

Ex. No:	1
Date:	26/24

## Introduction to various Data Visualization tools

### AIM

Study experiment on knowing the various visualization tools.

### DESCRIPTION

Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from. It is the representation of information and data through use of common graphics, such as charts, plots, infographics, and animations.

Data visualization is a powerful way for people, especially data professionals, to display data so that it can be interpreted easily.

Data Visualization enables decision-makers of any enterprise or industry to look into analytical reports and understand concepts that might otherwise be difficult to grasp.

**The following are some common types of data visualizations:**

**Table:** A table is data displayed in rows and columns, which can be easily created in a Word document or Excel spreadsheet.

**Chart or graph:** Information is presented in tabular form with data displayed along an x and y axis, usually with bars, points, or lines, to represent data in comparison.

**Geospatial visualization:** Data is depicted in map form with shapes and colors that illustrate the relationship between specific locations, such as a choropleth or heat map.

**Dashboard:** Data and visualizations are displayed, usually for business purposes, to help analysts understand and present data.

## Viva Questions:

1. Give Two Top companies where Tableau is Used for Data Visualization.

- ★ Nissan builds an enterprise data culture with tableau.
- ★ Henkel sales millions of euros a year with tableau.

2. What are the two uses of Google Charts?

Provides a perfect way to visualize data on your website.

3. Zoho analytics Specific use.

Charts, pivot tables, summary views and tabular view.

4. List the best 10 tools of Data Visualization Software of 2024.

- ★ Tableau
- ★ PowerBi
- ★ Google data studio
- ★ Infogram
- ★ Looker
- ★ Sisense
- ★ Qlik

5. Is Excel a data visualization tool?

MS Excel is a crucial data visualization tool.

6. What are Matplotlib, Seaborn ?

Matplotlib is library in python that enables users to generate visualization like histogram. Seaborn is a visualization library that is built on top of matplotlib.

7. Is Oracle a data visualization tool?

Oracle fusion analytics includes comprehensive self-service visualization capabilities for creating custom dashboards and reports.

8. Why does NASA use Matlab?

To verify algorithms before testing them aboard the space station.

9. How is data visualization used in healthcare?

Help medical staff interpret data analytics results faster, recognise trends and make better decisions.

10. How is data visualization used in healthcare?

Help medical staff interpret data analytics results faster, recognise trends and make better decisions.

## RESULT

Thus the study experiment on various Data visualization tools is successfully completed.

Ex. No:	2	Basic Visualization in Python
Date:	26/6/24	

Aim :

To implement different visualization effects with Python on data using matplotlib seaborn.

Description :

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. It presents data in 2D graphics. Seaborn is a visualization library that is built on top of Matplotlib. Matplotlib can be installed using the following command:

**pip install matplotlib**

Once the module installed, it must be imported into the program using the following command **import matplotlib as mpl**, where mpl is the alias name given to matplotlib library.matplotlib. **Pyplot** is a state-based interface to matplotlib.

**matplotlib.pyplot** is a collection of functions that make matplotlib work like MATLAB.

Each pyplot function makes some change to creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels etc.

pyplot can be imported into the program using following command

**import matplotlib.pyplot as plt**

Following are some of the basic data visualization plots

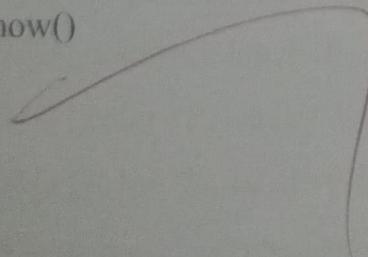
1. Line plots
2. Area plots
3. Histograms
4. Bar charts
5. Pie charts
6. Box plots
7. Scatter plots

### I ) Line Plots:

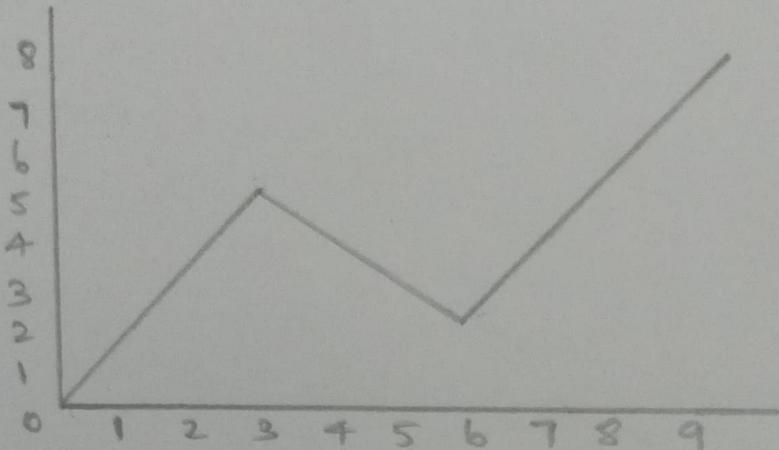
A line plot is used to represent quantitative values over a continuous interval or time period. It is generally used to depict trends on how the data has changed over time.

#### Program:

```
import matplotlib.pyplot as plt  
x = [1, 2, 3, 4, 5, 6]  
y = [1, 5, 3, 5, 7, 8]  
plt.plot(x, y)  
plt.show()
```



Output :



## II) Area Plots:

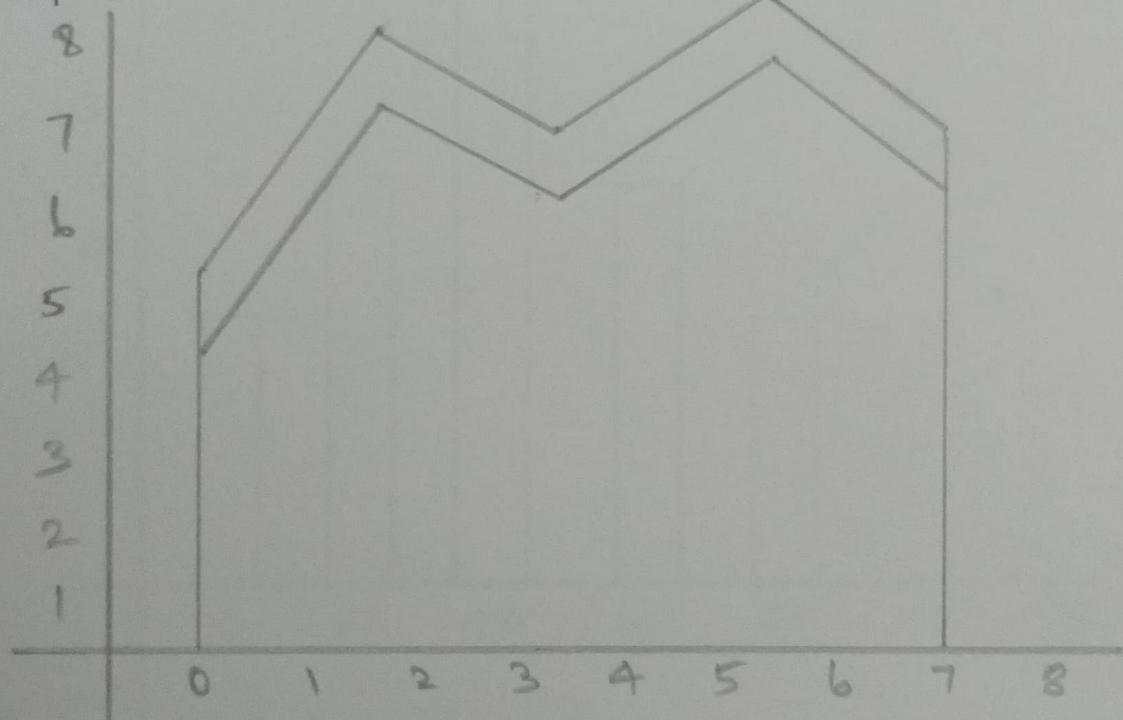
An Area Plot is also called as Area Chart which is used to display magnitude and proportion of multiple variables.

