SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)
DEPARTMENT OF INFORMATION TECHNOLOGY

COURSE CODE: DJS22ITL5013

COURSE NAME: Statistical Analysis Lab **CLASS:** T.Y. BTech

NAME: Anish Sharma

EXPERIMENT NO.02

CO 1: Interpret the data using Descriptive Statistics.

AIM / OBJECTIVE: To explore descriptive statistics- Measures of Central Tendency

DESCRIPTION OF EXPERIMENT:

Perform the following

1. Weight of babies (kg) below 6 months taken from a hospital record is given below. Calculate Mean, Median, Mode.

3.0	4.5	4.3	2.5	3.5	2.5	4.0	4.5	6.5	5.0
4.0	5.0	4.1	4.2	4.3	4.5	3.3	3.5	3.6	5.3
5.4	5.5	5.5	5.7	5.8	5.6	5.8	5.9	6.0	3.4
6.1	6.2	6.3	5.5	6.3	6.3	7.0	4.0	3.4	5.0

```
import numpy as np
from scipy import stats

# Data
weights = [3.0, 4.5, 4.3, 2.5, 3.5, 2.5, 4.0, 4.5, 6.5, 5.0, 4.0, 5.0, 4.1,
4.2, 4.3, 4.5,
           3.3, 3.5, 3.6, 5.3, 5.4, 5.5, 5.5, 5.7, 5.8, 5.6, 5.8, 5.9, 6.0,
3.4, 6.1, 6.2,
           6.3, 5.5, 6.3, 6.3, 7.0, 4.0, 3.4, 5.0]

# Mean
mean_weight = np.mean(weights)

# Median
median_weight = np.median(weights)

# Mode
mode_result = stats.mode(weights, keepdims=False)
mode_weight = mode_result.mode  # Access mode value directly
mode_count = mode_result.count  # Access count value directly
```

```
print(f"Mean: {mean_weight:.2f}")
print(f"Median: {median_weight:.2f}")
print(f"Mode: {mode_weight:.2f} (Count: {mode_count})")
```

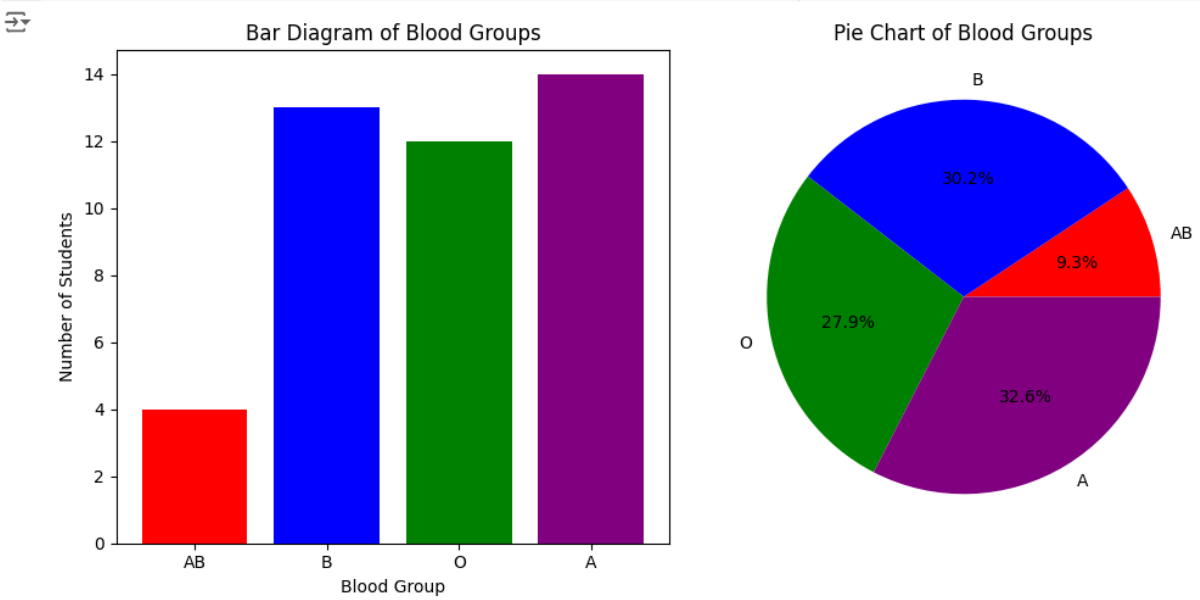
What are your observations about the data, based on the central tendency values.

```
Mean: 4.82
Median: 5.00
Mode: 4.00 (Count: 3)
```

