.read_csv("C://Users//MALVIKA//Desktop//18mcl2//Project_datascience/cyber_Crime.csv")
df = pd.DataFrame(table,columns=['STATE/UT','CRIME HEAD','YEAR','Below 18 Years',
'Between 18-30 Years','Between 30-45 Years','Between 45-60 Years','Above 60 Years','Total'])

#df.hist(column=['YEAR','Total'])
df.hist(column=['YEAR','Below 18 Years'])
#df.hist(column=['YEAR','Between 18-30 Years'])
#df.hist(column=['YEAR','Between 30-45 Years'])
#df.hist(column=['YEAR','Between 45-60 Years'])
#df.hist(column=['YEAR','Above 60 Years'])
plt.xlabel('YEAR')
plt.ylabel('TOTAL CRIME')
plt.title('TOTAL CRIME RATIO AGE WISE PER YEAR')
plt.show()
```

Output:

> ➢ Histogram show the crime ratio age wise per year in india.

2. Column Chart/Bar graph:

A column chart is a graphic representation of data. Column charts display vertical bars going across the chart horizontally, with the values axis being displayed on the left side of the chart.
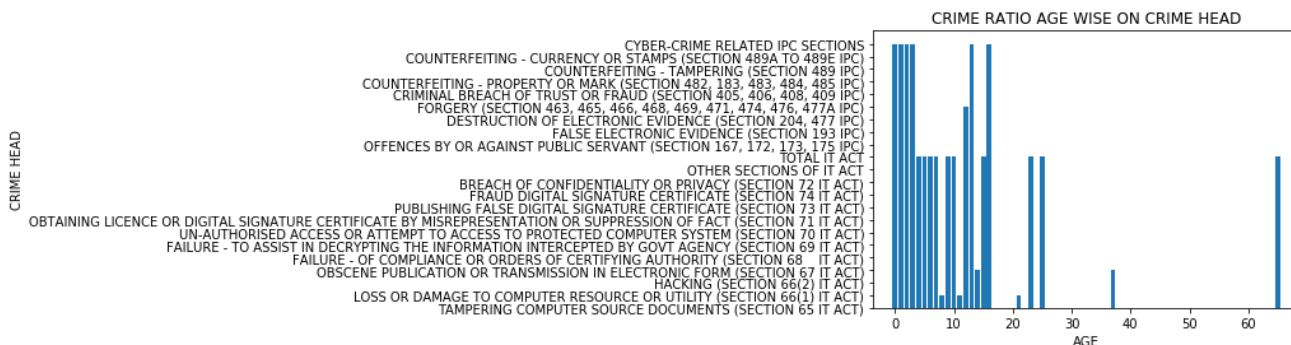
Code:

```python
import pandas as pd
import matplotlib.pyplot as plt


table = pd.read_csv("C://Users//MALVIKA//Desktop//18mcl2//Project_datascience/cyber_Crime.csv")
df = pd.DataFrame(table,columns=['STATE/UT','CRIME HEAD','YEAR','Below 18 Years',
'Between 18-30 Years','Between 30-45 Years','Between 45-60 Years','Above 60 Years','Total'])

plt.bar(df['Below 18 Years'], df['CRIME HEAD'])
#plt.bar(df['Between 18-30 Years'], df['CRIME HEAD'])
#plt.bar(df['Between 30-45 Years'], df['CRIME HEAD'])
#plt.bar(df['Between 45-60 Years'], df['CRIME HEAD'])
#plt.bar(df['Above 60 Years'], df['CRIME HEAD'])
#plt.bar(df['Total'], df['CRIME HEAD'])
plt.xlabel("AGE")
plt.ylabel("CRIME HEAD")
plt.title('CRIME RATIO AGE WISE ON CRIME HEAD')
plt.show()
```

Output:

➢ **Bar graph show the crime ratio age wise crime heads in India.**

3. Box Plot Chart:

a box plot or boxplot is a method for graphically depicting groups of numerical data through their quartiles. Box plots may also have lines extending vertically from the boxes (*whiskers*) indicating variability outside the upper and lower quartiles, hence the terms box-and-whisker plot and box-and-whisker diagram. Outliers may be plotted as individual points.

Code:

```python
import pandas as pd
import matplotlib.pyplot as plt


table = pd.read_csv("C://Users//MALVIKA//Desktop//18mcl2//Project_datascience/cyber_Crime.csv")
df = pd.DataFrame(table,columns=['STATE/UT','CRIME HEAD','YEAR','Below 18 Years',
'Between 18-30 Years','Between 30-45 Years','Between 45-60 Years','Above 60 Years','Total'])


df.boxplot(column=['Below 18 Years'])
plt.xlabel('AGE')
plt.ylabel('CRIME RATIO')
plt.title('CRIME RATIO AGE WISE')
plt.show()
```

Output:
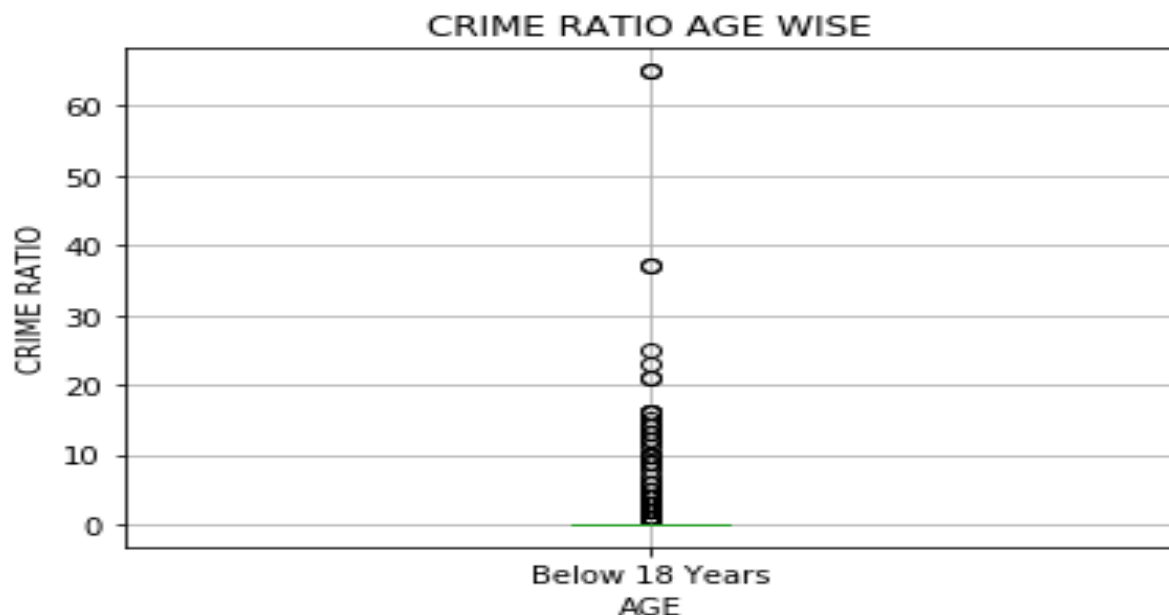
> ➢ **Box plot show the crime ratio below 18 years in India.**

4. Pie Chart:

A pie chart (or a circle chart) is a circular statistical graphic, which is divided into slices to illustrate numerical proportion. In a pie chart, the arc length of each slice (and consequently its central angle and area), is proportional to the quantity it represents. While it is named for its resemblance to a pie which has been sliced, there are variations on the way it can be presented. The earliest known pie chart is generally credited to William Playfair's *Statistical Breviary* of 1801.

Code:

```python
import pandas as pd
import matplotlib.pyplot as plt


table = pd.read_csv("C://Users//MALVIKA//Desktop//18mcl2//Project_datascience/cyber_Crime2.csv")
df = pd.DataFrame(table,columns=['STATE/UT','CRIME HEAD','YEAR','Below 18 Years',
'Between 18-30 Years','Between 30-45 Years','Between 45-60 Years','Above 60 Years','Total'])


plt.title("CRIME RATIO OF BELOW 18 YEARS PER YEAR IN INDIA")
data=df['YEAR']
data1=df['Below 18 Years']
plt.pie(data1,labels=data,autopct ='% 1.1f %%', shadow = True)
plt.show()
```

Output:
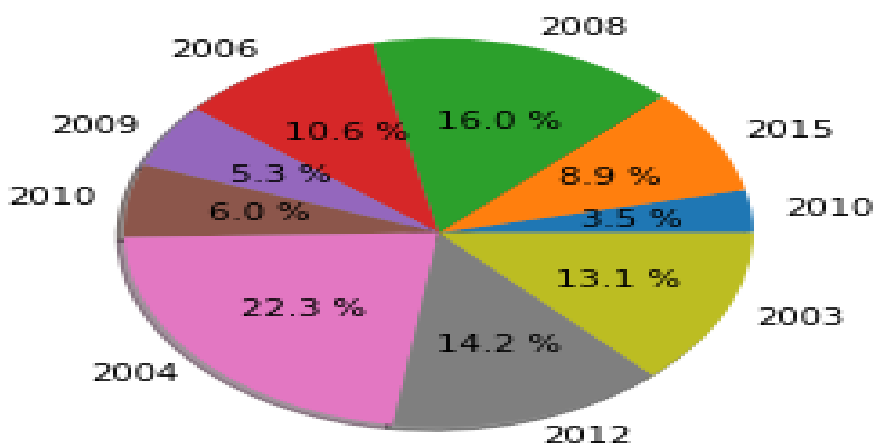
➤ **Pie chart show the crime ratio of below 18 years population year wise in India.**

5. Scatter plot

Scatter plot is a set of points plotted on a horizontal and vertical axes. Scatter plots are important in statistics because they can show the extent of correlation, if any, between the values of observed quantities or phenomena (called variables). If no correlation exists between the variables, the points appear randomly scattered on the coordinate plane. If a large correlation exists, the points concentrate near a straight line. Scatter plots are useful data visualization tools for illustrating a trend.

Code:

```python
import pandas as pd
import matplotlib.pyplot as plt


table = pd.read_csv("C://Users//MALVIKA//Desktop//18mcl2//Project_datascience/cyber_Crime2.csv")
df = pd.DataFrame(table,columns=['STATE/UT','CRIME HEAD','YEAR','Below 18 Years',
'Between 18-30 Years','Between 30-45 Years','Between 45-60 Years','Above 60 Years','Total'])

year = df['YEAR']
age = df['Below 18 Years']
plt.scatter(year, age, edgecolors='r')
plt.xlabel('YEAR')
plt.ylabel('AGE')
plt.title('CRIME RATIO AGE WISE PER YEAR IN INDIA')
plt.show()
```

Output:
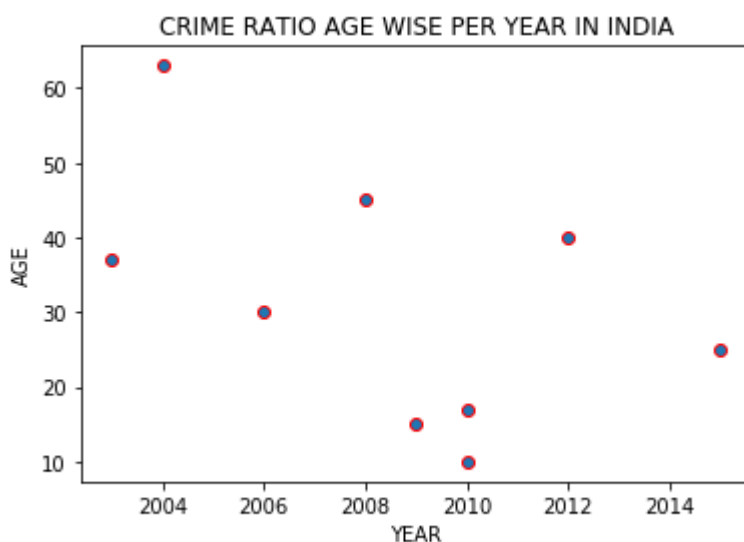
➤ **Scatter plot show the crime ratio of below 18 years population year wise in India.**