### Program:

```
import matplotlib.pyplot as plt  
days = [1,2,3,4,5]  
sleeping =[7,8,6,11,7]  
eating = [2,3,4,3,2]  
working =[7,8,7,2,2]  
playing = [8,5,7,8,13]  
plt.plot([],[],color='m', label='Sleeping', linewidth=5)  
plt.plot([],[],color='c', label='Eating', linewidth=5)  
plt.plot([],[],color='r', label='Working', linewidth=5)  
plt.plot([],[],color='k', label='Playing', linewidth=5)  
plt.stackplot(days, sleeping,eating,working,playing, colors=['m','c','r','k'])
```

```
plt.xlabel('x')
plt.ylabel('y')
plt.title('Stack Plot')
plt.legend()
plt.show()
```

Output :



### iii) Histograms:

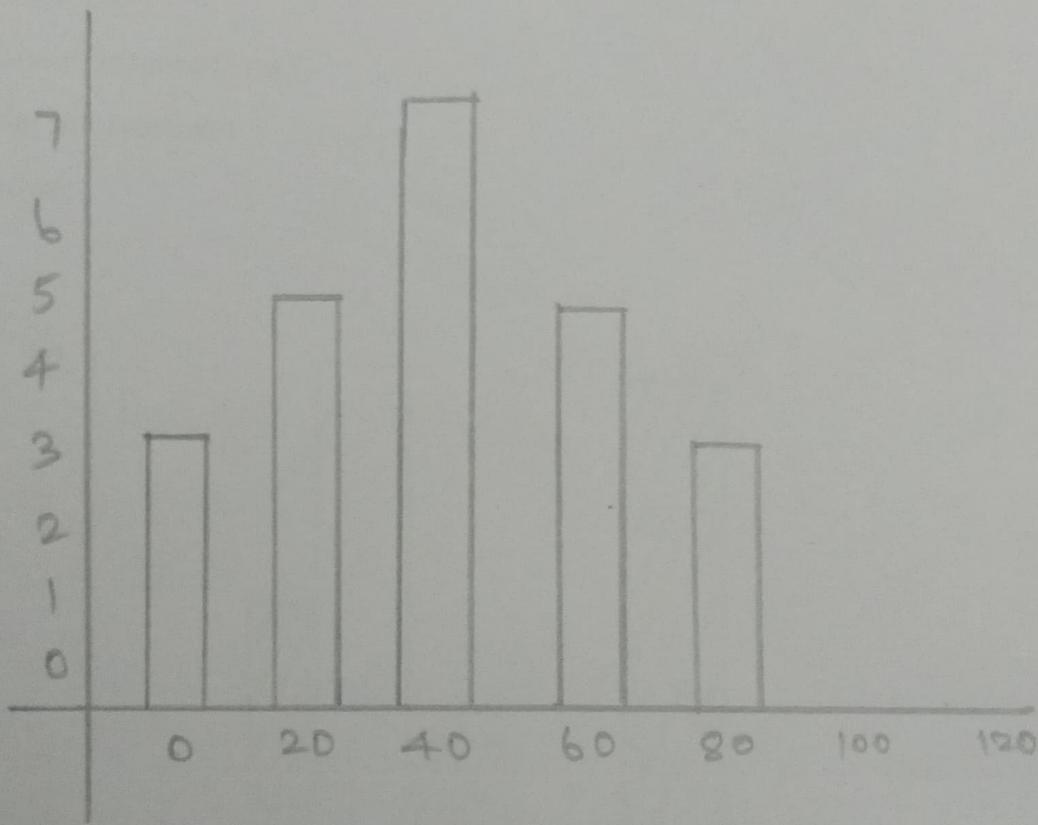
Histograms represents the frequency distribution of a dataset. It is a graph showing the number of observations within each given interval.

#### Program:

```
import matplotlib.pyplot as plt
population_age=[22,55,62,45,21,22,34,42,42,42,4,2,102,95,85,55,110,120,70,65,55,111,115,80]
bins = [0,10,20,30,40,50,60,70,80,90,100]
plt.hist(population_age, bins, histtype='bar', rwidth=0.8)
```

```
plt.xlabel('age groups')
plt.ylabel('Number of people')
plt.title('Histogram')
plt.show()
```

Output:



#### iv) Bar Charts:

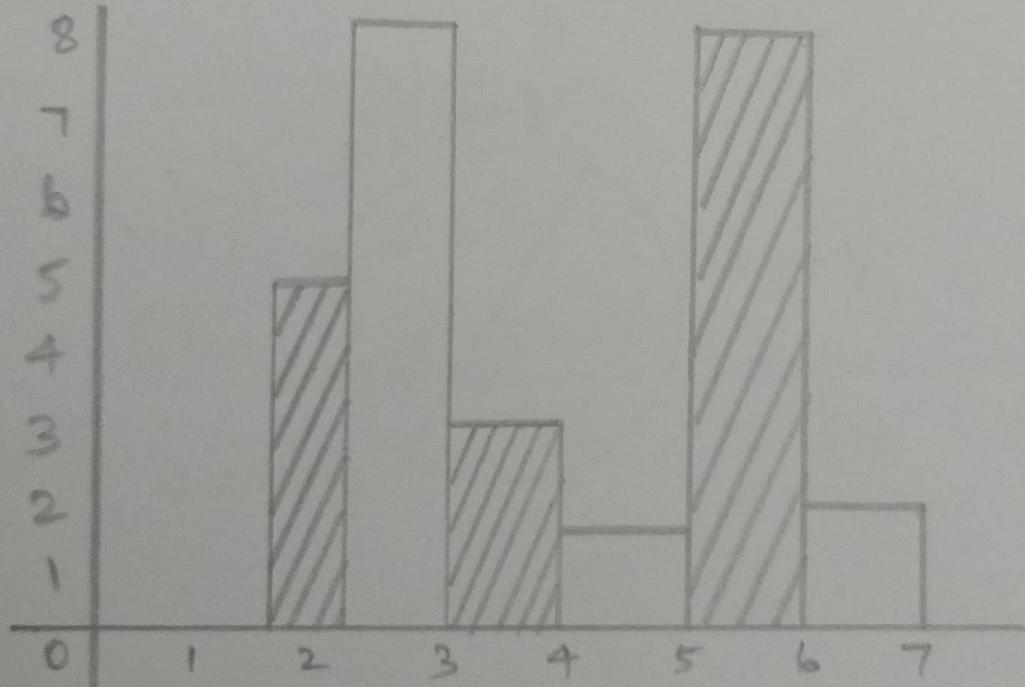
A Bar chart or bar graph is a chart or graph that presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent.