2. Draw a Bar diagram, Pie chart for the following data on the blood group of 45 students in a class.

AB	B	O	A	O	O	A	O	B
AB	B	A	B	A	B	OB	AB	A
O	O	A	O	AB	O	O	A	A
B	A	A	AB	O	A	A	O	A
O	A	A	O	A	O	O	B	A

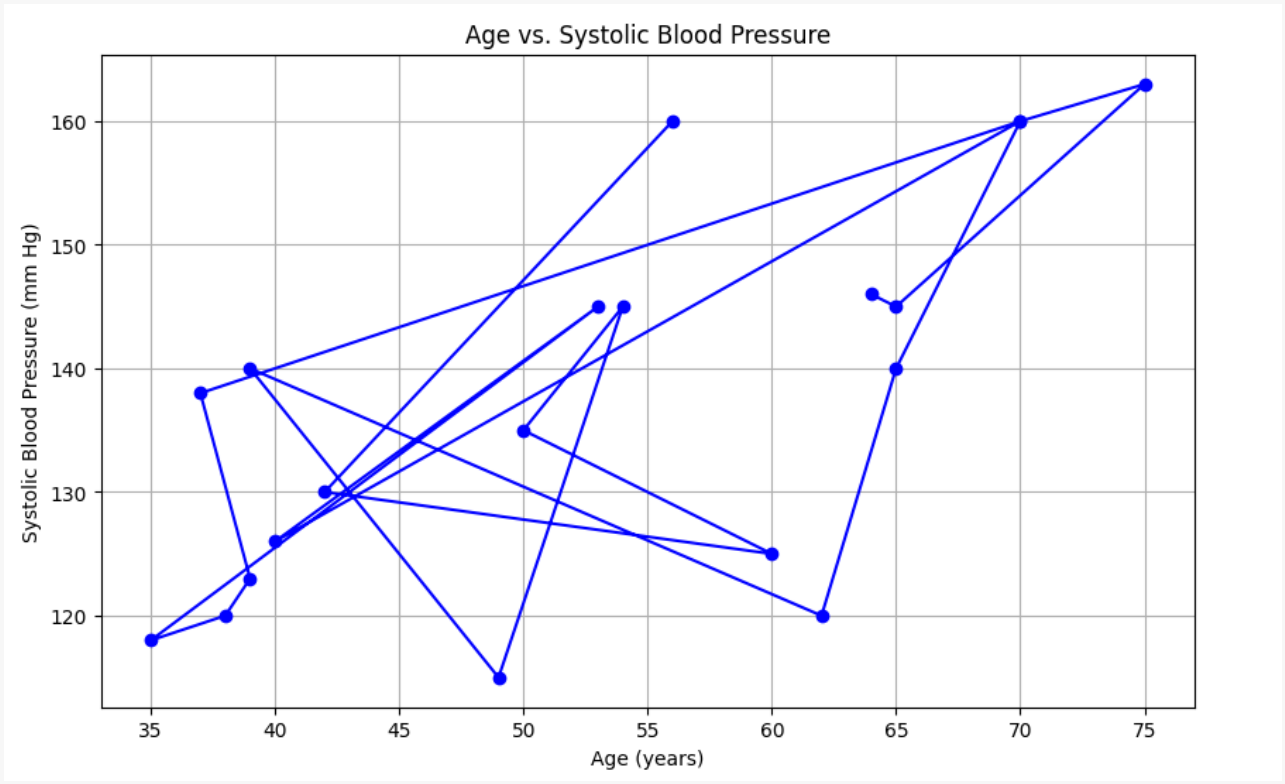
```
import matplotlib.pyplot as plt

# Data
blood_groups = ['AB', 'B', 'O', 'A']
counts = [4, 13, 12, 14]

# Bar Diagram
plt.figure(figsize=(10, 5))

plt.subplot(1, 2, 1)
plt.bar(blood_groups, counts, color=['red', 'blue', 'green', 'purple'])
plt.xlabel('Blood Group')
plt.ylabel('Number of Students')
plt.title('Bar Diagram of Blood Groups')

# Pie Chart
plt.subplot(1, 2, 2)
plt.pie(counts, labels=blood_groups, autopct='%1.1f%%', colors=['red',
'blue', 'green', 'purple'])
plt.title('Pie Chart of Blood Groups')

plt.tight_layout()
plt.show()
```