A bar plot is a way of representing data where the length of the bars represents the magnitude/size of the feature/variable.

**Program:**

```
import matplotlib.pyplot as plt  
plt.bar([0.25,1.25,2.25,3.25,4.25],[50,40,70,80,20],label="BMW",width=.5)  
plt.bar([.75,1.75,2.75,3.75,4.75],[80,20,20,50,60],label="Audi", color='r',width=.5)  
plt.legend()  
plt.xlabel('Days')  
plt.ylabel('Distance (kms)')  
plt.title('Information')  
plt.show()
```

**Output:**



#### V) Pie Charts:

A Pie chart is a circular statistical chart, which is divided into sectors to illustrate numerical proportion.

**Program :**

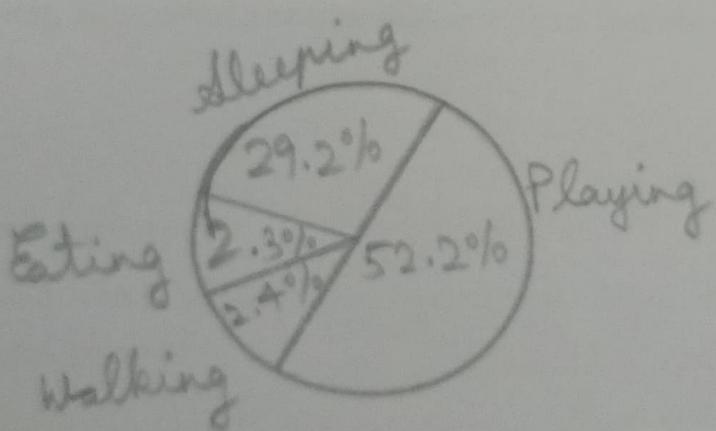
```
import matplotlib.pyplot as plt
```

```

days = [1,2,3,4,5]
sleeping =[7,8,6,11,7]
eating = [2,3,4,3,2]
working =[7,8,7,2,2]
playing = [8,5,7,8,13]
slices = [7,2,2,13]
activities = ['sleeping','eating','working','playing'] cols = ['c','m','r','b']
plt.pie(slices,labels=activities, colors=cols,start angle=90,shadow=True, explode=(0,0,1,0,0),
autopct='%.1f%%')
plt.title('Pie Plot')
plt.show()

```

Output:



## VI) Box Plots:

A Box plot (or box-and-whisker plot) shows the distribution of quantitative data in a way that facilitates comparisons between variables or across levels of a categorical variable.

Box plot shows the quartiles of the dataset while the whiskers extend encompass

the rest of the distribution but leave out the points that are the outliers.

**Program:**

```
import matplotlib.pyplot as plt  
x=[1,2,3,4,5,6,7]  
y=[1,2,4,5,3,6,9]  
z=[x,y] plt.boxplot(z,labels=['A','B'],showmeans=True)  
plt.show()
```

**Output:**

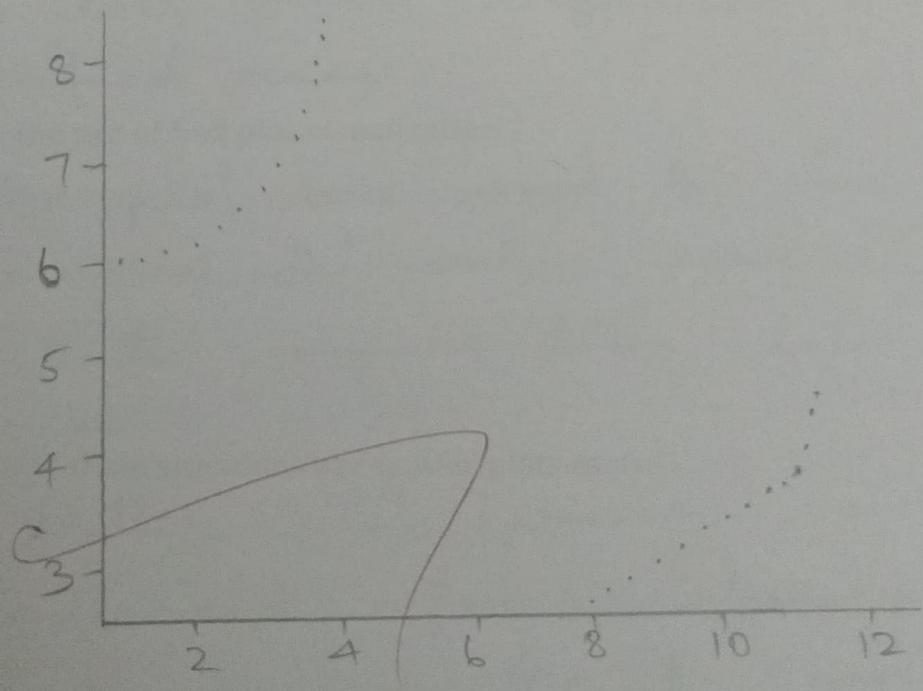
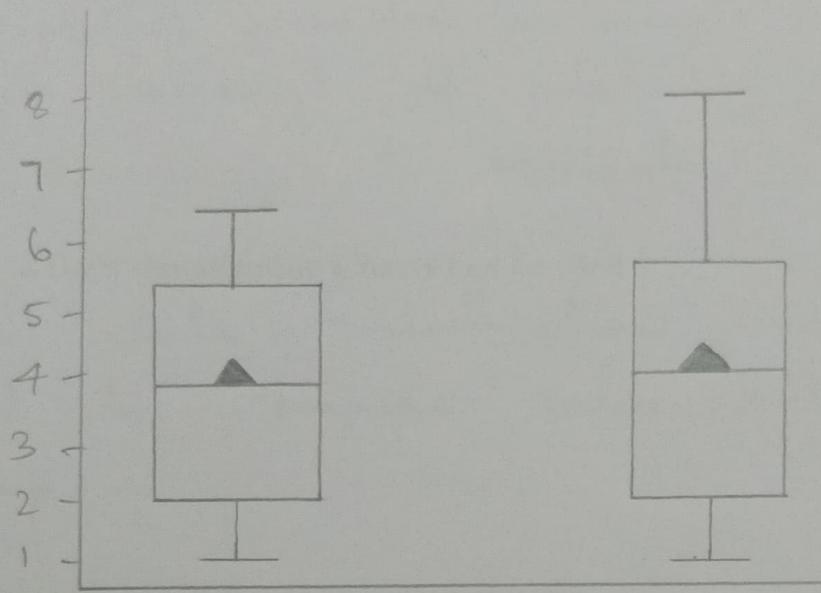
## VII) Scatter Plots:

A Scatter chart, also called a scatter plot, is a chart that shows the relationship between two variables.