3. Draw a line graph for age(years) versus systolic blood pressure (mm Hg) for the following data given below

Age	56	42	60	50	54	49	39	62	65	70
BP	160	130	125	135	145	115	140	120	140	160
Age	40	53	35	38	39	37	70	75	65	64
BP	126	145	118	120	123	138	160	163	145	146

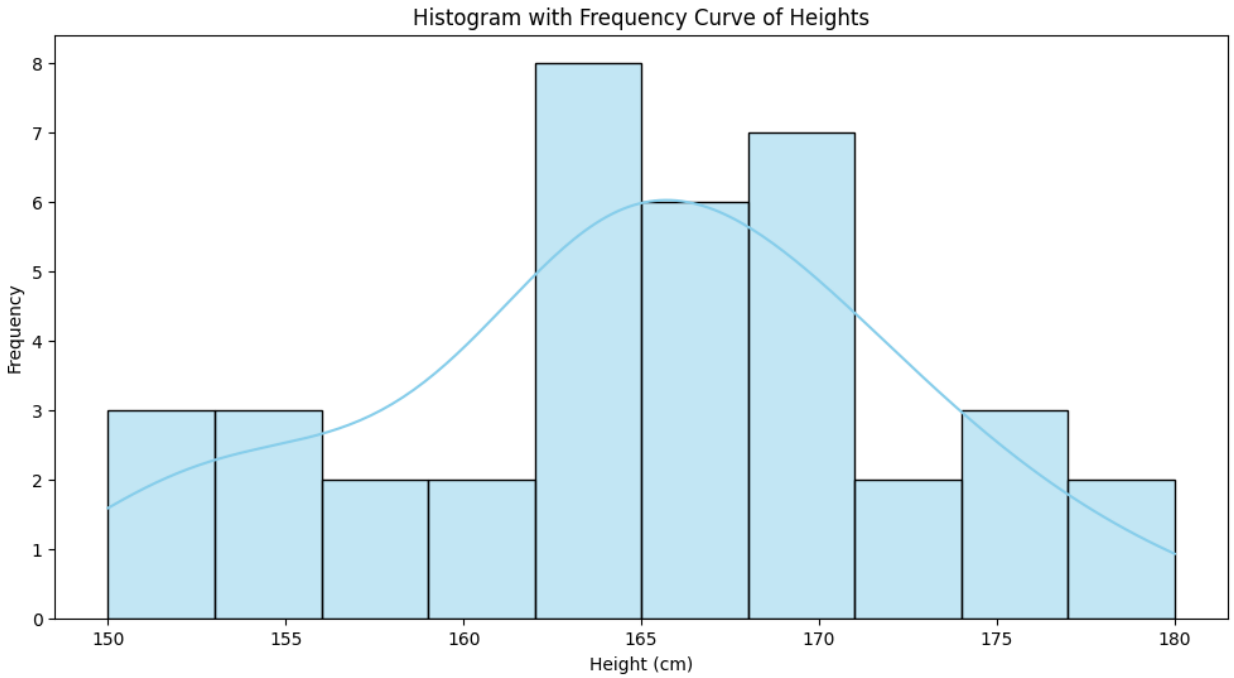
```
import matplotlib.pyplot as plt

# Data
ages = [56, 42, 60, 50, 54, 49, 39, 62, 65, 70, 40, 53, 35, 38, 39, 37, 70,
75, 65, 64]
bps = [160, 130, 125, 135, 145, 115, 140, 120, 140, 160, 126, 145, 118, 120,
123, 138,
       160, 163, 145, 146]

# Plot
plt.figure(figsize=(10, 6))
plt.plot(ages, bps, marker='o', linestyle='-', color='b')
plt.xlabel('Age (years)')
plt.ylabel('Systolic Blood Pressure (mm Hg)')
plt.title('Age vs. Systolic Blood Pressure')
plt.grid(True)
plt.show()
```