**Program:**

```
import matplotlib.pyplot as plt  
x=[1,1.5,2,2.5,3,3.5,3.6]  
y=[7.5,8,8.5,9,9.5,10,10.5]  
x1=[8,8.5,9,9.5,10,10.5,11]  
y1=[3,3.5,3.7,4,4.5,5,5.2]  
plt.scatter(x,y, label='high income low saving',color='r')  
plt.scatter(x1,y1,label='low income high savings',color='b')  
plt.xlabel('saving*100')  
plt.ylabel('income*1000')  
plt.title('Scatter Plot')  
plt.legend()  
plt.show()
```

## Output



## Viva Questions :

1. List the few advantages of Matplotlib.

Matplotlib provides a simple way to access large amount of data.

It is easy to navigate.

2. Where in Data visualization Charts can be used?

In data visualization charts can be used to compare value between distinct groups.

3. Histograms are used for what in visualization.

A histogram is a graphical representation of data points organised into user-specified ranges.

4. What is the use of box plot visualization?

Box plot are used to show distributions of numeric data values, especially when you want to compare them between multiple groups.

5. In which of these situations are scatter plots useful?

Scatter plots primary uses are to observe and show relationship between two numeric values. It reports the patterns when they take as whole data.

Ex. No:	3	Basic Visualization in R
Date:	3/7/24	

## AIM

To implement the various fundamental data visualizations using R language .

### DESCRIPTION:

ggplot2 is an open-source data visualization package for the statistical programming language R. ggplot is enriched with customized features to make visualization better. ggplot2 is a system for declaratively creating graphics, based on the Grammar Of Graphics .

The ggplot2 package can be easily installed using the following R function:

```
install.packages(ggplot2)
```

then the following command must be used in program to use

```
ggplot package: library(ggplot2)
```

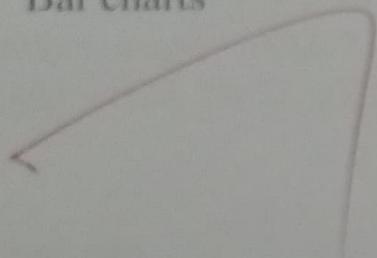
Consider the following dataset named **surveys**.

All the visualizations mentioned above are applied on this dataset.

```
Surveys<-data.frame(
  record_id=c(1,2,3,4,5),
  month=c(7,7,7,7,7),
  day=c(16,16,16,17,17),
  year=c(1977,1977,1977,1977,1977),
  plot_id=c(2,3,2,7,3),
  species_id=c(NL,NL,DM,DM,DM),
  sex=c(M,M,F,M,M),
  hindfoot_length=c(32,33,37, 36,35))
weight = c(20, 22, 25, 23, 24) # Added weight variable for demonstration
```

)  
The visualizations are carried out for the following:

1. Scatter plots
2. Line plots
3. Box plots
4. Histograms
5. Bar charts



## 1. Scatter plot

Program:

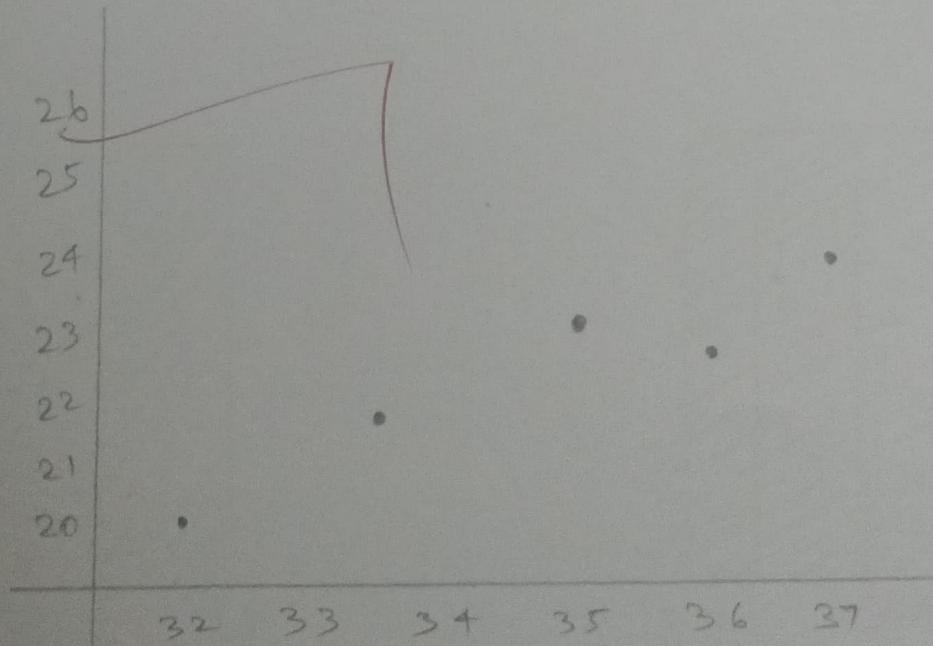
```
library(ggplot2)
```

```
Surveys <- data.frame(  
  record_id = c(1, 2, 3, 4, 5),  
  month = c(7, 7, 7, 7, 7),  
  day = c(16, 16, 16, 17, 17),  
  year = c(1977, 1977, 1977, 1977, 1977),  
  plot_id = c(2, 3, 2, 7, 3),  
  species_id = c("NL", "NL", "DM", "DM", "DM"),  
  sex = c("M", "M", "F", "M", "M"),  
  hindfoot_length = c(32, 33, 37, 36, 35),  
  weight = c(20, 22, 25, 23, 24) # Added weight variable for demonstration  
)
```

```
# Scatter plot
```

```
ggplot(data = Surveys, mapping = aes(x = hindfoot_length, y = weight)) +  
  geom_point(alpha = 0.1, color = "blue")
```

OutPut:



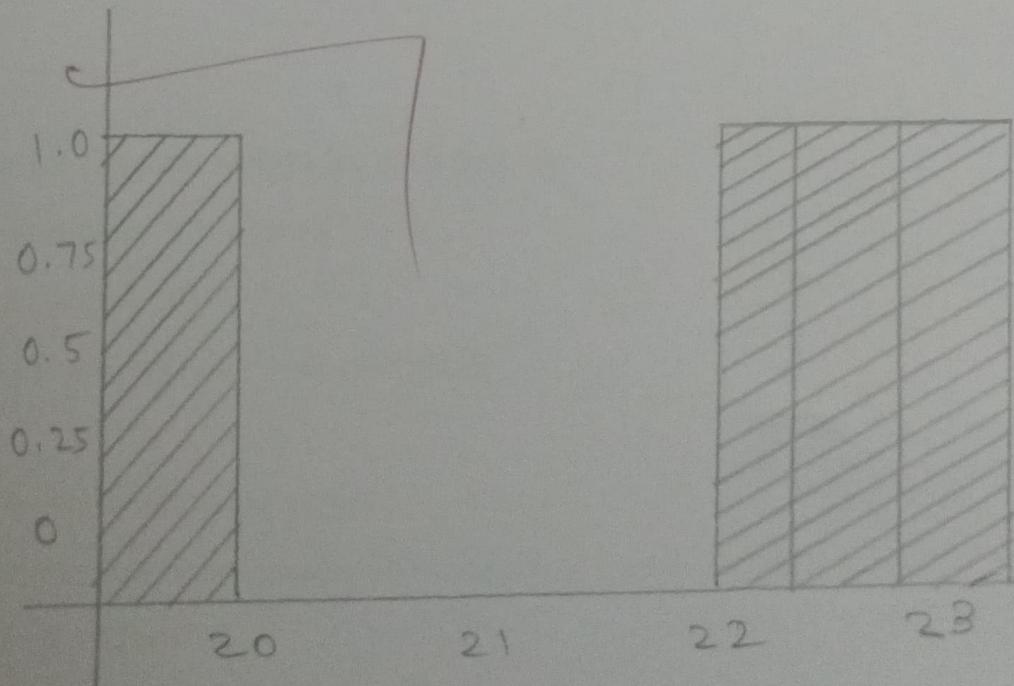
## 2) Histogram

Program

```
library(ggplot2)
Surveys <- data.frame(
  record_id = c(1, 2, 3, 4, 5),
  month = c(7, 7, 7, 7, 7),
  day = c(16, 16, 16, 17, 17),
  year = c(1977, 1977, 1977, 1977, 1977),
  plot_id = c(2, 3, 2, 7, 3),
  species_id = c("NL", "NL", "DM", "DM", "DM"),
  sex = c("M", "M", "F", "M", "M"),
  hindfoot_length = c(32, 33, 37, 36, 35),
  weight = c(20, 22, 25, 23, 24) # Added weight variable for demonstration
)

# Create a histogram of the weight variable
ggplot(data = Surveys, aes(x = weight)) +
  geom_histogram(binwidth = 1, fill = "blue", color = "black") +
  labs(x = "Weight", y = "Frequency", title = "Histogram of Weight")
```

Output:



### 3) Bar Chart

Program

```
library(ggplot2)
library(dplyr)

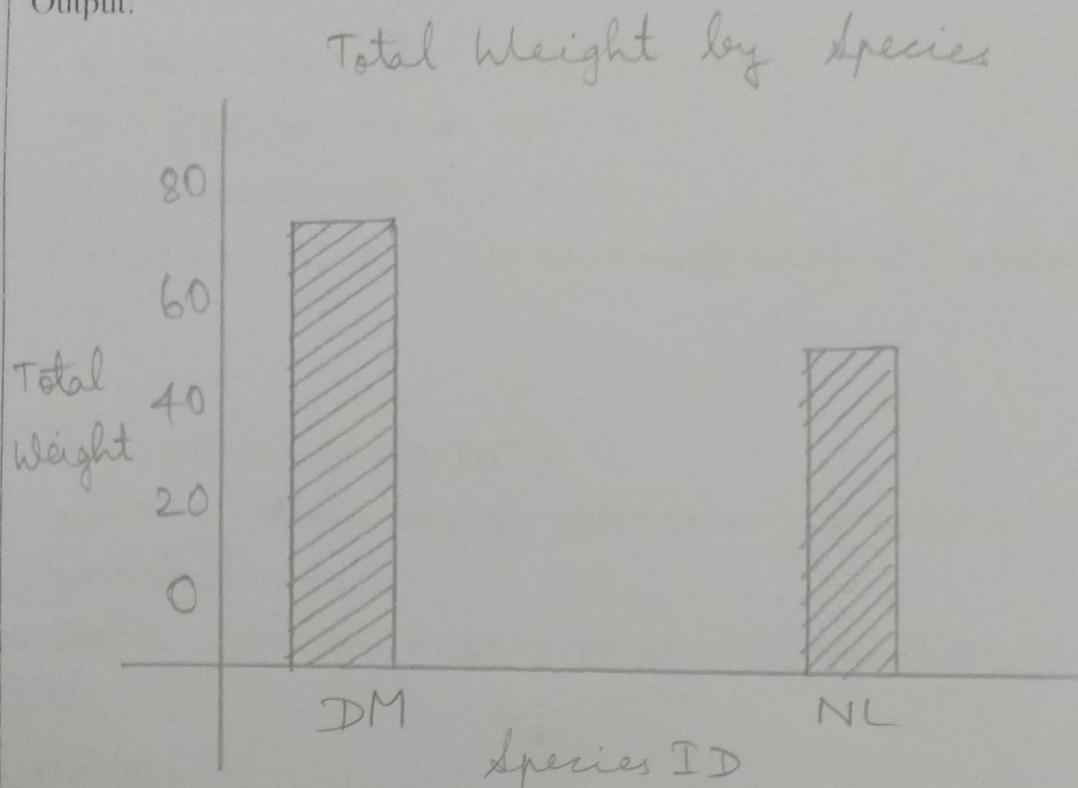
# Create the data frame
Surveys <- data.frame(
  record_id = c(1, 2, 3, 4, 5),
  month = c(7, 7, 7, 7, 7),
  day = c(16, 16, 16, 17, 17),
  year = c(1977, 1977, 1977, 1977, 1977),
  plot_id = c(2, 3, 2, 7, 3),
  species_id = c("NL", "NL", "DM", "DM", "DM"),
  sex = c("M", "M", "F", "M", "M"),
  hindfoot length = c(32, 33, 37, 36, 35),
  weight = c(20, 22, 25, 23, 24) # Added weight variable for demonstration
)

# Summarize the total weight for each species
Surveys_summary <- Surveys %>%
  group_by(species_id) %>%
  summarise(total_weight = sum(weight))

# Create the bar chart using ggplot2
ggplot(Surveys_summary, aes(x = species_id, y = total_weight)) +
  geom_col(fill = "Red") +
```

```
labs(title = "Total Weight by Species", x = "Species ID", y = "Total Weight")
```

Output:



#### 4.Box Plot:

##### Program

```
library(ggplot2)  
# Load required package  
library(ggplot2)
```

```
# Create the data frame  
Surveys <- data.frame(  
  record_id = c(1, 2, 3, 4, 5),  
  month = c(7, 7, 7, 7, 7),  
  day = c(16, 16, 16, 17, 17),
```

```

year = c(1977, 1977, 1977, 1977, 1977),
plot_id = c(2, 3, 2, 7, 3),
species_id = c("NL", "NL", "DM", "DM", "DM"),
sex = c("M", "M", "F", "M", "M"),
hindfoot_length = c(32, 33, 37, 36, 35),
weight = c(20, 22, 25, 23, 24) # Added weight variable for demonstration
)

```

# Create the box plot using ggplot2

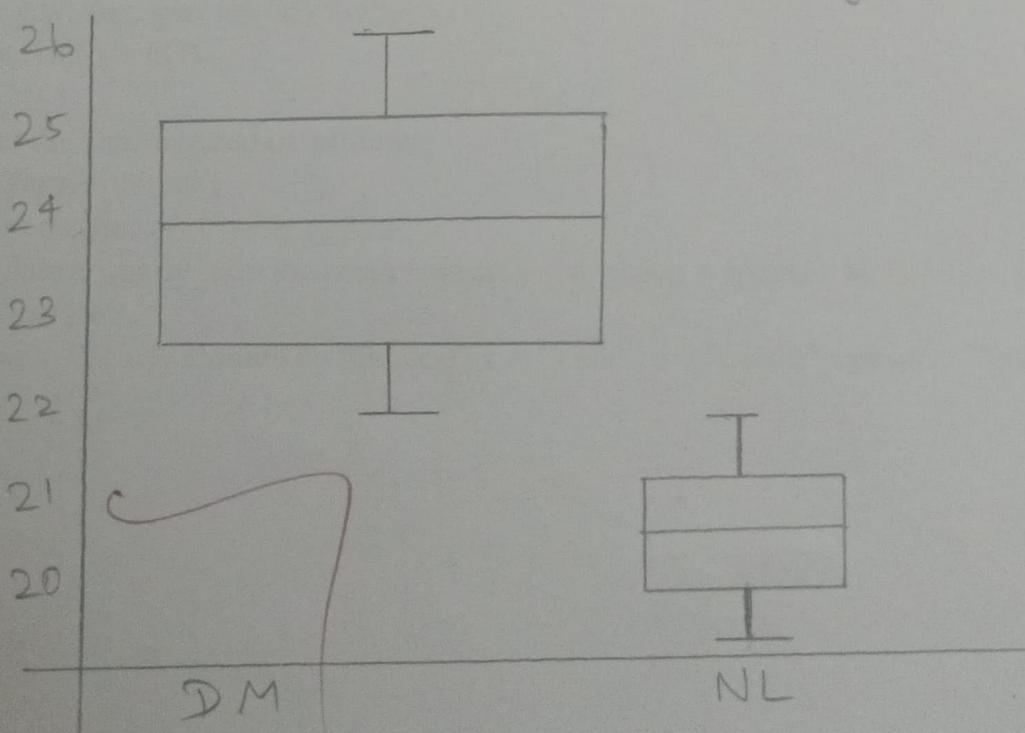
```

ggplot(data = Surveys, mapping = aes(x = species_id, y = weight)) +
geom_boxplot() +
labs(title = "Box Plot of Weight by Species", x = "Species ID", y = "Weight")

```

Output:

Box plot of weight by species



## 5) Line Plot

### Program

```
# Load required packages
library(ggplot2)
library(dplyr)

# Create example Surveys dataframe
Surveys <- data.frame(
  record_id = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10),
  month = c(7, 7, 7, 7, 7, 8, 8, 8, 8, 8),
  day = c(16, 16, 16, 17, 17, 16, 16, 16, 17, 17),
  year = c(1977, 1977, 1978, 1978, 1978, 1979, 1979, 1979, 1980, 1980),
  plot_id = c(2, 3, 2, 7, 3, 2, 3, 2, 7, 3),
  species_id = c("NL", "NL", "DM", "DM", "DM", "NL", "NL", "DM", "DM", "DM"),
  sex = c("M", "M", "F", "M", "M", "F", "M", "M", "F", "M"),
  hindfoot_length = c(32, 33, 37, 36, 35, 34, 32, 38, 37, 36),
  weight = c(20, 22, 25, 23, 24, 21, 23, 26, 24, 25)
)
# Summarize counts per year and species
yearly_counts <- Surveys %>%
  group_by(year, species_id) %>%
  summarise(n = n())

# Print the summarized dataframe
print(yearly_counts)

# Create the line plot using ggplot2
ggplot(data = yearly_counts, aes(x = year, y = n, group = species_id, colour = species_id)) +
  geom_line() +
  labs(title = "Yearly Counts by Species", x = "Year", y = "Count", colour = "Species ID")
```

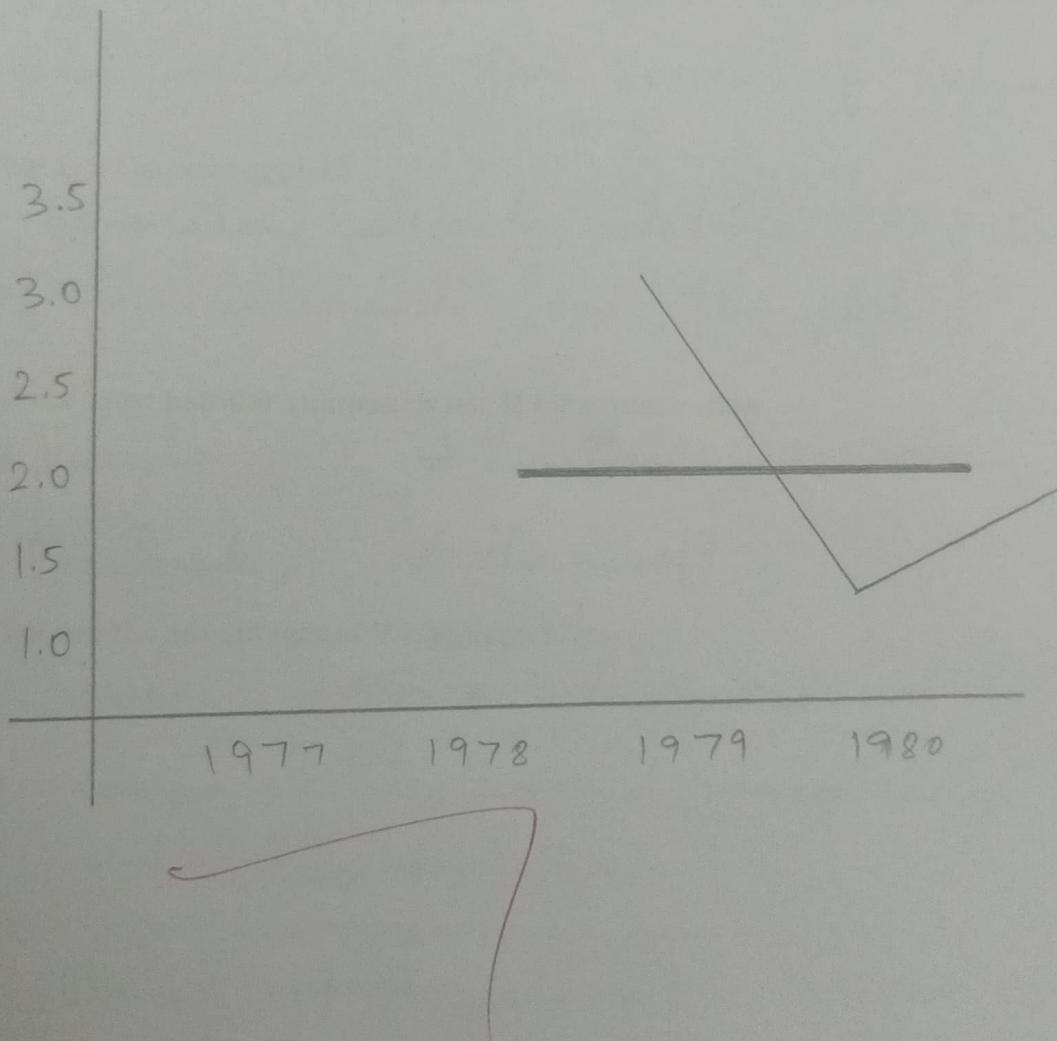
Output:

Species ID

— DM

— NL

Yearly count by species



### Viva Questions :

#### 1. What is the Use of R ?

R is widely used in data science by statistical and data miners for data analysis and the development.

#### 2. What is the difference between plot and ggplot2 in R?

ggplot2 is used for static plots.

plot is used for creating dynamic plots.