4. Draw a histogram with frequency curve for the following data on height in cm

155	171	170	169	167	180	175	150	164	165	163	170
158	153	172	152	163	164	167	159	163	162	168	154
177	164	165	174	164	160	170	157	169	168	165	151
165	174										

```
import matplotlib.pyplot as plt
import seaborn as sns

# Data
heights = [155, 171, 170, 169, 158, 153, 172, 152, 177, 164, 165, 174, 165,
174, 167, 180,
           175, 150, 164, 165, 163, 170, 163, 164, 164, 167, 159, 163, 162,
168, 154, 160,
           170, 157, 169, 168, 165, 151]

# Plot
plt.figure(figsize=(12, 6))
sns.histplot(heights, bins=10, kde=True, color='skyblue')
plt.xlabel('Height (cm)')
plt.ylabel('Frequency')
plt.title('Histogram with Frequency Curve of Heights')
plt.show()
```



5. Eight coins were tossed together and the number of heads resulting was noted. The operation was repeated 256 times and the frequencies (f) that were obtained for different values of x, the number of heads, are shown in the following table. Calculate median, quartiles, and 27th percentile.

x	0	1	2	3	4	5	6	7	8
f	1	9	26	59	72	52	29	7	1

```
import numpy as np

# Data
heads = [0, 1, 2, 3, 4, 5, 6, 7, 8]
frequencies = [1, 9, 26, 59, 72, 52, 29, 7, 1]

# Create dataset from frequencies
data = np.repeat(heads, frequencies)

# Median
median = np.median(data)

# Quartiles
Q1 = np.percentile(data, 25)
Q3 = np.percentile(data, 75)

# 27th Percentile
percentile_27 = np.percentile(data, 27)

print(f"Median: {median}")
print(f"1st Quartile (Q1): {Q1}")
print(f"3rd Quartile (Q3): {Q3}")
print(f"27th Percentile: {percentile_27}")
```

```
Median: 4.0
1st Quartile (Q1): 3.0
3rd Quartile (Q3): 5.0
27th Percentile: 3.0
```

6. Find the mean, median and mode for the following collection of responses to the question: "How many parking tickets have you received this semester?"