#### 3. What is a Geom in ggplot?

Geometric objects define the basic shape of the elements on the plot .

#### 4. List the most popular companies use R for visualization .

\* Google                  \* Twitter

\* Facebook                \* Microsoft

#### 5. What are the advantages of R language ?

Open source language

Compatibility and versatility

Extensions

Data visualization

Ex. No:	4
Date:	10/7/24

## Introduction to Tableau and Installation

### AIM

To implement the Tableau installation tool for performing the Data visualization.

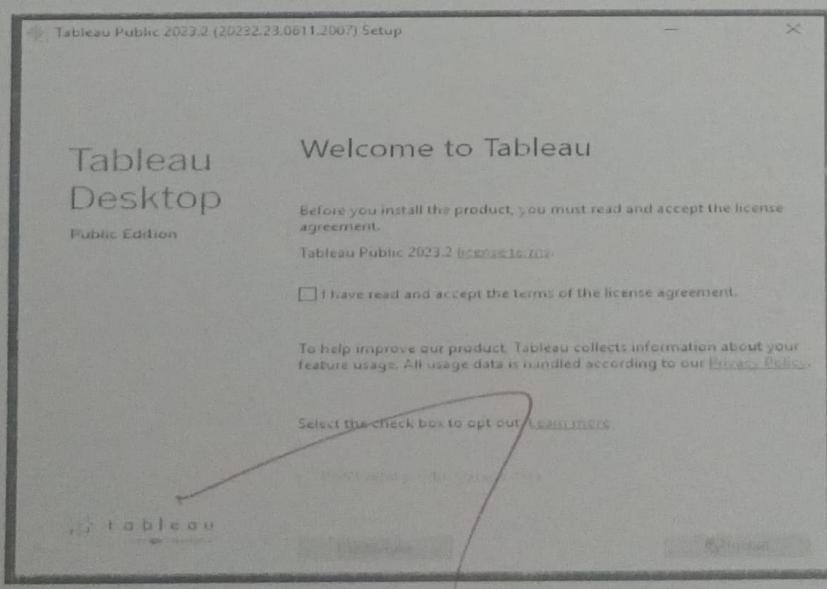
### DESCRIPTION:

**Tableau** is a data visualization tool that provides pictorial and graphical representations of data. It is used for data analytics and business intelligence. Tableau provides limitless data exploration without interrupting flow of analysis. With an intuitive drag and drop interface, user can uncover hidden insights in data and make smarter decisions faster.

Tableau can be downloaded from the following website:

<https://www.tableau.com/products/public/download>

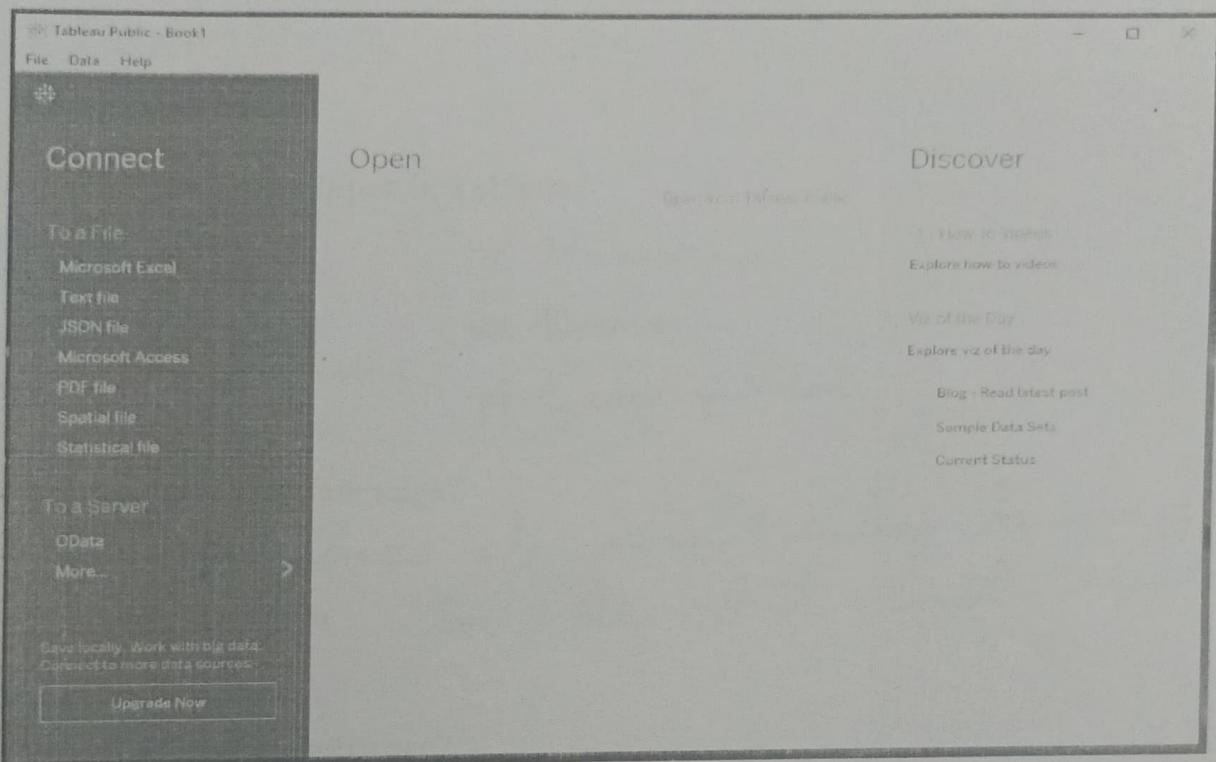
after downloading, the following is the screen appears.



Click the licence agreement checkbox and then click on install button.

After installation,click on Tableau Public icon to run Tableau.

Following is the Tableau Public home screen.



## Viva Questions

1. What are three types Tableau ?

Creator

Explorer

Viewer

2. What are the 7 data types in Tableau?

- ★ String                      ★ Numeric
- ★ Date                        ★ Boolean
- ★ Date & time              ★ Geographical

3. Is Tableau a coding language?

Tableau uses its own coding language. It is similar to SQL but not quite the same.

4. Is Tableau a tool or software?

Tableau software is an american interactive data visualization software company formed on business intelligence.

5. List few Best Features of Tableau.

Bins - Easily group data points to simplify analysis.

Catalog - enhanced data trust, visibility.

Result :

Thus the Tableau for Public was installed successfully.