   1, 1, 0,1, 2, 2, 0, 0, 0, 3, 3,0, 3, 3, 0,2, 2, 2, 1, 1,4, 1, 1,0,3, 0, 0, 0, 1, 1, 2, 2, 2, 2,1, 1, 1, 1, 4, 4, 4,1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2,1, 1, 1, 1, 1, 3,3,0, 3, 3, 1, 1, 1, 1,0, 0, 1, 1, 1, 1, 3, 3, 3, 2, 3, 3, 1, 1, 1,2, 2, 2,4, 5, 5, 4, 4, 1, 1, 1, 4,1, 1, 1,3, 3, 5,3, 3, 3, 2,3, 3, 0, 0, 0, 0, 3, 3, 3, 3, 3, 3, 0, 2, 2, 2, 2, 1, 1, 1,3, 1, 0, 0, 0,1, 1, 3,1, 1, 1, 2, 2, 2, 4, 2, 2, 2, 1, 1, 1, 1,0, 0, 2, 2, 3, 3,2, 2, 3,2, 0, 0, 1, 1,3, 3, 3, 1, 1, 1, 1, 1,2, 2, 2, 2, 1, 1, 1, 1, 0,1, 1, 1, 3,1, 1, 1, 2, 2, 2, 1, 1, 1,2, 1, 1, 1,3, 3,5, 3, 3, 1, 1, 1, 3, 3, 3, 3, 1, 1, 1,4, 1, 1, 4, 4, 4, 4, 4, 4,1, 1, 1,2, 2,5, 5, 2, 3, 3, 4, 4,3,2, 2, 2, 1,5, 1,2, 2, 1, 1, 1, 2, 2, 2, 2, 2,1, 1, 0,1, 1, 1,3, 3, 3, 3, 3

```
from scipy import stats

# Data
tickets = [1, 1, 0, 1, 2, 2, 0, 0, 0, 3, 3, 0, 3, 3, 0, 2, 2, 2, 1, 1, 4, 1, 1, 0, 3, 0,
```

```
           0, 0, 1, 1, 2, 2, 2, 2, 1, 1, 1, 1, 4, 4, 4, 1, 1, 1, 2, 2, 2, 2,
2, 2, 2, 2,
           1, 1, 1, 1, 1, 3, 3, 0, 3, 3, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 3, 3,
3, 2, 3, 3,
           1, 1, 1, 2, 2, 2, 4, 2, 2, 2, 1, 1, 1, 1, 0, 0, 2, 2, 3, 3, 2, 2,
3, 2, 0, 0,
           1, 1, 3, 3, 3, 1, 1, 1, 1, 1, 2, 2, 2, 2, 1, 1, 1, 1, 0, 1, 1, 1,
3, 1, 1, 1,
           2, 2, 2, 1, 1, 1, 2, 1, 1, 1, 3, 3, 5, 3, 3, 1, 1, 1, 3, 3, 3, 3,
1, 1, 1, 4,
           1, 1, 4, 4, 4, 4, 4, 4, 1, 1, 1, 2, 2, 5, 5, 2, 3, 3, 4, 4, 3, 2,
2, 2, 1, 5,
           1, 2, 2, 1, 1, 1, 2, 2, 2, 2, 2, 1, 1, 0, 1, 1, 1, 3, 3, 3, 3,
3]

# Mean
mean_tickets = np.mean(tickets)

# Median
median_tickets = np.median(tickets)

# Mode
mode_tickets = stats.mode(tickets)

print(f"Mean: {mean_tickets:.2f}")
print(f"Median: {median_tickets:.2f}")
# Access the mode and count directly from the ModeResult object
print(f"Mode: {mode_tickets.mode} (Count: {mode_tickets.count})")
```

```
Mean: 1.80
Median: 2.00
Mode: 1 (Count: 80)
```

7. Select any data set. (Kaggle, UCI Machine Learning Repository, Google data sets)
8. Generate the mean, median, mode, range, midrange for the chosen dataset.
9. Generate various types of graphs for the selected dataset.

SOURCE CODE (OPTIONAL):

OBSERVATIONS / DISCUSSION OF RESULT:

CONCLUSION:

Observation Sheet Questions:

1. Based on all conclusions on your actual results; describe the meaning of the experiment and the implications of your results.
2. Give some real-life examples, where these measures are applied.

REFERENCES:

Website References:

Chapter 2 Lab 2: Descriptive Statistics | Answering questions with data: Lab Manual (crumplab.com)

https://www.webpages.uidaho.edu/~stevel/251/Utts/R/chapter%2011.pdf

https://onlinestatbook.com/2/estimation/mean.html

https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_confidence_intervals/bs704_confidence_intervals_print.html

https://courses.lumenlearning.com/suny-natural-resources-biometrics/chapter/chapter-2-sampling-distributions-and-confidence-intervals/