10/7/24

Date: 24/7/24

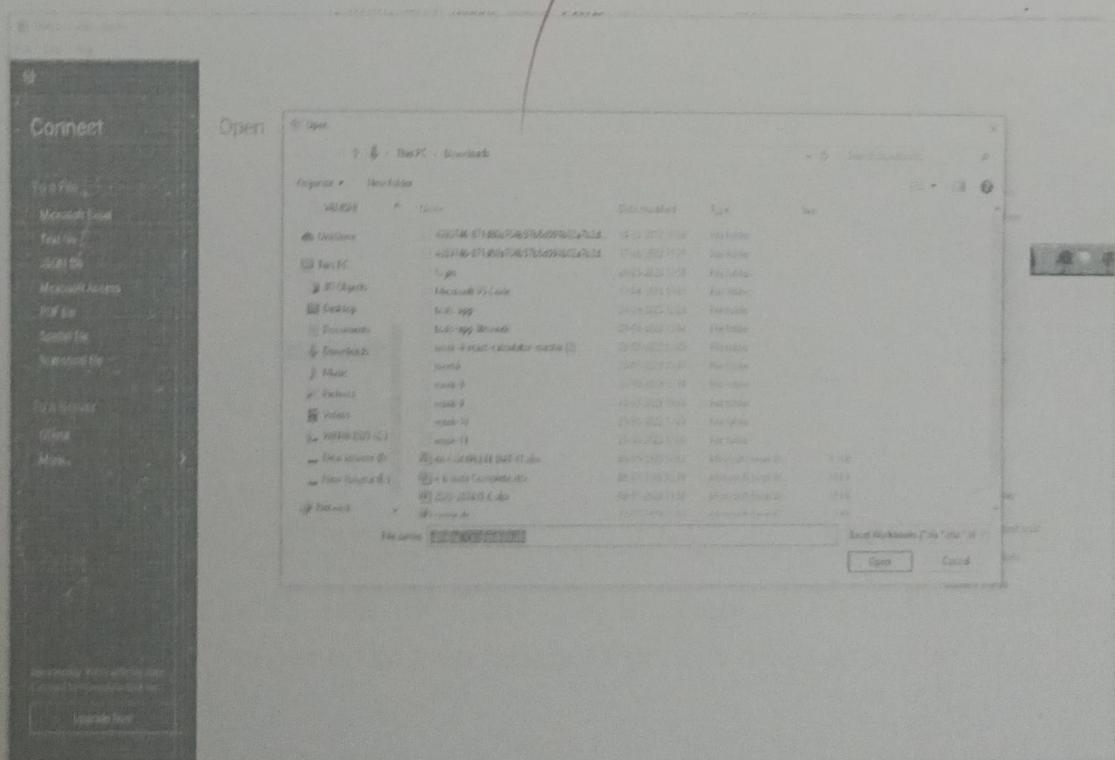
## Aim

To connect the data from the external source and prepare it for visualization

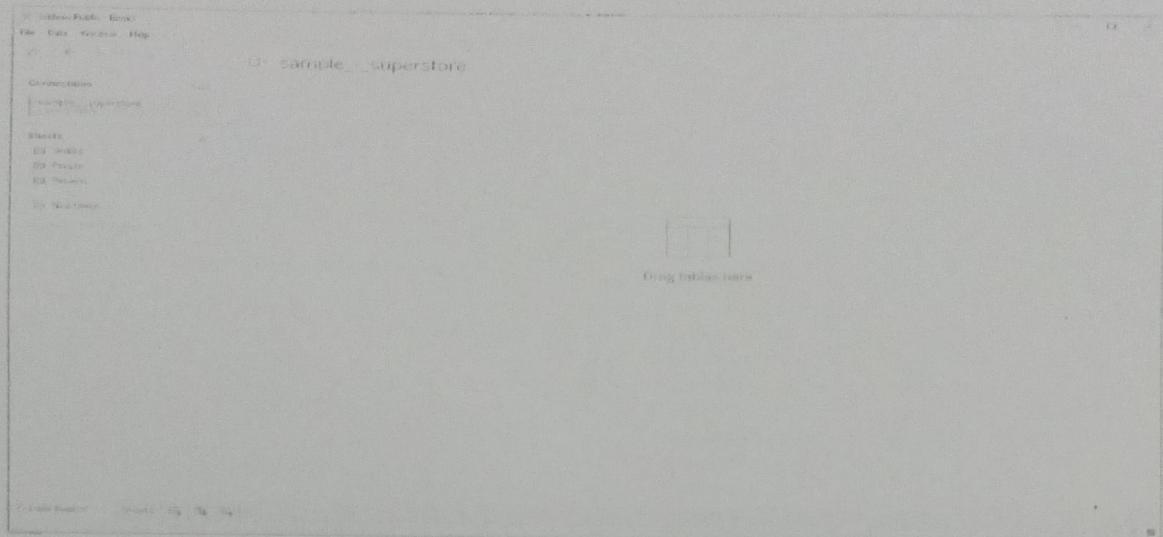
### DESCRIPTION:

**Tableau** supports connecting to a wide variety of data, stored in a variety of places. For example, data might be stored on computer in a spread sheet or a text file, or in a big data, relational, or cube (multidimensional) database on a server in enterprise or the data can be from a public domain available on the web.

Data can be imported in Tableau Public from Connect panel on left side. For example, an Excel sample data set was loaded into Tableau as follows:



After clicking on open, screen is as follows:



The data store page appears as above. The left pan shows that above dataset consists of 3 worksheets. If we drag orders table, screen appears as follows: Tableau automatically identifies the data type of each column.

This screenshot shows the 'Orders' table from the 'sample\_superstore' dataset. The table has 10 columns: Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Status, Type, Order ID, and Order Date. The 'Customer ID' and 'Customer Name' columns are identified as 'String' type, while the others are 'Date'. A preview of the data is shown below:

Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Status	Type	Order ID	Order Date
CA-2009-152356	06-11-2010	18-11-2010	Second Class	LG41250	Glare	Shipped	Shipped	CA-2009-152356	06-11-2010
CA-2009-152356	06-11-2010	11-11-2010	Second Class	LG41250	Glare	Shipped	Shipped	CA-2009-152356	06-11-2010
CA-2009-152356	12-09-2010	16-09-2010	Second Class	GW41043	Glare	Shipped	Shipped	CA-2009-152356	12-09-2010
US-2010-038266	11-10-2010	16-10-2010	Standard Class	SV43393	Glare	Shipped	Shipped	US-2010-038266	11-10-2010
US-2010-038266	11-10-2010	18-10-2010	Standard Class	SV43393	Glare	Shipped	Shipped	US-2010-038266	11-10-2010
CA-2014-128212	26-06-2014	14-07-2014	Standard Class	SI41270	Glare	Shipped	Shipped	CA-2014-128212	26-06-2014
CA-2014-128212	26-06-2014	14-07-2014	Standard Class	SI41270	Glare	Shipped	Shipped	CA-2014-128212	26-06-2014

Now drag Returns table onto the Canvas to the right of Orders table. This shows the relation between the two tables Orders and Returns.

If we click on the link between Orders and Returns table names at the top gives the summary of the relationship between the tables. Now rename the data store and click on Sheet1 at the

bottom left to proceed. This step creates a data extract which improves query performance.

**1. What is the use of connect in tableau**

Data connect service include cluster monitoring and trouble shooting. Tableau monitors the health and usage.

**2. How to load the data into tableau sheet**

on the data menu select a data source and then select extract and append data from file.

**3. What are dimensions in tableau with example?**

Dimensions split up the dataset into different categories and several different levels of details in the data.

**4. What is a measurement in tableau?**

According to tableau's knowledge, base a measure is a field that is a dependent variable, that is its value is a function of one or more dimensions.

**5. What is the use of SHOW ME in tableau?**

Tableau equipped with a function called 'show me' which seeks to teach and enable users to employ visual best practice.